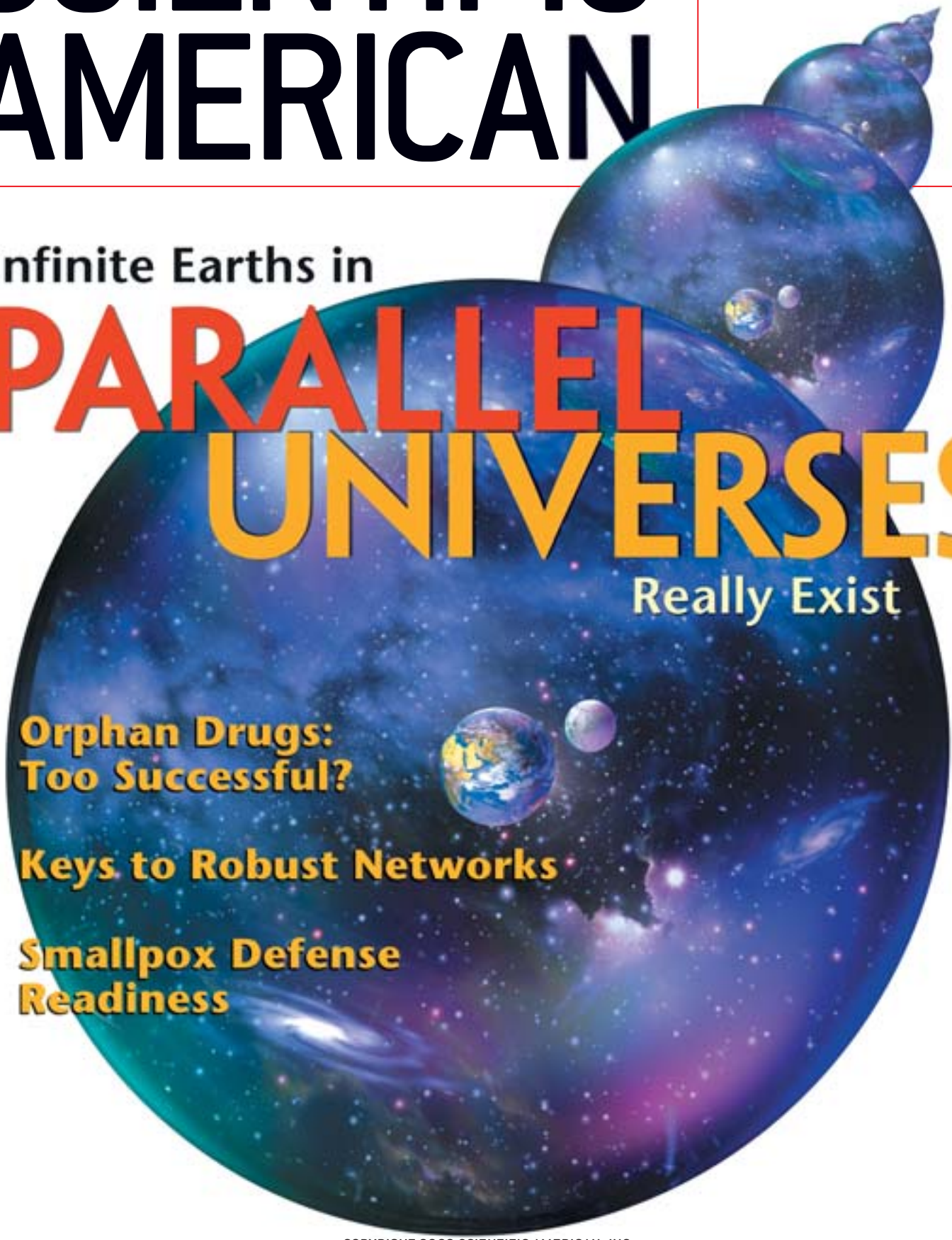
HEARING COLORS, TASTING SHAPES • ICEMAN REVISITED

SCIENTIFIC AMERICAN

MAY 2003 $4.95
WWW.SCIAM.COM



Infinite Earths in

PARALLEL UNIVERSES

Really Exist

Orphan Drugs: Too Successful?

Keys to Robust Networks

Smallpox Defense Readiness

may 2003

contents

features

SCIENTIFIC AMERICAN Volume 288 Number 5

COSMOLOGY

40 Parallel Universes

BY MAX TEGMARK

Not only are parallel universes—a staple of science fiction—probably real, but they could exist in four different ways. Somewhere out there our universe has a twin.

NEUROSCIENCE

52 Hearing Colors, Tasting Shapes

BY VILAYANUR S. RAMACHANDRAN AND EDWARD M. HUBBARD

In the extraordinary world of synesthesia, senses mingle together—revealing some of the brain's mysteries.

INFORMATION TECHNOLOGY

60 Scale-Free Networks

BY ALBERT-LÁSZLÓ BARABÁSI AND ERIC BONABEAU

Fundamental laws that organize complex networks are the key to defending against computer hackers, developing better drugs, and much more.

ARCHAEOLOGY

70 The Iceman Reconsidered

BY JAMES H. DICKSON, KLAUS OEGGL AND LINDA L. HANDLEY

Painstaking research contradicts many of the early speculations about the 5,300-year-old Alpine wanderer.

BIOTECHNOLOGY

80 The Orphan Drug Backlash

BY THOMAS MAEDER

Thanks to a 1983 law, pharmaceutical makers have turned drugs for rare diseases into profitable blockbusters. Has that law gone too far?

40 Infinite Earths in the multiverse

www.sciam.com

SCIENTIFIC AMERICAN Volume 288 Number 5

departments

8 SA Perspectives
Misguided missile shield.

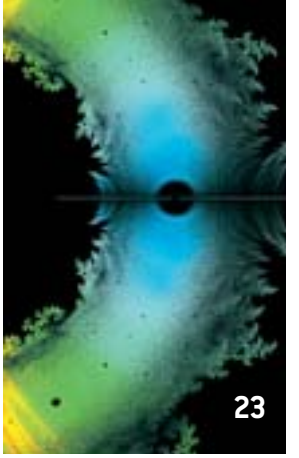
10 How to Contact Us

10 On the Web

12 Letters

18 50, 100 & 150 Years Ago

20 News Scan
- Resistance to smallpox vaccines.
- Cutting-edge math with supercomputers.
- The spam-filter challenge.
- Disappointment over VaxGen's AIDS vaccine.
- Will whale worries sink underwater acoustics?
- How Earth sweeps up interstellar dust.
- By the Numbers: Recidivism.
- Data Points: Advancing glaciers, rising seas.



34 Innovations
To save himself from radiation, a physician enters the rag trade.

36 Staking Claims
Patents let private parties take law into their own hands.

38 Insights
Paul Ginsparg started a Western Union for physicists. Now his idea is changing how scientific information is communicated worldwide.

88 Working Knowledge
Antennas, from rabbit ears to satellite dishes.

90 Voyages
Namibia's arid expanses are home to a menagerie of creatures that live nowhere else.

94 Reviews
A trio of books traces the quest to prove a prime-number hypothesis.

columns

37 Skeptic BY MICHAEL SHERMER
No body of evidence for cryptic critters.

96 Puzzling Adventures BY DENNIS E. SHASHA
Competitive analysis and the regret ratio.

98 Anti Gravity BY STEVE MIRSKY
From sheep to sheepskins in the field of genes.

99 Ask the Experts
Why do computers crash? What causes thunder?

100 Fuzzy Logic BY ROZ CHAST

Cover image by Alfred T. Kamajian

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. 242764. Canadian BN No. 127387652RT; QST No. Q1015332537. Subscription rates: one year $34.97, Canada $49 USD, International $55 USD. Postmaster: Send address changes to Scientific American, Box 3187, Harlan, Iowa 51537. **Reprints available:** write Reprint Department, Scientific American, Inc., 415 Madison Avenue, New York, N.Y. 10017-1111; (212) 451-8877; fax: (212) 355-0408 or send e-mail to sacust@sciam.com **Subscription inquiries: U.S. and Canada (800) 333-1199; other (515) 247-7631.** Printed in U.S.A.

Misguided Missile Shield

Imagine that you are a police officer in a tough neighborhood where the criminals are heavily armed. You go to a maker of bulletproof vests, who proudly claims that his latest product has passed five of its past eight tests. Somewhat anxious, you ask, "Did three of the bullets go through the vest?" The vest maker looks sheepish: "Well, we didn't actually fire bullets at it. We fired BBs. But don't worry, we're going to keep working on it. And, hey, it's better than nothing, right?"

The faulty vest is roughly analogous to America's unproved system for shooting down nuclear-tipped missiles. Over the next two years the Bush administration plans to deploy 20 ground-based missile interceptors in Alaska and California and 20 sea-based interceptors on U.S. Navy Aegis cruisers. The interceptors are designed to smash into incoming warheads in midflight. Ordinarily, the Department of Defense would be required to fully test the interceptors before installing them in their silos. The Pentagon, however, has asked Congress to waive this requirement. The reason for the rush is North Korea, which is believed to already possess two nuclear devices and is trying to develop intercontinental missiles that could hit the U.S.

The administration's approach might make sense if the missile shield showed true promise. The Pentagon's Missile Defense Agency (MDA) has conducted eight flight tests since 1999, launching mock warheads from California and interceptors from Kwajalein Atoll in the Pacific. In five of the attempts, the interceptor homed in on and destroyed the warhead; in two trials, the interceptor did not separate from its booster rocket, and in one, its infrared sensors failed. These exercises, however, have been far from realistic. Because the MDA's high-resolution radar system is still in development, the agency tracked the incoming missiles with the help of radar beacons placed on the mock warheads. The three-stage boosters planned for the interceptors are also not ready yet, so the MDA used two-stage Minuteman boosters instead. As a result, the interceptors traveled much more slowly than they would in an actual encounter and thus had more time to distinguish between the mock warheads and the decoys launched with them. Furthermore, the spherical balloons used as decoys in the tests did not resemble the mock warheads; the infrared signatures of the balloons were either much brighter or much dimmer.


MISSILE INTERCEPTOR begins a test flight.

Secretary of Defense Donald Rumsfeld says the MDA will fix the missile shield's problems as the system becomes operational. But many defense analysts believe it is simply infeasible at this time to build a missile interceptor that cannot be outwitted by clever decoys or other countermeasures [see "Why National Missile Defense Won't Work," by George N. Lewis, Theodore A. Postol and John Pike; Scientific American, August 1999]. A patchy missile shield could be more dangerous than none at all. It could give presidents and generals a false sense of security, encouraging them to pursue reckless policies and military actions that just might trigger the first real test of their interceptors.

Moreover, the most immediate peril from North Korea does not involve intercontinental missiles. It would be much easier for North Korea (or Iran or Al Qaeda) to smuggle a nuclear device into the U.S. in a truck or a container ship. Instead of spending $1.5 billion to deploy missile interceptors, the Bush administration should direct the money to homeland security and local counterterrorism programs, which are still woefully underfunded. And the Pentagon should evaluate the prospects of missile defense objectively rather than blindly promoting it.

THE EDITORS editors@sciam.com

How to Contact Us

EDITORIAL

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

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

SUBSCRIPTIONS

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

REPRINTS

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

PERMISSIONS

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

ADVERTISING

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

New York
Scientific American
415 Madison Ave.
New York, NY 10017-1111
212-451-8893
fax: 212-754-1138
Los Angeles
310-234-2699
fax: 310-234-2670
San Francisco
415-403-9030
fax: 415-403-9033
Midwest
Derr Media Group
847-615-1921
fax: 847-735-1457
Southeast/Southwest
MancheeMedia
972-662-2503
fax: 972-662-2577
Detroit
Karen Teegarden & Associates
248-642-1773
fax: 248-642-6138
Canada
Derr Media Group
847-615-1921
fax: 847-735-1457
U.K.
The Powers Turner Group
+44-207-592-8331
fax: +44-207-630-9922
France and Switzerland
PEM-PEMA
+33-1-46-37-2117
fax: +33-1-47-38-6329
Germany
Publicitas Germany GmbH
+49-211-862-092-0
fax: +49-211-862-092-21
Sweden
Publicitas Nordic AB
+46-8-442-7050
fax: +49-8-442-7059
Belgium
Publicitas Media S.A.
+32-(0)2-639-8420
fax: +32-(0)2-639-8430
Middle East and India
Peter Smith Media & Marketing
+44-140-484-1321
fax: +44-140-484-1320
Japan
Pacific Business, Inc.
+813-3661-6138
fax: +813-3661-6139
Korea
Biscom, Inc.
+822-739-7840
fax: +822-732-3662
Hong Kong
Hutton Media Limited
+852-2528-9135
fax: +852-2528-9281

On the Web

WWW.SCIENTIFICAMERICAN.COM

FEATURED THIS MONTH

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

Tanzanian Fossil May Trim Human Family Tree

A long-standing debate among scholars of human evolution centers on the number of hominid species that existed in the past. Whereas some paleoanthropologists favor a sleek family tree, others liken the known fossil record of humans to a tangled bush. The latter view has gained popularity in recent years, but a new fossil from Tanzania suggests that a bit of pruning might be in order. Researchers report that a specimen unearthed from Olduvai Gorge—a site made famous several decades ago by Louis and Mary Leakey—bridges two previously established species, indicating that they are instead one and the same.

The Economics of Science

After months of delay and uncertainty, the U.S. Congress finished work on the 2003 budget in February, approving large spending increases for the National Institutes of Health and the National Science Foundation. Science advocates worry that 2004 could still see a dramatically smaller boost. But would science necessarily suffer if government spending stopped rising? No, says Terence Kealey, a clinical biochemist and vice-chancellor of the University of Buckingham in England. His 1996 book, *The Economic Laws of Scientific Research*, claims that government science funding is not critical to economic growth, because science flourishes under the free market.

Ask the Experts

How does relativity theory resolve the Twin Paradox?

Ronald C. Lasky of Dartmouth College explains.
www.sciam.com/askexpert_directory.cfm

Scientific American DIGITAL

More than just a digital magazine!

SIGN UP NOW AND GET:

- The current issue each month, before it hits the newsstand
- Every issue of *Scientific American* from the past 10 years
- Exclusive online issues FREE (a savings of $30 alone)

Subscribe to Scientific American DIGITAL today and save!
www.sciamdigital.com/index.cfm?sc=ontheweb

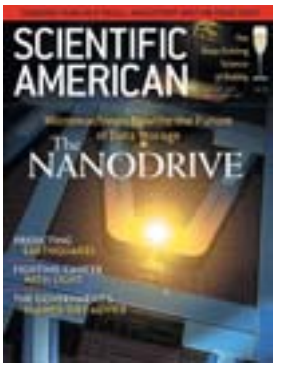
COURTESY OF R. J. BLUMENSCHINE

SCIENTIFIC AMERICAN®
Established 1845

Letters

EDITORS@SCIAM.COM

IS READING *Scientific American* good for you? Several articles in January educated readers about various health matters. "New Light on Medicine," by Nick Lane, described how light could activate compounds for treating certain ailments. A feature proposing a revised food pyramid put regular exercise at the foundation of a healthful lifestyle. Even housework counts—that activity helped to reduce the risk of dying for the elderly by almost 60 percent in one study, noted in News Scan's Brief Points. In response, Richard Hardwick sent an offer via e-mail that may be—okay, we'll say it—nothing to sneeze at: "As the occupant of one of Europe's major dust traps, I feel I can sustain a whole army of elderly duster-wielding would-be immortals. I offer access to my dust on a first-come, first-served basis; vacuum cleaner supplied, but applicants must bring their own dusters." Other reactions to the fitness of the January issue follow.



THERAPY WITH LIGHT

Nick Lane's otherwise excellent article on photodynamic therapy (PDT), "New Light on Medicine," fails to credit the scientific founders of the field, who deserve to be better known. These were the medical student Otto Raab and his professor Hermann von Tappeiner of the Pharmacological Institute of Ludwig-Maximilians University in Munich, Germany. They were active in the opening years of the 20th century. Von Tappeiner and another colleague later published the case history of a patient with basal cell carcinoma who was cured through an early form of PDT that used the coal tar dye eosin as a photosensitizer.

Ralph W. Moss
State College, Pa.

Surely, as Lane speculates, the rare sighting of a porphyria victim scuttling out at night might have strengthened vampire or werewolf beliefs in specific locales and could have stimulated a craze. It's also possible that a heme-deprived porphyriac might crave blood. But we don't need actual victims of porphyria to explain legends of bloodsucking humanoid creatures of the night. Such beliefs are widespread and part of fundamental human fears that are probably deeply rooted in our evolutionary biology.

Phillips Stevens, Jr.
Department of Anthropology
State University of New York at Buffalo

FOOD FIGHT

"Rebuilding the Food Pyramid," by Walter C. Willett and Meir J. Stampfer, discourages the consumption of dairy products, presumably because of the fat content. Does this hold true for nonfat milk, yogurt and other low- or reduced-fat dairy products?

Maureen Breakiron-Evans
Atherton, Calif.

Where does corn fit on the new food pyramid? Is it a grain or a vegetable?

Robin Cramer
Solana Beach, Calif.

The authors state that the starch in potatoes is metabolized into glucose more readily than table sugar, spiking blood sugar levels and contributing to insulin resistance and the onset of diabetes. I've heard that combining carbohydrates with proteins or fats in a single meal can slow the absorption of the carbohydrates, reducing that effect. Would it follow that french fries and potato chips cooked in healthful monounsaturated or polyunsaturated oils are better for you than a boiled potato? Can decent french fries and potato chips be made using the healthful oils instead of trans-fats?

Phil Thompson
Los Altos, Calif.

One of the main arguments made in the food pyramid article is that the 1992 USDA

Letters

Food Guide Pyramid oversimplified dietary recommendations. Ironically, the article itself falls prey to similar problems in its discussion of carbohydrates and vegetables.

The authors discuss the detrimental effects of diets high in carbohydrates, especially "refined carbohydrates," and imply that potatoes should fall into that category. To be fair, the starch in potatoes should be treated with the same consideration as the starch in grain. A key aspect that differentiates whole grains from refined grains is the greater amount of fiber in the former; whole potatoes have about as much fiber per calorie as whole grains.

The article also misrepresents the nutritional value of potatoes. It says that the potato should not be considered a vegetable, but whole potatoes contain plenty of the nutrients that Willett and Stampfer attribute to what they call vegetables. Although each vegetable has its strong and weak points, potatoes compare favorably with other vegetables nutritionally. If potatoes were such an empty food, how did many Irish peasants live almost exclusively on them in the 18th and 19th centuries?

Besides these points, Willett and Stampfer's position may benefit from a review of the literature regarding the antioxidant content of potatoes. Much research shows that potatoes are high in certain classes of antioxidants.

Andrew Jensen
Washington State Potato Commission

WILLETT AND STAMPFER REPLY: Clearly, nonfat dairy products are preferable to those with full fat. Other concerns remain, however. Several studies find that high calcium intake, from dairy products or supplements, is associated with a higher risk of prostate cancer; preliminary evidence also suggests a link with ovarian cancer. We recommend consuming dairy products in moderation.

Corn should be considered a grain. It has a lower glycemic index than potatoes, thus raising blood sugar to a lesser extent. Popcorn has a similar nutritional profile to corn

and can be a good snack food, depending on how it is prepared. Nuts, however, would be a superior choice.

It is certainly possible to prepare good-tasting french fries using healthful oils instead of those loaded with trans-fats. The extent to which mixed meals raise blood sugar is a function of the different foods in the meal. Thus, replacing some of the calories from a baked potato with those from healthful fats used in frying a spud would probably have an overall health benefit. Eating foods that have a lower glycemic index would be even better.

Many of the potato's nutrients are in its skin, which is rarely eaten. Even with the skin, potatoes contain a relatively large amount of high-glycemic carbohydrates. The basis of our placement of potatoes comes not just from this evidence but also from the epidemiology data. In a major review by the World Cancer Research Fund, potatoes were the only vegetable found not to help in reducing the risk of cancer. Our studies show that potatoes are the food most frequently associated with type 2 diabetes risk. Unlike other vegetables, potatoes do not appear to reduce the risk of coronary heart disease but have a weak positive effect. When we compare potatoes with other sources of starch, such as whole grains, they do not fare well either: unlike potatoes, whole grains are consistently associated with lower risks of diabetes and coronary heart disease.

Potatoes appear to be at best empty calories compared with alternatives and thus a lost opportunity for improved health. Of course, they could enable you to survive famine, but that hardly describes our current situation: the glycemic load was much less of an issue for lean, highly active farmers in Ireland or in this country 100 years ago than it is today.

DETECTING NUCLEAR TESTS

I read Ross S. Stein's article on stress transfer and seismicity, "Earthquake Conversations." Having just finished a class paper on seismic detection of nuclear tests, I began wondering about possible connections. I know that nuclear tests often result in shock waves of magnitude 4 to 6. I also read that although 20 to 30 percent of this energy is "earthquakelike," nuclear tests generally do not cause earthquakes. Could the tests change regional seismicity through a process similar to the one Stein describes? Would it be possible, for instance, to plug nuclear-test blasts, such as the hundreds that took place in Nevada, into his stress-transfer model to see if the changes in seismicity that it predicts correspond to real-world changes?

Dan Koik
Georgetown University

STEIN REPLIES: *It is certainly possible that regional seismicity has been affected by nuclear blasts. Volcanic eruptions share some similarities to nuclear blasts, and they clearly have altered seismicity. That interaction has been especially notable between historical eruptions of Mount Vesuvius and large Apennine earthquakes in Italy, according to some of my team's recent work. But to accurately detect a possible change in seismicity rate around the site of a nuclear blast would require a very dense seismic network, which was not used for any past test blasts. Nuclear blasts are explosion or implosion sources, rather than shear sources. Our downloadable Mac program, Coulomb 2.2, can calculate the static stress changes imparted by a point source of expansion or contraction on surrounding faults. These results would reveal on which faults near a nuclear blast failure is promoted. I haven't looked at this problem, but someone should.*

ERRATUM "The Captain Kirk Principle," by Michael Shermer [Skeptic, December 2002], should have attributed the study of the effects of showing emotionally charged images to subjects to "Subliminal Conditioning of Attitudes," by Jon A. Krosnick, Andrew L. Betz, Lee J. Jussim and Ann R. Lynn in *Personality and Social Psychology Bulletin,* Vol. 18, No. 2, pages 152–162; April 1992.

50, 100 & 150 Years Ago

FROM SCIENTIFIC AMERICAN

Martian Reality ▪ Zeppelin Dreams ▪ Creationist Dogma

MAY 1953

OBJECTIVE MARS—"For nearly a century Mars has captivated the passionate interest of astronomers and the credulous imagination of the public—of which we had an example not long ago in the great 'Martian' scare instigated by a radio program. The facts, although not as exciting as the former speculations, are interesting enough. Easily the most conspicuous feature of the planet is the white caps that cover its polar regions. They display a fascinating rhythm of advance and retreat. At the end of winter in each hemisphere the polar cap covers some four million square miles. But even in midsummer a tiny dazzling spot remains near the pole. As to the fine structure of the 'canals' much uncertainty remains.—Gérard de Vaucouleurs"

ELECTIONS GO LIVE—"The presidential campaign of 1952 was the first in which television played a major part. In a University of Michigan study, the first noteworthy fact is that the public went out of its way to watch the campaign on television. Only about 40 per cent of the homes in the U.S. have TV sets, but some 53 per cent of the population saw TV programs on the campaign—a reflection of 'television visiting.' As to how television affected the voting itself, we have no clear evidence. Those who rated television their most important source of information voted for Dwight D. Eisenhower in about the same proportion as those who relied mainly on radio or newspapers. Adlai Stevenson did somewhat better among the television devotees."

MAY 1903

DUST STORM—"Elaborate researches have been carried out by two eminent scientists, Profs. Hellmann and Meinardus, relative to the dust storm which swept over the coasts of Northern Africa, Sicily, Italy, Austria-Hungary, Prussia and the British Isles between March 12 and 19 of 1901. The dust originated in storms occurring on March 8, 9 and 10 in the desert of El Erg, situated in the southern part of Algeria. Roughly 1,800,000 tons of dust were carried by a large mass of air which moved with great velocity from Northern Africa to the north of Europe. All the microscopic and chemical analyses point to this dust being neither volcanic nor cosmic."

YAKUT BELLE, Siberia, 1903

SIBERIAN EXPEDITION—"The Jesup North Pacific Expedition, sent out under the auspices of the American Museum of Natural History, has completed its field work. Remarkable ethnological specimens and discoveries were obtained in Siberia by the Russian explorers and scientists, Messrs. Waldemar Jochelson and Waldemar Bogoras. Our illustration shows the costume of a rich Yakut belle, the Yakuts being the largest and richest of the Siberian races. The striking feature of the garment, besides the genuine wealth of fur, is the lavish display of silver ornaments which adorn the front. The neck and shoulder bands of solid filigree-work are three inches wide and several yards long, finely executed. The object of the expedition, under the general supervision of Dr. Franz Boas, was to investigate the obscure tribes of northeastern Asia, and to compare their customs with the inhabitants of the extreme northwestern part of North America."

IN THE RED ZEPPELIN—"It is announced in Berlin that Count Zeppelin's airship shed on Lake Constance, together with his apparatus, will be sold at auction. The count is a poor man. He sank over one million marks in the enterprise."

MAY 1853

CREATIONISM DEVOLVES—"Prof. Louis Agassiz, in his recent course of lectures, delivered in Charleston, S.C., taught and proclaimed his disbelief in all men having descended by ordinary generation from Adam, or from one pair, or two or three pairs. He believes, as we learn from the 'Charleston Mercury,' that men were created in separate nations, each distinct nationality having had a separate origin. Prof. Agassiz has been bearding the lion in his den—we mean the Rev. Dr. Smyth, of Charleston, who has written a very able work on the unity of the human race, the Bible doctrine of all men being descended from a single pair, Adam and Eve. This is a scientific question, which, within a few years, has created no small amount of discussion among the lovers of the natural sciences."

BIOTERRORISM

Spotty Defense

BIG CITIES ARE LATE TO VACCINATE AGAINST SMALLPOX **BY MARK ALPERT**

The effort to build a defensive line against a terrorist smallpox attack is off to a slow start. Under the plan outlined last December by President George W. Bush, nearly half a million doctors, nurses and epidemiologists were supposed to be vaccinated against smallpox in a voluntary 30-day program beginning in late January. If terrorists were to bring smallpox to the U.S.—possibly by spraying the virus in airports or sending infected "smallpox martyrs" into crowded areas—the vaccinated health care workers would be responsible for treating the exposed individuals, tracking down anyone who may have come into contact with them, and running the emergency clinics for vaccinating the general public.

SMALLPOX VACCINE called Dryvax is being administered to health care workers across the U.S. In the event of a smallpox attack, vaccinated workers would treat exposed individuals.

TAMI CHAPPELL *Reuters/Corbis*

By mid-March, however, local health departments across the U.S. had vaccinated only 21,698 people. Some states responded promptly: for example, Florida (which inoculated 2,649 people in less than six weeks), Tennessee (2,373 people) and Nebraska (1,388). But health departments in America's largest cities, which are surely among the most likely targets of a bioterror attack, were lagging. By March 14 the New York City Department of Health had vaccinated only 51 people—50 members of its staff, plus Mayor Michael R. Bloomberg. The department planned to inoculate between 5,000 and 10,000 people to form smallpox response teams at 68 hospitals, but vaccinations at the first eight hospitals did not begin until March 17.

The pace was also slow in Los Angeles (134 inoculated by March 14) and Chicago (18). Washington, D.C., had vaccinated just four people, including the health department's director. "A lot of hospital administrators are still very wary," says Laurene Mascola, chief of the disease control program at the Los Angeles County Department of

Health Services. Much of the concern stems from the health risks of the vaccine itself, which caused one to two deaths and 14 to 52 life-threatening complications for every million doses when it was last used in the 1960s. The vaccine's fatality risk, however, is one hundredth the average death rate from motor vehicle accidents in the U.S. and one 200,000th the mortality rate from smallpox, which would be likely to kill 30 percent of the people infected.

U.S. intelligence officials suspect that both Iraq and North Korea possess stocks of smallpox. The big uncertainty is whether terrorists could spread the disease effectively—spraying the live virus over a wide area is technically difficult, and a smallpox martyr could not infect others until he or she was quite ill. Smallpox experts note, though, that the public would demand mass vaccinations even if only one case appeared in the U.S. and that health care workers might be unwilling to perform that task if they had not been previously vaccinated themselves. Says William J. Bicknell of the Boston University School of Public Health: "To vaccinate the whole country in 10 days, we'd need two to three million workers."

Only a few states have come close to that level of preparedness. Nebraska, which had one of the highest per-capita smallpox vaccination rates as of mid-March, benefited from the zeal of Richard A. Raymond, the state's chief medical officer, who personally lobbied administrators at dozens of hospitals. "Government is all about priorities, and this was a priority for us," Raymond says. "An attack may start in a big city, but because Americans are so mobile, the entire country is at risk."

Joseph M. Henderson, associate director for terrorism preparedness at the Centers for Disease Control and Prevention, notes that vaccinations are not the only defense against smallpox. New York City, for instance, has an excellent disease surveillance program, increasing the chances that epidemiologists would be able to identify and contain a smallpox outbreak. "Overall, New York gets a passing grade," Henderson says. "But they should have a lot more people vaccinated. They're doing it, but not as fast as we'd like."

BEYOND SMALLPOX

Smallpox is not the only bioterror agent that Iraq is believed to possess. Under pressure from the United Nations, Iraqi officials admitted in 1995 that their laboratories had churned out these bioweapons:

- **Botulinum toxin: nerve agent produced by the bacteria that cause botulism**
- **Anthrax: bacteria that lie dormant in spores; if inhaled, the bacteria multiply rapidly in the body, causing internal bleeding and respiratory failure**
- **Aflatoxin: chemical produced by fungi that grow on peanuts and corn; causes liver cancer**
- **Perfringens toxin: compound released by the bacteria that cause gas gangrene**

MATH

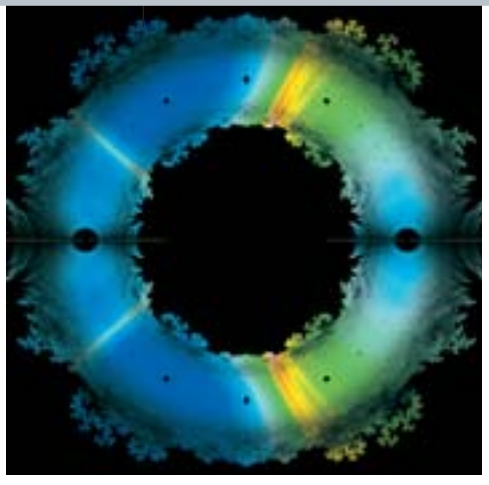
A Digital Slice of Pi

THE NEW WAY TO DO PURE MATH: EXPERIMENTALLY **BY W. WAYT GIBBS**

"One of the greatest ironies of the information technology revolution is that while the computer was conceived and born in the field of pure mathematics, through the genius of giants such as John von Neumann and Alan Turing, until recently this marvelous technology had only a minor impact within the field that gave it birth." So begins *Experimentation in Mathematics,* a book by Jonathan M. Borwein and David H. Bailey due out in September that documents how all that has begun to change. Computers, once looked on by mathematical researchers with disdain as mere calculators, have gained enough power to enable an entirely new way to make fundamental discoveries: by running experiments and observing what happens.

The first clear evidence of this shift emerged in 1996. Bailey, who is chief technologist at the National Energy Research Scientific Computing Center in Berkeley, Calif., and several colleagues developed a computer program that could uncover integer relations among long chains of real numbers. It was a problem that had long vexed mathematicians. Euclid discovered the first integer relation scheme—a way to work out the greatest common divisor of any two integers—around 300 B.C. But it wasn't until 1977 that Helaman Ferguson and Rodney W. Forcade at last found a method to detect relations among an arbitrarily large set of numbers. Building on that work, in 1995 Bailey's group turned its computers loose on some of the fundamental constants of math, such as log 2 and pi.

To the researchers' great surprise, after months of calculations the machines came up with novel formulas for these and other nat-



COMPUTER RENDERINGS of mathematical constructs can reveal hidden structure. The bands of color that appear in this plot of all solutions to a certain class of polynomials (specifically, those of the form $\pm 1 \pm x \pm x^2 \pm x^3 \pm \ldots \pm x^n = 0$, up to $n = 18$) have yet to be explained by conventional analysis.

COURTESY OF JONATHAN M. BORWEIN AND PETER BORWEIN *Simon Fraser University*

CRUNCHING NUMBERS

Mathematical experiments require software that can manipulate numbers thousands of digits long. David H. Bailey has written a program that can do math with arbitrary precision. That and the PSLQ algorithm that uncovered a new formula for pi are available at www.nersc.gov/~dhbailey/mpdist/

A volunteer effort is under way to verify the famous Riemann Hypothesis by using distributed computer software to search for the zeros of the Riemann zeta function. (German mathematician Bernhard Riemann hypothesized in 1859 that all the nontrivial zeros of the function fall on a particular line. See "Math's Most Wanted," Reviews, on page 94.) To date, more than 5,000 participating computers have found more than 300 billion zeros. For more information, visit www.zetagrid.net

ural constants. And the new formulas made it possible to calculate any digit of pi or log 2 without having to know any of the preceding digits, a feat assumed for millennia to be impossible.

There are hardly any practical uses for such an algorithm. A Japanese team used it to check very rapidly a much slower supercomputer calculation of the first 1.2 trillion digits of pi, completed last December. A pickup group of amateurs incorporated it into a widely distributed program that let them tease out the quadrillionth digit of pi. But mathematicians, stunned by the discovery, began looking hard at what else experimentation could do for them.

Recently, for example, the mathematical empiricists have advanced on a deeper question about pi: whether or not it is normal. The constant is clearly normal in the conventional sense of belonging to a common class. Pi is a transcendental number—its digits run on forever, and it cannot be expressed as a fraction of integers (such as $^{355}/_{113}$) or as the solution to an algebraic equation (such as $x^2 - 2 = 0$). In the universe of all known numbers, transcendental numbers are in the majority.

But to mathematicians, the "normality" of pi means that the infinite stream of digits that follow 3.14159… must be truly random, in the sense that the digit 1 is there exactly one tenth of the time, 22 appears one hundredth of the time, and so on. No particular string of digits should be overrepresented, whether pi is expressed in decimal, binary or any other base.

Empirically that seems true, not only for pi but for almost all transcendental numbers. "Yet we have had no ability to prove that even a single natural constant is normal," laments Borwein, who directs the Center for Experimental and Constructive Mathematics at Simon Fraser University in British Columbia.

"It now appears that this formula for pi found by the computer program may be the key that unlocks that door," Bailey says. He and Richard E. Crandall of Reed College have shown that the algorithm links the normality problem to other, more tractable areas of mathematics, such as chaos theory and pseudorandom number theory. Solve these related (and easier) problems, and you prove that pi is normal. "That would open the floodgates to a variety of results in number theory that have eluded researchers for centuries," Borwein predicts.

INTERNET

A Man, a Plan, Spam

A STANFORD LAWYER PITS HIS JOB AGAINST JUNK E-MAIL

BY WENDY M. GROSSMAN

Like most Internet users, Stanford University law professor Lawrence Lessig hates junk e-mail—or, as it is formally known, unsolicited commercial e-mail (UCE). In fact, he hates it so much, he's put his job on the line. "I think it will work," he says of his scheme for defeating the megabyte loads of penis extenders, Viagra offers, invitations to work at home, discount inkjet cartridges, and requests for "urgent assistance" to get yet another $20 million out of Nigeria.

Lessig, who wrote two influential books about the Internet and recently argued before the U.S. Supreme Court against the extension of copyright protection, has developed a two-part plan. The first part is legislative: pass federal laws mandating consistent labeling so that it would be trivial for users and Internet service providers (ISPs) to prefilter junk. Federal antispam legislation hasn't been tried yet, and unlike state laws—which have been enacted in 26 states since 1997, to little effect—it would have a chance at deterring American spammers operating outside the nation's borders. Second: offer a bounty to the world's computer users for every proven violator they turn in. Just try it, he says, and if it doesn't work, he'll quit his job. He gets to decide on the particular schemes; longtime sparring partner and CNET reporter Declan McCullagh will decide whether it has worked.

"Spam only pays now because [spammers] get to send 10 million e-mails and [they] know five million will be delivered and 0.1 percent will be considered and responded to," Lessig explains. "If all of a sudden you make

it very easy for people who don't want it to filter it out, then it doesn't pay to play the game anymore." The European Commission reported in February 2001 that junk e-mail costs Internet users some 10 billion euros (about $10.6 billion) worldwide, mostly in terms of lost time and clogged bandwidth. But that's almost certainly too low an estimate now. In 2002 the volume of junk went through the roof, as anyone who keeps an e-mail address can attest. Accounts now receive multiple copies of the same ad. AOL reported this past February that it filters out up to 780 million pieces of junk daily—an average of 22 per account.

Relying on spam blockers has led to an escalating e-mail-filtering arms race as UCE becomes ever more evasive. Because constructing effective filters is time-consuming, the trend is toward collaboration. SpamCop, for example, is a Microsoft Outlook–only peer-to-peer version of spam blocking: users report known junk to a pooled database, which is applied to everyone's e-mail. SpamAssassin is an open-source bit of heuristics that has been incorporated into plug-ins for most e-mail software; it works in a way similar to antivirus software, identifying junk mail that uses generic signatures ("You opted to receive this").

Some services maintain a "white list" of accepted correspondents and challenge e-mail messages from anyone new. If a person does not respond, the e-mail is discarded. But this approach is too hostile for businesses and organizations that must accept messages from strangers, who might after all be new customers.

So until Lessig's gambit pays off, the best strategy may be a combination of filters. As a parallel experiment, I've set up a mail server with its own filters and integrated SpamAssassin through my service provider. The jokes in my in-box may go out of date, but at least I'll be able to find Lessig's announcement of whether he's still employed.

Wendy M. Grossman, based in London, is at wendyg@pelicancrossing.net. Anyone sending UCE will be hunted down and made to work at home stuffing Viagra into inkjet cartridges.

SPAMMERS: MAKING THEM PAY

Clive Feather, a policy specialist for the U.K.'s oldest ISP, Demon Internet, thinks spammers need to pay for their antics. "If we had micropayments," he says, "and there were some way people could attach money to a message, you could take the view that you will accept e-mail from a known contact or one that has five pennies on it."

Even a charge of as little as a tenth of a penny, he argues, would cost the average spammer $1 million for a mailing—surely enough to deter untargeted junk. The user could choose to void the payment. The problem: micropayments, though technically feasible, have yet to find widespread acceptance.

AIDS

The Race Card

DOES AN HIV VACCINE WORK DIFFERENTLY IN VARIOUS RACES? **BY CAROL EZZELL**

At the end of February, VaxGen—a biotechnology company based in Brisbane, Calif.—announced the long-awaited test results of AIDSVAX, the first AIDS vaccine to have its effectiveness evaluated in large numbers of people. Unfortunately, the bottom line was that the vaccine didn't work. Of the 3,330 people who received AIDSVAX, 5.7 percent had nonetheless become infected with HIV within three years, a rate almost identical to the 5.8 percent seen among 1,679 individuals who received a placebo.

PHILIP JONES GRIFFITHS *Magnum Photos*

GAY MEN constituted most of the AIDSVAX participants. The drug showed no overall protection in whites but offered a hint of efficacy in blacks and Asians.

But intriguingly, the company reported, AIDSVAX appeared to work better among the small numbers of African- and Asian-Americans in the study. Although only 327 blacks, Asians and people of other ethnicities received the vaccine, VaxGen said it protected 67 percent of them (3.7 percent got infected as compared with 9.9 percent of controls). AIDSVAX was particularly effective among African-Americans, preventing 78 percent of the 203 individuals in the study from contracting HIV. (Only two of the 53 Asians became infected, whereas six of the 71 people classified as "other minorities" did.)

Exactly how AIDSVAX might elicit disparate effects among people of various races is unclear. It consists of pieces of gp120, the outer envelope of HIV. Vaccines made of such fragments typically cause the body to make antibodies that latch onto microbes and cause their destruction. But scientists disagree about whether the process necessarily involves so-called tissue-type antigens, which vary among races and whose usual function is to help the body distinguish parts of itself from foreign invaders.

In fact, the racial differences observed by VaxGen could have resulted from any number of reasons, according to Richard A. Kaslow, an AIDS researcher at the University of Alabama at Birmingham who studies why HIV infects some people more readily than others. Because the numbers of blacks and Asians were so small, random factors such as the amount of virus circulating within the sexual partners of the study participants could have had an effect. "Chance could have distorted the results," Kaslow suggests. "But [VaxGen] perhaps has some additional data that we haven't seen yet."

He points out that VaxGen is still analyzing its numbers—it only "broke the code" to learn which clinical trial volunteers had gotten the real vaccine and which the sham vaccine in mid-February—and it has not yet published the results in a scientific journal for other researchers to scrutinize. Nevertheless, he says, it strikes him as "unlikely" that AIDSVAX could have been so selectively effective in two racial groups: no other vaccine has been.

Biostatisticians, including Steven G. Self of the University of Washington, claim that the positive news in blacks and Asians could also have resulted from honest statistical errors in making the adjustments required to analyze such data subsets. In response, VaxGen has issued a statement that its analysis "followed a statistical analysis plan that was agreed on in advance with the U.S. Food and Drug Administration" and that the results "remain accurate as stated, and the analysis continues." The company said it planned to report additional findings at a scientific conference in early April, after this issue of *Scientific American* went to press.

A QUESTION OF SEXUAL PRACTICES

Another variable confounding the new AIDS vaccine results is that most of the African-Americans participating in the study were women whose risk of HIV infection was having sex with men. In contrast, the great majority of the other study volunteers were white gay men. Accordingly, the vaccine's apparent ability to protect blacks and Asians more readily than Caucasians and Hispanics could suggest that it might work best in preventing heterosexual transmission.

ACOUSTICS

Sounding Off

LAWSUITS BLOCK SCIENCE OVER FEARS THAT SONAR HARMS WHALES **BY KRISTA WEST**

The new year didn't start off so well for conservation biologist Peter L. Tyack of the Woods Hole Oceanographic Institution. In January a judge stopped his tests of a new, high-frequency sonar system intended to act as a "whale finder," for fear that the bursts of sound might harm gray whales migrating close by.

The decision is the latest in a rash of court cases in which public concern for marine mammals has stopped acoustic research. Last October a judge halted seismic operations in the Gulf of California after whales became stranded nearby, and in November a court order limited the U.S. Navy's sonar tests, citing multiple suspicious strandings in years past. Yet the recent rulings have nothing to do with any new science; sound has been used to explore the seas for decades. Rather the national media have tuned in, and the subsequent legal activity is putting scientists in a catch-22: the laws need to be improved to protect marine life from harmful acoustic research, but more acoustic research is needed to determine what is harmful to marine life so that the laws can be improved.

Tyack's experience this winter is a perfect example of the circular debate. His project off the coast of California was intended to help marine mammals by giving boats a tool to detect the sea creatures and thereby avoid exposing them to potentially harmful man-made noises. Tyack's whale finder got the legal go-ahead from the National Marine Fisheries Service (NMFS) for testing. Then an attorney representing six environmental groups convinced a San Francisco judge to stop the research. The judge ruled that the NMFS must go back and com-

JOE MOBLEY

WHALE SURVEYS, which spotted this sperm whale in 2002, were done near North Pacific Acoustic Laboratory operations to see if sound affected the mammals.

plete an environmental impact assessment, even though fish finders (to which Tyack's whale finder is similar) do not require such approval and are unregulated.

Joel Reynolds, an attorney with the Natural Resources Defense Council, says Tyack and his team were "hung out to dry" by the NMFS, which did not adequately complete its part of the permit process. And although Reynolds is a staunch defender of the current system, calling U.S. marine laws among the strongest in the world, he says they are not perfect. The permit process can be expensive and slow, and it is not always applied equally to academic research, industry and the military.

One of the earliest tussles between academics and whale defenders involved the Acoustic Thermometry of the Ocean Climate (ATOC) project. In 1995 acoustic sources off the coast of Kauai, Hawaii, and Point Sur, Calif., began transmitting low-frequency sound waves across the North Pacific to measure large-scale changes in ocean temperature. The ATOC, now known as the North Pacific Acoustic Laboratory (NPAL), transmitted sound for several years before stopping in 1999 for the renewal of marine mammal permits; operations resumed in Hawaii last year.

Using aerial surveys to better understand marine life near NPAL operations, researchers counted significantly more marine mammals in 2002, when the sound was on, compared with 2001, when the sound was off. Good ocean conditions and an increase in humpback whale populations probably explain the increase in sightings. NPAL transmissions have not had any obvious effects on marine mammals, remarks NPAL's Peter Worcester of the Scripps Institution of Oceanography. (As for the experiment itself, Worcester is excited about finally obtaining temperature data: "The Pacific north of Hawaii is warming, but between Hawaii and the mainland it's cooling.")

More specific knowledge about how sound affects marine mammals may come this summer, when Tyack will team up with researchers from Columbia University's Lamont-Doherty Earth Observatory to measure the effect of sound on sperm whale behavior. One ship will fire an array of airguns, and the research vesssel *Maurice Ewing* will tag and track the response of the whales.

Perhaps not surprisingly, Tyack's group may find itself in another bind. Operations of the *Maurice Ewing,* long regarded as one of the quietest in the fleet, were stopped last October after two beaked whales were stranded in the Gulf of California near the vessel. The pending legal action against the *Maurice Ewing,* says Maya Tolstoy, a lead researcher at Lamont-Doherty, may threaten work planned for this season.

Krista West, based in Las Cruces, N.M., wrote about Ted Turner's conservation efforts in the August 2002 issue.

THE NAVY AGAINST THE LAW

The U.S. Navy is one of the oldest and loudest producers of sound in the sea and has been testing high-frequency sonar systems designed to detect enemy vessels for decades. In 1994 attorney Joel Reynolds of the Natural Resources Defense Council discovered that the navy was testing sonar without the required sound permits and has been engaged in litigation with it ever since.

Most recently, the navy proposed changing the legal definition of marine mammal harassment to encompass only "significant" changes in behavior. The NRDC is fighting this proposal because, Reynolds says, it could greatly reduce the effectiveness of current laws by making them subjective rather than objective. At the end of 2002 a federal judge restricted navy sonar testing to a relatively small swath of the Pacific, where operations continue today.

ASTRONOMY

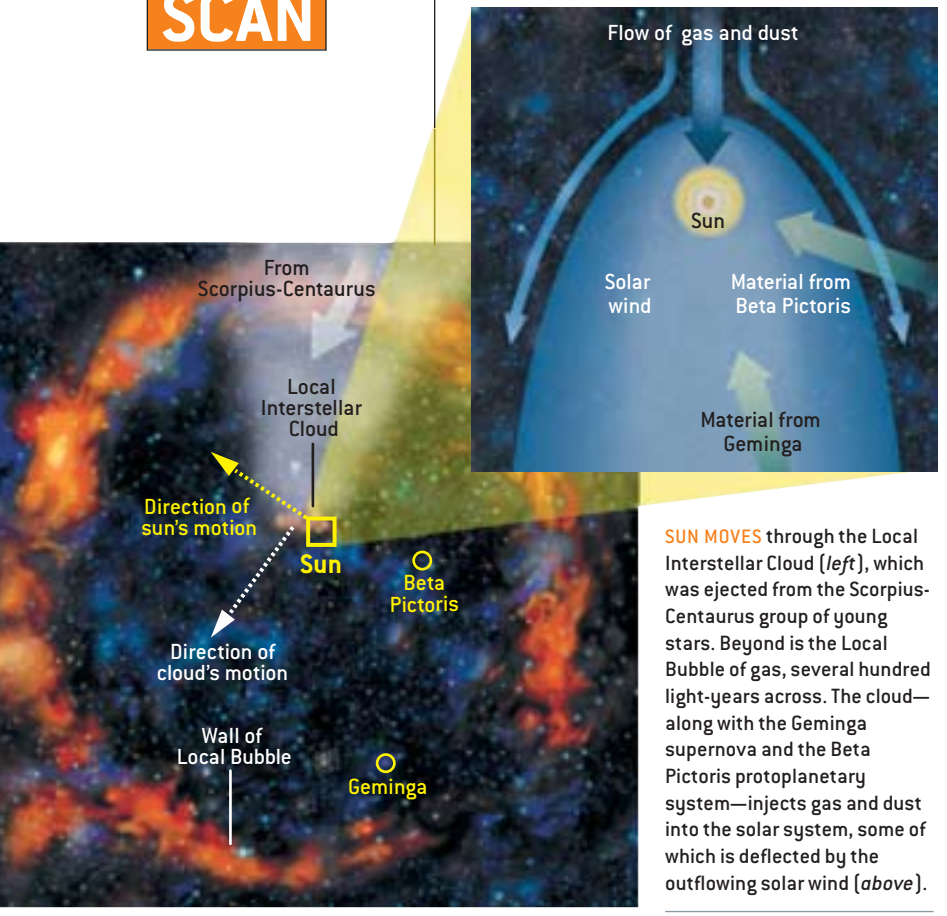
Interstellar Pelting

EXTRASOLAR PLANET AND CLIMATE CLUES FROM ALIEN MATTER

BY GEORGE MUSSER

For quite the longest time, astronomers thought of the galaxy as a kingdom of independent principalities. Each star held sway in its own little area, mostly cut off from all the others. The Milky Way at large determined the grand course of cosmic history, but the sun ran the day-to-day affairs of the solar system. Gradually, though, it has dawned on researchers that the sun's sovereignty is not so inviolable after all. Observations have shown that 98 percent of the gas within the solar system is not of the solar system—it is foreign material that slipped through the sun's Maginot Line. One of every 100 meteoroids entering Earth's atmosphere on an average night is an interstellar intruder.

"When I was an astronomy grad student in Berkeley in the late '60s, interstellar matter was what you observed towards other stars," says Priscilla C. Frisch of the University of Chicago, a pioneer in this subfield of astronomy. "No one dreamed that it was inside of the solar system today." Telescopes have cobbled together a map of our neighborhood;



SUN MOVES through the Local Interstellar Cloud (*left*), which was ejected from the Scorpius-Centaurus group of young stars. Beyond is the Local Bubble of gas, several hundred light-years across. The cloud—along with the Geminga supernova and the Beta Pictoris protoplanetary system—injects gas and dust into the solar system, some of which is deflected by the outflowing solar wind (*above*).

OLD DUST, YOUNG DUST

The word "interstellar" has been applied to two types of material within the solar system. There are the interstellar grains found in meteorites or comets, but these tidbits are presolar—they got swept up during the formation of the sun and planets 4.5 billion years ago and have survived unchanged ever since. They reveal which kinds of stars seeded the solar system. And there is the brand-new stuff, much of it arriving in the headwind that the solar system encounters as it moves through the galaxy. That headwind pours in at 26 kilometers per second from the direction of the constellation Sagittarius. Flecks also arrive from other directions. A recent study at Arecibo Observatory in Puerto Rico attributed some dust to Geminga, a supernova that took place 650,000 years ago about 230 light-years away.

deep space probes have sampled trespassing dust and gas; and radar facilities have tracked interstellar meteors, which distinguish themselves by their unusually high speeds.

Until humanity builds its first starship, these interstellar interlopers will be our only specimens of the rest of the galaxy. They could provide some important ground-truth for theories. For instance, some of the incoming dust could be pieces of other planetary systems. Three years ago W. Jack Baggaley of the University of Canterbury in New Zealand traced a batch of interstellar meteors to Beta Pictoris, a star famous for its disk of dust and planetesimals. The comparatively massive grains that Baggaley considered are barely deflected by radiation pressure or magnetic fields, so they travel in nearly straight lines. Their trajectories point back to where Beta Pictoris was located about 600,000 years ago, implying that the system ejected them—presumably by a gravitational slingshot effect around a planet—at 30 kilometers per second.

Joseph C. Weingartner and Norman Murray of the University of Toronto have now corroborated Baggaley's basic concept. Beta Pictoris and half a dozen other known systems should indeed fling dust our way. But the scientists doubt Baggaley has seen any such dust; the implied ejection speed was much higher than gravitational effects typically manage. Weingartner and Murray suggest putting together a network of radars, monitoring an area the size of Alaska, over which about 20 grains should arrive from each system every year.

Not only do interstellar invaders bring news of distant events, they might change the course of events on Earth. Some astronomers think that the ever changing galactic environment could affect the planet's climate. Right now the sun and its retinue are passing through the Local Interstellar Cloud, but as recently as 10,000 years ago, we found ourselves in the lower-density Local Bubble. Frisch and her colleagues recently pinpointed two higher-density clouds that might engulf us over the coming millennia. Once every 100 million years or so, the solar system wades through one of the galaxy's spiral arms, where the density of stuff is especially high. The higher the external density, the more material will push past the sun's outflowing matter and intrude into the realm of the planets. In extreme cases, the sun's writ does not extend even as far as the outer planets.

Last year Nir J. Shaviv of the Hebrew University of Jerusalem argued that the 100-million-year galactic cycle matches a 100-million-year cycle of broadly higher or lower temperatures on Earth. The connection could be cosmic rays: as more of these energetic particles get through to Earth, they may seed the formation of more low-altitude clouds, which cool the planet. But the evidence is inconclusive, and climatologists are less smitten with the hypothesis than astronomers are.

This past January, NASA launched CHIPSat, dedicated to measuring the Local Bubble. Stardust, a NASA mission to collect samples from Comet Wild-2 next January, has been making chemical analyses of the interstellar dust it bumps into on the way. The European Space Agency is considering Galactic DUNE—a spaceborne "dust telescope." And NASA is pondering Interstellar Probe, which would make a break from the solar system using solar sails. If we cannot keep the rest of the galaxy off our turf, we might as well engage in a little imperialism of our own.

SAMUEL VELASCO

DATA POINTS: ICY SURGE

Researchers have wondered whether floating ice shelves along the Antarctic coast hold back interior glaciers and keep them from the ocean, where they could raise the sea level. They apparently do, according to work by Hernán De Angelis and Pedro Skvarca of the Argentine Antarctic Institute in Buenos Aires. Using aerial and satellite data from February 2000 and September 2001, the two found that after the collapse of the Larsen Ice Shelf in West Antarctica, inland glaciers have surged dramatically toward the coast in recent years.

Advance of the Sjögren and Boydell glaciers: 1.25 kilometers

Advance of the Bombardier and Edgeworth glaciers (net): 1.65 kilometers

Rate of advance of the Sjögren glacier, 1999: 1 meter per day

Rate of advance, 2001: 1.8 to 2.4 meters per day

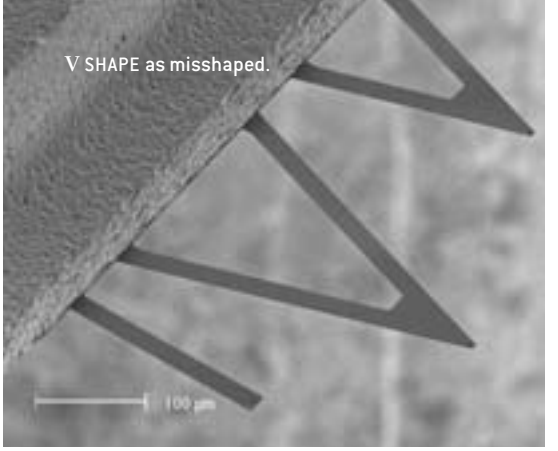
Rise in sea level per year: 2 millimeters

Rise in sea level if the West Antarctic ice sheet collapsed: 5 meters

SOURCES: Science, *March 7, 2003;* Scientific American, *December 2002*

MICROSCOPY

Pulling the Lever



Atomic-force microscopes have been the exquisitely delicate tools of choice for making three-dimensional images of atoms for almost two decades. One mathematician now concludes that its predominant design is fundamentally flawed. To form images, the microscopes rely on probes, as long as a human hair is wide, running over surfaces. In most instruments, the tip is mounted at the end of a V-shaped cantilever. Scientists believed that this chevron shape would resist the swaying that could lower image quality. John E. Sader of the University of Melbourne instead finds that the V shape enhances twisting and inadvertently degrades the performance of the instrument. "This came as a complete surprise, since intuition would dictate the opposite would be true," Sader says. He compares this result with a sheet of metal attacked by pliers: it is easier to bend the sheet at the corners than at the middle. Sader, whose calculations suggest that straight beams are better, reports his findings in the April *Review of Scientific Instruments*. —*Charles Choi*

PSYCHOLOGY

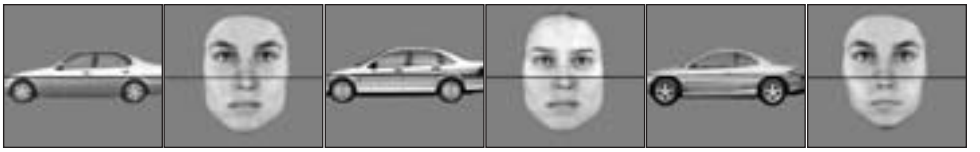
Not So Happy Together

Married people are on average happier than singles, but that extra happiness looks negligible. A 15-year study of 24,000 subjects in Germany found that married folks get a boost in satisfaction shortly after the nuptials, but their levels of happiness drop back to their single days: on an 11-point scale, marrieds rated themselves only 0.1 point happier. People who were most satisfied with their lives react least positively to marriage and, in a surprise, most negatively to divorce or widowhood. The study, in the March *Journal of Personality and Social Psychology,* supports the notion that over time people's sense of well-being reverts to their general level of happiness, no matter what life events have occurred. —*Philip Yam*

PERCEPTION

A Face in the Car Crowd

Car aficionados obsessed with the latest models apparently rely on the same neural circuits as those used to recognize faces. Psychologists tested 40 men—20 car fanciers and 20 car novices—by showing them alternating images of vehicles and faces. Subjects, whose brains were wired with electrodes, looked only at the lower half of the images and had to compare faces and autos with those seen previously in the sequence. The researchers found that novices had to assemble the automobile pieces mentally to identify the model, although they were able to recognize faces "holistically"—that is, all at once. Car lovers, in contrast, perceived the autos holistically, just as they did faces; their use of identical brain areas for car and face recognition also resulted in a perceptual traffic jam: auto enthusiasts had more difficulty recognizing faces than novices did. The work, which challenges the notion that a specialized area of the brain recognizes faces, was published online March 10 by *Nature Neuroscience*. —*Philip Yam*



FACES AND CARS had to be remembered during a sequence of images; subjects attended to the lower halves.

ILLUSTRATION BY MATT COLLINS; SIMON CRAWFORD *University of Melbourne (top)*; COURTESY OF *NATURE NEUROSCIENCE (bottom)*

LIGHT THERAPY

Seeing Red

RED LIGHT, GO: Light-emitting diodes stimulate plant growth in space and may also help repair tissue, including methanol-damaged retinas.

Someday red lights could stop more than cars—they could halt and even reverse blinding eye damage. To encourage plant growth in space, NASA designed a red light-emitting diode, the emission of which packs 10 times the energy of the sun's at the same wavelength. Astronauts found that the red LED, about the size of a pack of cards, also helped cuts to heal faster. Evidently the light stimulates mitochondria, the cell's powerhouses, through a still unknown mechanism. Neurotoxicologist Janis T. Eells of the Medical College of Wisconsin was studying methanol poisoning, believed to induce blindness by inhibiting mitochondria, which are especially active in the eyes. Despite Eells's initial skepticism, she observed that three 144-second bouts of red LED light over a 50-hour span given to methanol-poisoned rats enabled them to recover 60 to 70 percent of retinal function. Eells and her colleagues, whose report was published online March 7 by the *Proceedings of the National Academy of Sciences USA*, plan to see if the light can fight glaucoma. *—Charles Choi*

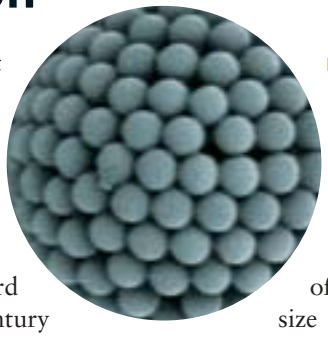
STEM CELLS

Insulin from Bone Marrow

Insulin-producing beta cells could be harvested from the stem cell–rich bone marrow, according to a study in mice by researchers at New York University. The team created male mice with marrow cells that made a fluorescent protein in the presence of an active insulin gene. The researchers removed the marrow cells and transplanted them into female mice whose marrow cells had been destroyed. After four to six weeks, some of the fluorescent protein–making cells had migrated to the pancreas, where they joined with existing beta cells and made insulin. Only 1.7 to 3 percent of the pancreatic beta cells actually came from the bone marrow, and scientists do not know which stem cells in the marrow actually produced them. But the strategy offers fresh hope for diabetics for a comparatively convenient source of the insulin-making cells. The study appears in the March *Journal of Clinical Investigation*. *—Philip Yam*

MATHEMATICAL PHYSICS

Packing 'Em On



POLYSTYRENE BEADS only microns wide coat a water droplet.

Particles crowded onto a flat surface will settle into a pattern resembling racked pool balls, but researchers have puzzled for years over the structure of those same particles wrapped around a sphere. Swiss mathematician Leonhard Euler proved in the 18th century that adjacent triangles wrapped onto a sphere must have at least 12 defects, or sites that have five neighbors instead of six. (That's why a soccer ball has 12 pentagons amid all its hexagons.) Now physicists have predicted and confirmed that spheres made of several hundred or more particles relieve strain by forcing additional particles to have five or seven neighbors, thereby creating defects beyond the original 12 that Euler stipulated. These neighbor defects are arranged in lines, or "scars," the lengths of which are proportional to the size of the sphere. The scientists used a microscope to view and trace images of micron-size polystyrene beads coating tiny water droplets. The scar lengths should be independent of the type of particles, the researchers report in the March 14 *Science*, so the result could help in designing self-assembling materials and in understanding biological protein shells and defects in fullerene molecules. *—JR Minkel*

NASA (top); DAVID A. WEITZ Harvard University (bottom)

BRIEF POINTS

■ **A drug called TNX-901 sops up immunoglobulin E antibodies that are triggered by peanuts, thereby tempering the sometimes deadly reaction in those with peanut allergies.**

New England Journal of Medicine, March 13, 2003

■ **Signed off: Pioneer 10, which was launched 30 years ago and was the first spacecraft to visit Jupiter and fly beyond Pluto, is now apparently too weak to signal home. Its last transmission was January 22.**

NASA statement, February 25, 2003

■ **Researchers in March used the Arecibo radio telescope to reobserve up to 150 of the most interesting objects identified by the SETI@home project, meant to find signals that might have originated from an extraterrestrial intelligence.**

Planetary Society; see http://planetary.org/stellarcountdown/

■ **The Bush administration announced plans to build the first zero-emissions power plant. The $1-billion, 10-year project will try to construct a coal-based plant; carbon emissions would be captured and sequestered.**

White House announcement, February 27, 2003

BY THE NUMBERS

Reducing Crime

REHABILITATION IS MAKING A COMEBACK BY RODGER DOYLE

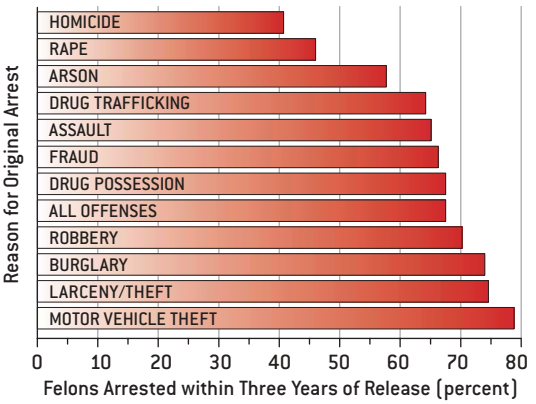
Two thirds of all felons released from state prisons are rearrested within three years, which helps to explain why U.S. imprisonment rates are so high. Another reason is the increased length of sentences, the result of "tough on crime" sentencing laws that became popular in the 1970s.

Before 1970, rehabilitation was the dominant philosophy among American criminologists. The change to a harsher regime was signaled by sociologist Robert Martinson of the City University of New York, who, in an influential article published in 1974, concluded that "with few exceptions, the rehabilitative efforts that have been reported so far have had no appreciable effect on recidivism." The press expressed this idea under headlines such as "NOTHING WORKS." In light of rehabilitation's supposed failure, James Q. Wilson of Harvard University and other neoconservatives urged longer prison sentences and, occasionally, capital punishment to fight crime. This view soon became the accepted wisdom—despite Martinson's repudiation in 1979 of his earlier conclusion. In 1985 Alfred S. Regnery, the administrator of the Office of Juvenile Justice and Delinquency Prevention, claimed that "rehabilitation … has failed miserably," and in 1987 Attorney General Edwin Meese referred to the "substantially discredited theory of rehabilitation." In 1989 the Supreme Court upheld federal sentencing guidelines that removed rehabilitation from serious consideration.

About 10 years ago opinion began to shift again, largely because a new research technique, meta-analysis, convincingly demonstrated that rehabilitation *does* work. The method combines the results of many studies, thereby averaging out extraneous and idiosyncratic factors. Meta-analyses of nearly 2,000 studies encompassing a variety of approaches aimed at reducing recidivism have revealed that the average effect of rehabilitation is positive, though fairly modest, in part because of the inclusion of a number of therapies that did not work. Certain behavioral modification programs for violent offenders and for medium-risk sex offenders have been particularly effective, achieving reductions in recidivism of 50 percent or more as compared with controls. Programs targeting juvenile offenders—including mentoring, skills instruction and, for teenage mothers, intensive home visiting to reduce child abuse—attained high success rates in preventing crime.

Research studies measure the effectiveness of therapies in an artificial setting, but in real-life situations the treatments are often less convincing. Nevertheless, results such as these, even if diluted by half, would still make a substantial dent in the U.S. crime rate. Rehabilitation therapy is expensive in the short term; still, it is far cheaper than the criminal justice system, which incurred direct costs of $147 billion in 1999 and has been growing by more than 5 percent annually in recent years.

One of the leading researchers on criminal behavior, James McGuire of the University of Liverpool in England, notes that, in general, punishment is not effective and may actually increase crime rates. Boot camps, three-strikes laws, so-called scared-straight programs and the death penalty are proving ineffective in preventing recidivism. Public notification of released sex offenders in the community—"Megan's Law" measures—has never been adequately tested for efficacy.



SOURCE: Bureau of Justice Statistics. Chart is based on data from 15 states, representing two thirds of all prisoners released in 1994 in the U.S. The original arrest does not necessarily match the reason for the rearrest.

Rodger Doyle can be reached at rdoyle2@adelphia.net

RODGER DOYLE

BACK TO THE BIG HOUSE

Percent of prisoners released in 1994 who were rearrested within three years:

Male	68
Female	58
White	63
Black	73
Hispanic	65
Ages	
14–17	82
18–24	75
25–29	71
30–34	69
35–39	66
40–44	58
45 and older	45

FURTHER READING

What Works: Reducing Reoffending. Edited by James McGuire. John Wiley & Sons, 1995.

Evidence-Based Programming Today. James McGuire. Paper delivered at the International Community Corrections Association annual conference, Boston, 2002.

Offender Rehabilitation and Treatment. Edited by James McGuire. John Wiley & Sons, 2002.

Recidivism of Prisoners Released in 1994. Patrick A. Langan and David J. Levin. Bureau of Justice Statistics, June 2002.

X-ray Proofing

To save himself, a physician enters the rag trade By STEVEN ASHLEY

If necessity is the mother of invention, then self-preservation is surely one of the family matriarchs. A case in point is the brainchild of Ronald F. DeMeo, a Florida-based anesthesiologist who regularly takes x-rays of his patients when treating chronic back and neck pain.

Concerned about the cumulative damage x-rays might be wreaking on his own body, DeMeo began searching years ago for a better way to protect himself—beyond the standard practice of donning a heavy lead medical vest or apron, gloves, a thyroid shield or lead-glass goggles or of having to leave the room frequently during x-ray imaging to keep a safe distance away from the radiation source.

NEW HAZMAT COUTURE is both radiation-resistant and comfortable to wear over extended periods.

After eight years of collaborative research, the physician-entrepreneur has developed a unique polymer composite–based fabric he calls Demron. It not only blocks x-rays and nuclear emissions (gamma rays, alpha particles and beta particles) as effectively as current standard lead-based apparel does, it is also significantly more flexible and wearable. Widely used lightweight plastic protective outerwear does not impede the passage of x-rays and gamma rays at all.

In addition, the new fabric seems to be impermeable to deadly chemical and biological warfare agents, so it can be used in jumpsuits for hazardous-materials emergency workers and "first responders" to disaster scenes. Experts at the U.S. Department of Defense are currently evaluating Demron's effectiveness when used in nuclear-biological-chemical suits against common chemical warfare agents such as mustard gas, VX nerve gas and sarin. A typical Demron full-body hazmat suit costs about $600. The new material could also be fashioned into radiation-proof tents, linings for aircraft and spacecraft, covers for sensitive equipment, and medical shielding garments.

Anxious about the steady rise of his own total radiation dosage, DeMeo sought to reduce exposure for himself and his staff. "I entered the radiation-shielding business for reasons of self-preservation—to allow me to live longer," he recalls.

For those who come into contact regularly with x-rays or nuclear material, limiting one's dosage is difficult. "Most practitioners, for example, work in different hospital facilities, each of which use different dosimeter badge sets," DeMeo notes. "Hardly anybody does the math and adds up all the separately measured doses." Complicating the situation is an abiding problem: regulations forbid medical and radiation workers from continuing in their jobs if they have exceeded safe cumulative dosage levels. "People often don't want to know what their total dose is because they don't want to be forced to stop working," he says.

RADIATION SHIELD TECHNOLOGIES

And few want to work wearing awkward lead aprons and vests (costing between $85 and $600), which are typically constructed of weighty, cumbrous sheets of powdered lead in a polymeric matrix.

Although the radiation-safety experts DeMeo consulted were skeptical, he began funding research projects in which he hired chemists and materials experts to search for lightweight, flexible substances that can stop x-rays. Eventually the physician formed a company in Miami, Radiation Shield Technologies (RST), to develop and market his products. Now the firm's chief executive officer, DeMeo continues his medical practice as well.

At first the small research group studied metal shielding, but that turned out to be just one of numerous dead ends. Lead is toxic, heavy and bulky, so that was out. Says DeMeo: "Copper and aluminum showed some [shielding] response, but nothing overly useful. Later we worked on embedding metal particles in fabric and obtained a few patents in that area. Then we got involved with trying to find polymers that attenuate radiation."

After considerable fruitless effort, the RST team came up with a polymer composite of polyurethane and polyvinylchloride that incorporates a variety of organic and inorganic salt particles that block radiation. Constituents of these salts have high atomic numbers (the number of protons in an atom of a particular element), so they tend to arrest radiation more effectively. "Our material looks and behaves like a heavy, dense rubber," DeMeo says.

Demron works in two ways, depending on the type of radiation. When x-rays or gamma rays meet these dispersed salt particles, DeMeo explains, they are either absorbed (via the photoelectric effect) and their energy dissipated through the generation of heat, or they are scattered at an altered energy level (via the Compton effect) and then absorbed or deflected by surrounding particles. This cascade of absorption and scattering stops harmful radiation from penetrating to body tissues. When alpha and beta particles strike Demron, intervening electrons in the salt atoms deflect and slow them down, whereupon they are absorbed into the material.

Because x-ray machines produce a spectrum of photons and common radionuclides emit particles with a range of energies, the radiation-blocking agents in the Demron fabric must be tailored to these various energies, a technique called spectral hardening. "Each attenuation material we've included has an energy level it's good at absorbing or scattering," DeMeo says. "It's something like installing soundproofing. A one-inch-thick panel of wood stops certain sound frequencies, but a similar-size sandwich comprising a quarter-inch-thick piece and a three-quarter-inch piece stops more frequencies."

The polymer composite can be made in two forms: as thin film sheets or as injection-molded shapes. RST's initial Demron offering is produced by laminating the film between two layers of fabric—one woven, the other nonwoven. The resulting material is about 0.43 millimeter thick and has a density of about 0.7 gram per square inch.

Though nearly as dense as the material in lead-based shielding vestments, Demron readily bends, creases and folds. The thin, compliant fabric has proved itself against both x-rays and nuclear emissions in tests at Lawrence Livermore National Laboratory, the Neely Nuclear Research Center at the Georgia Institute of Technology, and the department of radiology at Columbia University's College of Physicians and Surgeons. It is not yet clear, however, whether Demron degrades when subjected to extended radiation exposure. The material is impermeable to air and fluids and can withstand at least eight hours of exposure to corrosive chlorine and ammonia gas.

Toxic-site cleanup crews wore Demron hazmat suits for hours, even doing calisthenics in them.

Because it allows radiant heat loss, Demron feels cool to the touch and releases internal heat to the surrounding air. Therefore, "it can be used to cover 100 percent of your body surface area," DeMeo says. Last summer, toxic-site cleanup crews tested prototype Demron suits to see whether they would be comfortable when worn for long periods. "The ergonomic evaluation went well," he reports. "The crews could wear it for hours at a time, even do calisthenics in it. Current nuclear-biological-chemical suits are walking saunas. Troops wearing them could die of heatstroke in the desert."

In October 2002 RST contracted with a clothing manufacturer to make jumpsuits for first responders and cleanup workers. DeMeo is next considering producing injection-molded gloves as well as customized protective covers for equipment.

Orders for Demron hazmat outfits are backing up, he says: "We've had a fairly tremendous response to our product introduction." Thus far Demron has gone a long way toward proving that a thin, highly flexible and wearable radiation shield is not a technical impossibility after all. SA

Make Your Own Rules

Patents let private parties take the law into their own hands By GARY STIX

Biotechnology critics Jeremy Rifkin and Stewart Newman filed for a patent in 1997 for a method of combining various types of embryo cells to produce chimeras, creatures that would be part human, part animal. Rifkin and Newman had no intention of becoming biomedical entrepreneurs. Rather they wished to take advantage of the essential nature of patent law to press their case against what they consider an objectionable form of bioengineering. A patent permits someone to exclude others from making, using or selling an invention. If their application were granted, Rifkin and Newman could use their patent rights to become private regulators of chimera technology, which could be of interest to researchers engaged in creating replacement organs for human transplants. The two men would, in effect, have the power to ban chimeras for the term of the patent, the better part of 20 years.

The Rifkin-Newman application has been rejected several times already by the U.S. Patent and Trademark Office, although an altered application has been resubmitted and is still pending. Georgetown University law professor John R. Thomas sees the case as a demonstration of how the patent system is being commandeered by private individuals who then go on to make their own laws, free from the traditional safeguards that prevent the government from abusing its power. This trend emerges from the willingness of the U.S. patent office to approve what Thomas calls "post-industrial" patents that cover everything from methods of doing business to human behaviors.

A political party might claim that a soft-money campaign technique infringes its patent, or a human-rights organization could prohibit use of its patent on a racial-profiling process. The possibility of an antiabortion group obtaining a patent and using it to restrict access to, say, an abortion-inducing drug is a real one, Thomas notes. Patents can even hijack federal tax law for private ends. Signature Financial Group in Boston received a patent on a computerized method that allows certain partnerships to allocate profits, losses and expenses to individual mutual funds invested in such partnerships on a daily basis. By making allocations each day, the partnerships can obtain favorable tax treatment. Some of the language in the patent, Thomas says, closely parallels the tax code—what's new is merely that the process is carried out by a computer. "Congress presumably intends its laws to apply to all citizens," Thomas remarks. "Allowing one private entity to regulate access to a tax break is strikingly poor intellectual-property policy."

Using the patent system as a private regulatory vehicle circumvents the checks and balances to which government-made law is subjected. Constitutional guarantees of individual rights can be invoked only against the government, not against a plaintiff suing for patent infringement. Thomas gives the example of patents that have been granted that regulate the content of speech, including ones for making sales pitches or delivering advertising over networks. Government control of expression is strictly circumscribed. "Yet all indications from the courts are that privately held patents offer their owners the ability to suppress or punish speech without reference to these limitations," Thomas wrote last year in the *Houston Law Review*.

He suggests that a set of little-known Supreme Court decisions—which constitute what is called the nondelegation doctrine—might be invoked by federal courts to curb unwarranted attempts at private lawmaking. The Supreme Court decided in a number of cases before World War II that the government should not confer its lawmaking authority on private individuals or organizations. The courts' selective use of the nondelegation doctrine, Thomas contends, could provide a "backdoor to the Bill of Rights" if the ambitions of patent holders overstep the bounds that were intended by the framers of the Constitution. SA

JENNIFER KANE

Show Me the Body

Purported sightings of Bigfoot, Nessie and Ogopogo fire our imaginations. But anecdotes alone do not make a science By MICHAEL SHERMER

The world lost the creators of two of its most celebrated biohoaxes recently: Douglas Herrick, father of the risibly ridiculous jackalope (half jackrabbit, half antelope), and Ray L. Wallace, paternal guardian of the less absurd Bigfoot.

The jackalope enjoins laughter in response to such peripheral hokum as hunting licenses sold only to those whose IQs range between 50 and 72, bottles of the rare but rich jackalope milk, and additional evolutionary hybrids such as the jackapanda. Bigfoot, on the other hand, while occasionally eliciting an acerbic snicker, enjoys greater plausibility for a simple evolutionary reason: large hirsute apes currently roam the forests of Africa, and at least one species of a giant ape—*Gigantopithecus*—flourished some hundreds of thousands of years ago alongside our ancestors.

If such cryptids still survived in the hinterlands of North America and Asia, surely by now one would have turned up.

Is it possible that a real Bigfoot lives despite the posthumous confession by the Wallace family that it was just a practical joke? Certainly. After all, although Bigfoot proponents do not dispute the Wallace hoax, they correctly note that tales of the giant Yeti living in the Himalayas and Native American lore about Sasquatch wandering around the Pacific Northwest emerged long before Wallace pulled his prank in 1958.

In point of fact, throughout much of the 20th century it was entirely reasonable to speculate about and search for Bigfoot, as it was for the creatures of Loch Ness, Lake Champlain and Lake Okanagan (Scotland's Nessie, the northeastern U.S.'s Champ and British Columbia's Ogopogo, respectively). Science traffics in the soluble, so for a time these other chimeras warranted our limited exploratory resources. Why don't they now?

The study of animals whose existence has yet to be proved is known as cryptozoology, a term coined in the late 1950s by Belgian zoologist Bernard Heuvelmans. Cryptids, or "hidden animals," begin life as blurry photographs, grainy videos and countless stories about strange things that go bump in the night. Cryptids come in many forms, including the aforementioned giant pongid and lake monsters, as well as sea serpents, giant octopuses, snakes, birds and even living dinosaurs.

The reason cryptids merit our attention is that enough successful discoveries have been made by scientists based on local anecdotes and folklore that we cannot dismiss all claims a priori. The most famous examples include the gorilla in 1847 (and the mountain gorilla in 1902), the giant panda in 1869, the okapi (a short-necked relative of the giraffe) in 1901, the Komodo dragon in 1912, the bonobo (or pygmy chimpanzee) in 1929, the megamouth shark in 1976 and the giant gecko in 1984. Cryptozoologists are especially proud of the catch in 1938 of a coelacanth, an archaic-looking species of fish that had been thought to have gone extinct in the Cretaceous.

Although discoveries of previously unrecorded species of bugs and bacteria are routinely published in the annals of biology, these instances are startling because of their recency, size, and similarity to cryptid cousins Bigfoot, Nessie, et al. They also have in common—a body! In order to name a new species, one must have a type specimen—a holotype—from which a detailed description can be made, photographs taken, models cast and a professional scientific analysis prepared.

If such cryptids still survived in the hinterlands of North America and Asia, surely by now one would have turned up. So far all we have are the accounts. Anecdotes are a good place to begin an investigation—which by themselves cannot verify a new species. In fact, in the words of social scientist Frank J. Sulloway of the University of California at Berkeley—words that should be elevated to a maxim: "Anecdotes do not make a science. Ten anecdotes are no better than one, and a hundred anecdotes are no better than ten."

I employ Sulloway's maxim every time I encounter Bigfoot hunters and Nessie seekers. Their tales make for gripping narratives, but they do not make sound science. A century has been spent searching for these chimerical creatures. Until a body is produced, skepticism is the appropriate response. SA

Michael Shermer is publisher of Skeptic *magazine (www.skeptic.com) and general editor of* The Skeptic Encyclopedia of Pseudoscience.

BRAD HINES

Wired Superstrings

His networked computer became the equivalent of a Western Union for physicists. Now Paul Ginsparg watches how his idea is changing the way science is communicated By GARY STIX

Cornell University physicist N. David Mermin remembers a student in the late 1970s who would occasionally attend his advanced graduate class on how a branch of topology, called homotopy theory, could be applied in condensed-matter physics. The first-year student would show up every two weeks or so, sit for 10 minutes and then, having ascertained that the class still wasn't covering material that he didn't already know, quietly pick up and leave. After a while, the drop-in stopped appearing at all, but he would sometimes come around to Mermin's office to give the professor advice. "I learned a lot from him," Mermin recounts.

That same independent streak manifested itself 13 years later when the former student, Paul Ginsparg, took a few hours to program a NeXT computer at Los Alamos National Laboratory. The program directed the computer to accept prepublication copies of physics papers automatically and to send out e-mail abstracts of the papers. The full text of the preprint could then be retrieved by querying the computer. Within weeks after the server (then called xxx.lanl.gov) became active in 1991, communication within the high-energy-physics community underwent a transformation. The preprints, which had been available to only an elite few, could now be picked over by anyone instantaneously, whether in Cambridge, Kraków or Calcutta.

The server radically democratized some of the most esoteric pursuits in contemporary science and changed lives—scientists in eastern Europe, the Middle East, South Asia and Latin America suddenly became contributors to or critics of the latest paper on "exact black-string solutions in three dimensions." A self-taught Czech string theorist even won a U.S. graduate scholarship after posting several papers. The importance of Ginsparg's achievement was recognized last year when the researcher won a $500,000 MacArthur fellowship.

Physics, computers and communications have constituted parallel themes throughout Ginsparg's life. The son of a mechanical engineer, Ginsparg built and operated ham radios as a youth in Syosset, Long Island, and later became a Harvard classmate of future Microsoft magnates Bill Gates and Steve Ballmer. His graduate thesis at Cornell dealt in part with the incorporation of fermions, a type of subatomic particle, into lattice theory, a computational means of attacking difficult challenges in high-energy physics. Moving on to a career as a fellow and later a junior professor at Harvard, Ginsparg often found himself enlisted to concoct hastily fashioned software programs that would solve, say, a problem in superstring theory—in which all the fundamental forces, including gravity, are explained in terms of vibrating strings. "The average physicist wasn't into doing this—it was just so alien to them," the 47-year-old

FOREST McMULLIN

PAUL GINSPARG: THE ACTIVE ARCHIVIST

- **In 2002 arXiv.org received 36,000 submissions of scientific papers.**
- **The archive boasts 60,000 registered contributors and is growing by 1,200 per month; there were more than 20 million full-text downloads in 2002.**

Ginsparg muses, his shoeless feet propped on a chair in his office at the Information Sciences building at Cornell.

Nobelist Sheldon L. Glashow, a father figure of sorts, tapped Ginsparg to co-author a scathing 1986 critique in *Physics Today* of the nascent field of superstrings, the aspirations of which, at least for a time, became horribly overblown. The excitement was such, Ginsparg recalls, that superstring practitioners thought they would be able to derive a theory of everything in six months. "A naive comparison suggests that to calculate the electron mass from superstrings would be a trillion times more difficult than to explain human behavior in terms of atomic physics," Glashow and Ginsparg wrote. Before helping to pen the paper, entitled "Desperately Seeking Superstrings," Ginsparg recalls telling Glashow that he was working on superstrings but that he was capable of stopping at any time. "You sound like a smoker," Glashow replied.

Ginsparg was accomplished enough as a physicist to have a subatomic particle, the Ginsparg-Wilson fermion, named after him and his thesis adviser, 1982 Nobelist Kenneth G. Wilson. But he had not, in Harvard's view, initiated a significant new field of research, a prerequisite for tenure. The university's denial was a disappointment, but based on what happened later, he says, "it was the best thing that could possibly have happened."

In 1990, with the fall of the Berlin Wall, Los Alamos National Laboratory began embracing new missions. It extended an open invitation to Ginsparg, and he accepted. It was a logical place for a theoretical physicist with a passion for mountain climbing and cycling. The preprint server came about serendipitously, as Ginsparg's spur-of-the-moment reaction to a colleague's complaint about his electronic mailbox not accepting new messages when it became overloaded with e-mail transmittals of preprints. When the e-print archive was born, in 1991, it was intended to hold papers for only three months. The project became such an instant success that it was turned into a permanent archive—a meticulous record of everything that has occurred in high-energy physics since then. The pace of interaction in the high-energy community—and other disciplines that were added later, such as astrophysics and condensed-matter studies—quickened appreciably. "String theory made unprecedented progress over the last 10 years perhaps due more to Ginsparg than any other individual," says Harvard physicist Andrew Strominger.

Early on, the physics publishing establishment did not know what to make of the archive. "I felt like a visitor from the future showing 19th-century mathematicians the power of the pocket calculator," Ginsparg remarks. At the time, the American Physical Society (APS) had yet to consolidate plans for bringing its journals online, and officials worried about copyright issues and a loss in subscribers. But after Martin Blume took over as editor in chief of the APS six years ago, the tension subsided. Blume paid a visit to Ginsparg at Los Alamos to, as he put it, "make peace." He scrapped APS's own preprint archive and changed copyright terms to allow posting of society-published articles in the archive.

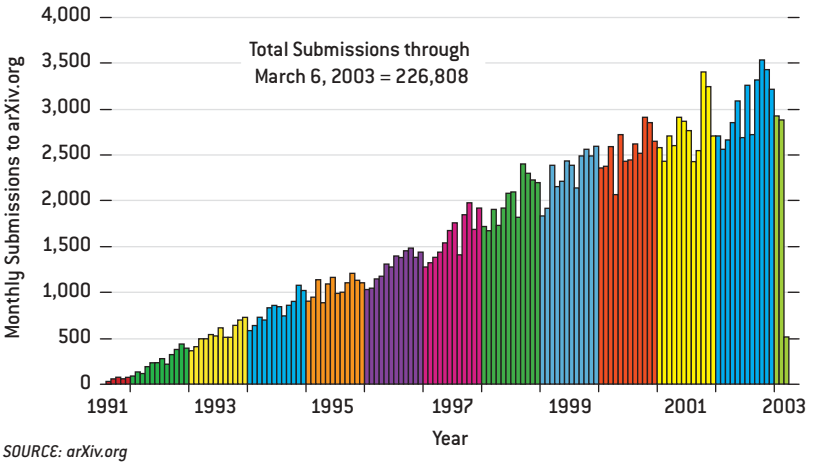
Ginsparg's ideas are now actively solicited by the APS and other organizations, such as PubMed Central, a free archive of life sciences journals. One part of his initial vision that remains unfulfilled is to do away entirely with physics print journals by instituting online peer review of papers submitted to the server. Ginsparg proposes a two-tiered system: all papers submitted to the archive would get cursory scrutiny, but only the most valued findings would go on to receive full peer review.

The biology community, which has been slow to adopt electronic preprints, may actually pioneer the next wave of online publishing. The Public Library of Science, a nonprofit based in San Francisco, will create free access to peer-reviewed electronic biology and medicine journals as an alternative to subscription-based science publications.

The originally welcoming atmosphere at Los Alamos mutated as the 1990s wore on. Nonweapons projects started getting short shrift, and the Wen Ho Lee scandal fostered a paranoia inimical to the open exchange of ideas that the archive embodies. Disenchanted, Ginsparg accepted a joint professorship in information sciences and physics at Cornell in 2001. He brought with him the archive (now called arXiv.org), delegating its maintenance to the university library.

The move to Cornell has afforded Ginsparg an opportunity to return to physics. He is also exploring techniques to facilitate the navigation of the archive for relevant papers but has given up day-to-day responsibility for the repository with a sense of relief: "I think I have overstayed the 15 minutes that Andy Warhol allotted me by a decade." The combination of the MacArthur grant and tenure should permit him to do what he loves best: calculations and problem solving that no one else has touched, whether in physics or computer science. SA

CONTRIBUTIONS TO GINSPARG'S ARCHIVE



SOURCE: arXiv.org

JOHNNY JOHNSON

By Max Tegmark

Parallel Universes

Not just a staple of science fiction, other universes are a direct implication of cosmological observations

Is there a copy of you

reading this article? A person who is not you but who lives on a planet called Earth, with misty mountains, fertile fields and sprawling cities, in a solar system with eight other planets? The life of this person has been identical to yours in every respect. But perhaps he or she now decides to put down this article without finishing it, while you read on.

The idea of such an alter ego seems strange and implausible, but it looks as if we will just have to live with it, because it is supported by astronomical observations. The simplest and most popular cosmological model today predicts that you have a twin in a galaxy about 10 to the 10^{28} meters from here. This distance is so large that it is beyond astronomical, but that does not make your doppelgänger any less real. The estimate is derived from elementary probability and does not even assume speculative modern physics, merely that space is infinite (or at least sufficiently large) in size and almost uniformly filled with matter, as observations indicate. In infinite space, even the most unlikely events must take place somewhere. There are infinitely many other inhabited planets, including not just one but infinitely many that have people with the same appearance, name and memories as you, who play out every possible permutation of your life choices.

You will probably never see your other selves. The farthest you can observe is the distance that light has been able to travel during the 14 billion years since the big bang expansion began. The most distant visible objects are now about 4×10^{26} meters away—a distance that defines our observable universe, also called our Hubble volume, our horizon volume or simply our universe. Likewise, the universes of your other selves are spheres of the same size centered on their planets. They are the most straightforward example of parallel universes. Each universe is merely a small part of a larger "multiverse."

By this very definition of "universe," one might expect the notion of a multiverse to be forever in the domain of metaphysics. Yet the borderline between physics and metaphysics is defined by whether a theory is experimentally testable, not by whether it is weird or involves unobservable entities. The frontiers of physics have gradually expanded to incorporate ever more abstract (and once metaphysical) concepts such as a round Earth, invisible electromagnetic fields, time slowdown at high speeds, quantum superpositions, curved space, and black holes. Over the past several years the concept of a multiverse has joined this list. It is grounded in well-tested theories such as relativity and quantum mechanics, and it fulfills both of the basic criteria

ALFRED T. KAMAJIAN

of an empirical science: it makes predictions, and it can be falsified. Scientists have discussed as many as four distinct types of parallel universes. The key question is not whether the multiverse exists but rather how many levels it has.

Level I: Beyond Our Cosmic Horizon

THE PARALLEL UNIVERSES of your alter egos constitute the Level I multiverse. It is the least controversial type. We all accept the existence of things that we cannot see but could see if we moved to a different vantage point or merely waited, like people watching for ships to come over the horizon. Objects beyond the cosmic horizon have a similar status. The observable universe grows by a light-year every year as light from farther away has time to reach us. An infinity lies out there, waiting to be seen. You will probably die long before your alter egos come into view, but in principle, and if cosmic expansion cooperates, your descendants could observe them through a sufficiently powerful telescope.

If anything, the Level I multiverse sounds trivially obvious. How could space *not* be infinite? Is there a sign somewhere saying "Space Ends Here—Mind the Gap"? If so, what lies beyond it? In fact, Einstein's theory of gravity calls this intuition into question. Space could be finite if it has a convex curvature or an unusual topology (that is, interconnectedness). A spherical, doughnut-shaped or pretzel-shaped universe would have a limited volume and no edges. The cosmic microwave background radiation allows sensitive tests of such scenarios [see "Is Space Finite?" by Jean-Pierre Luminet, Glenn D. Starkman and Jeffrey R. Weeks; SCIENTIFIC AMERICAN, April 1999]. So far, however, the evidence is against them. Infinite models fit the data, and strong limits have been placed on the alternatives.

Another possibility is that space is infinite but matter is confined to a finite region around us—the historically popular "island universe" model. In a variant on this model, matter thins out on large scales in a fractal pattern. In both cases, almost all universes in the Level I multiverse would be empty and dead. But recent observations of the three-dimensional galaxy distribution and the microwave background have shown that the arrangement of matter gives way to dull uniformity on large scales, with no coherent structures larger than about 10^{24} meters. Assuming that this pattern continues, space beyond our observable universe teems with galaxies, stars and planets.

Observers living in Level I parallel universes experience the same laws of physics as we do but with different initial conditions. According to current theories, processes early in the big bang spread matter around with a degree of randomness, generating all possible arrangements with nonzero probability. Cosmologists assume that our universe, with an almost uniform distribution of matter and initial density fluctuations of one part in 100,000, is a fairly typical one (at least among those that contain observers). That assumption underlies the estimate that your closest identical copy is 10 to the 10^{28} meters away. About 10 to the 10^{92} meters away, there should be a sphere of radius 100 light-years identical to the one centered here, so all perceptions that we have during the next century will be identical to those of our counterparts over there. About 10 to the 10^{118} meters away should be an entire Hubble volume identical to ours.

These are extremely conservative estimates, derived simply by counting all possible quantum states that a Hubble volume can have if it is no hotter than 10^8 kelvins. One way to do the calculation is to ask how many protons could be packed into a Hubble volume at that temperature. The answer is 10^{118} protons. Each of those particles may or may not, in fact, be present, which makes for 2 to the 10^{118} possible arrangements of protons. A box containing that many Hubble volumes exhausts all the possibilities. If you round off the numbers, such a box is about 10 to the 10^{118} meters across. Beyond that box, universes—including ours—must repeat. Roughly the same number could be derived by using thermodynamic or quantum-gravitational estimates of the total information content of the universe.

Your nearest doppelgänger is most likely to be much closer than these numbers suggest, given the processes of planet formation and biological evolution that tip the odds in your favor. Astronomers suspect that our Hubble volume has at least 10^{20} habitable planets; some might well look like Earth.

The Level I multiverse framework is used routinely to evaluate theories in modern cosmology, although this procedure is rarely spelled out explicitly. For instance, consider how cosmologists used the microwave background to rule out a finite spherical geometry. Hot and cold spots in microwave background maps have a characteristic size that depends on the curvature of space, and the observed spots appear too small to be consistent with a spherical shape. But it is important to be statistically rigorous. The average spot size varies randomly from one Hubble volume to another, so it is possible that our universe is fooling us—it could be spherical but happen to have abnormally small spots. When cosmologists say they have ruled out the spherical model with 99.9 percent confidence, they really mean that if this model were true, fewer than one in 1,000 Hubble volumes would show spots as small as those we observe.

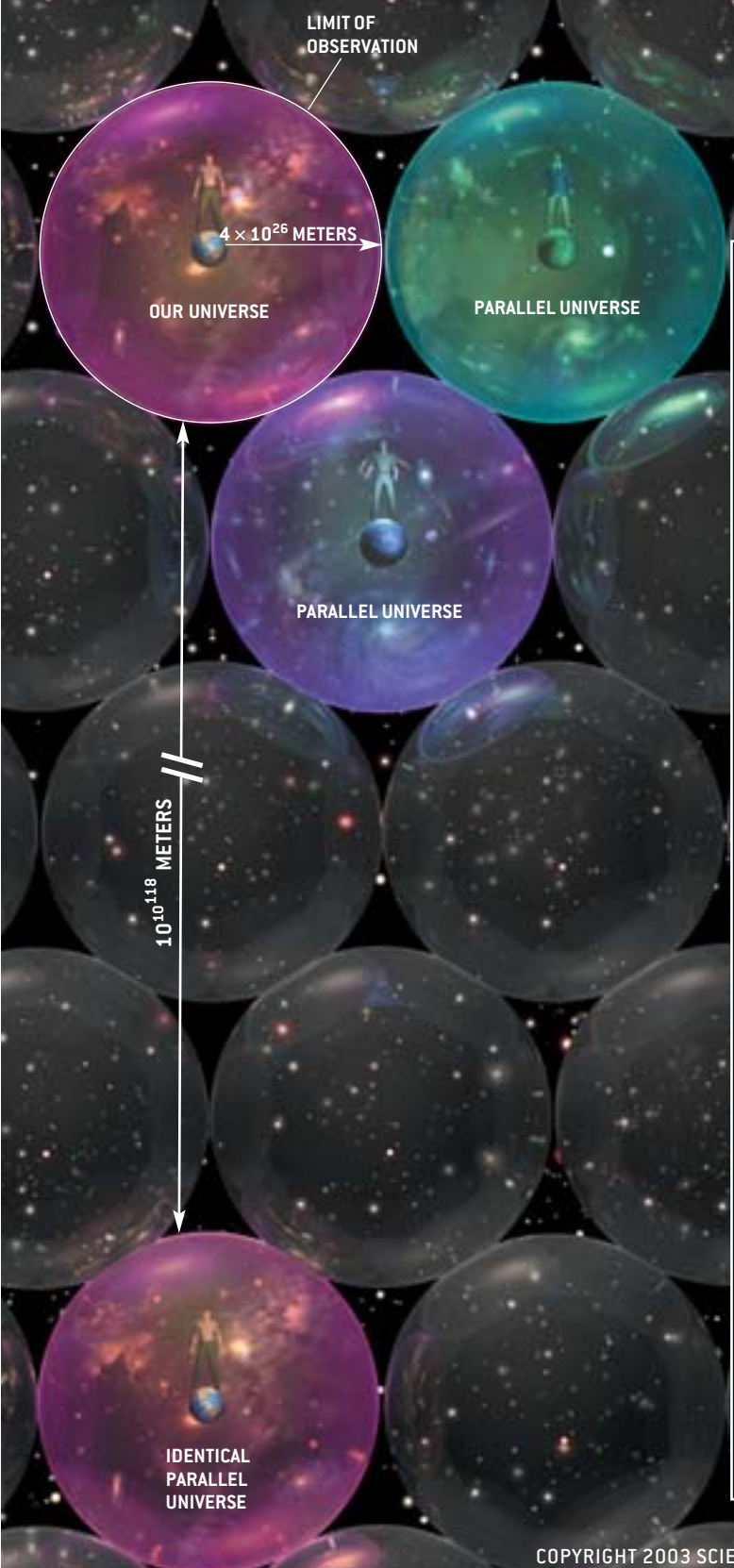
Overview/*Multiverses*

- One of the many implications of recent cosmological observations is that the concept of parallel universes is no mere metaphor. Space appears to be infinite in size. If so, then somewhere out there, everything that is possible becomes real, no matter how improbable it is. Beyond the range of our telescopes are other regions of space that are identical to ours. Those regions are a type of parallel universe. Scientists can even calculate how distant these universes are, on average.
- And that is fairly solid physics. When cosmologists consider theories that are less well established, they conclude that other universes can have entirely different properties and laws of physics. The presence of those universes would explain various strange aspects of our own. It could even answer fundamental questions about the nature of time and the comprehensibility of the physical world.

ALFRED T. KAMAJIAN (*background*); SARA CHEN (*inset*)

LEVEL I MULTIVERSE

THE SIMPLEST TYPE of parallel universe is simply a region of space that is too far away for us to have seen yet. The farthest that we can observe is currently about 4×10^{26} meters, or 42 billion light-years—the distance that light has been able to travel since the big bang began. (The distance is greater than 14 billion light-years because cosmic expansion has lengthened distances.) Each of the Level I parallel universes is basically the same as ours. All the differences stem from variations in the initial arrangement of matter.



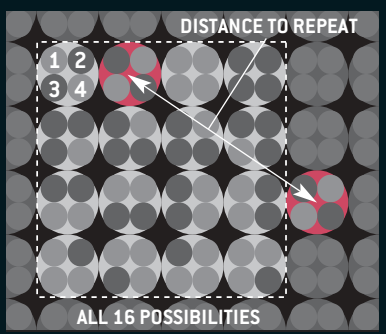
How Far Away Is a Duplicate Universe?

EXAMPLE UNIVERSE

Imagine a two-dimensional universe with space for four particles. Such a universe has 2^4, or 16, possible arrangements of matter. If more than 16 of these universes exist, they must begin to repeat. In this example, the distance to the nearest duplicate is roughly four times the diameter of each universe.

4 particles

2^4 arrangements

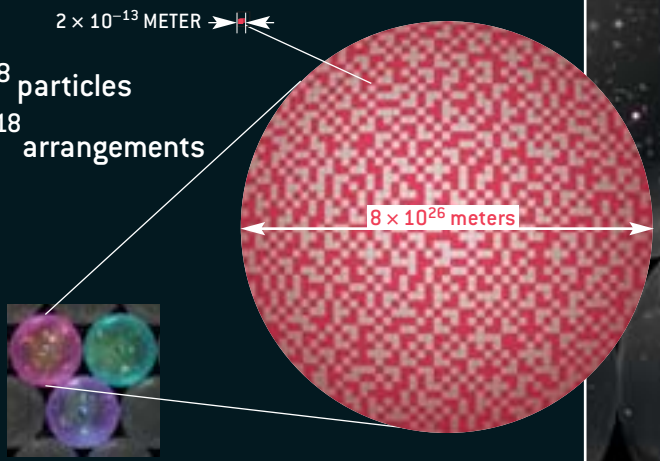


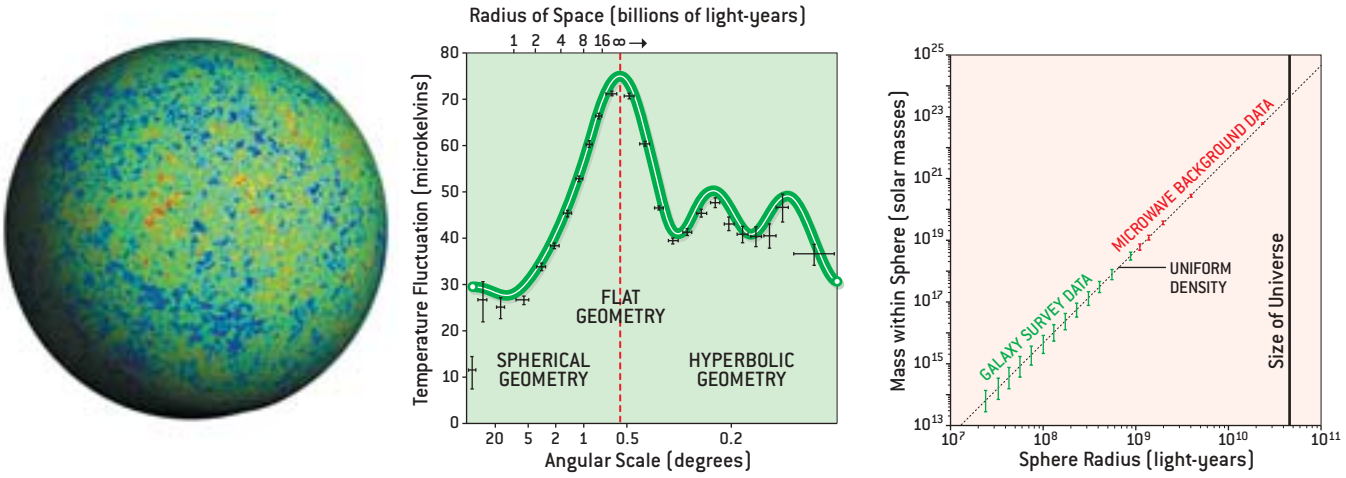
OUR UNIVERSE

The same argument applies to our universe, which has space for about 10^{118} subatomic particles. The number of possible arrangements is therefore 2 to the 10^{118}, or approximately 10 to the 10^{118}. Multiplying by the diameter of the universe gives an average distance to the nearest duplicate of 10 to the 10^{118} meters.

10^{118} particles

$2^{10^{118}}$ arrangements





COSMOLOGICAL DATA support the idea that space continues beyond the confines of our observable universe. The WMAP satellite recently measured the fluctuations in the microwave background (*left*). The strongest fluctuations are just over half a degree across, which indicates—after applying the rules of geometry—that space is very large or infinite (*center*). (One caveat: some cosmologists speculate that the discrepant point on the left of the graph is evidence for a finite volume.) In addition, WMAP and the 2dF Galaxy Redshift Survey have found that space on large scales is filled with matter uniformly (*right*), meaning that other universes should look basically like ours.

The lesson is that the multiverse theory can be tested and falsified even though we cannot see the other universes. The key is to predict what the ensemble of parallel universes is and to specify a probability distribution, or what mathematicians call a "measure," over that ensemble. Our universe should emerge as one of the most probable. If not—if, according to the multiverse theory, we live in an improbable universe—then the theory is in trouble. As I will discuss later, this measure problem can become quite challenging.

Level II: Other Postinflation Bubbles

IF THE LEVEL I MULTIVERSE was hard to stomach, try imagining an infinite set of distinct Level I multiverses, some perhaps with different spacetime dimensionality and different physical constants. Those other multiverses—which constitute a Level II multiverse—are predicted by the currently popular theory of chaotic eternal inflation.

Inflation is an extension of the big bang theory and ties up many of the loose ends of that theory, such as why the universe is so big, so uniform and so flat. A rapid stretching of space long ago can explain all these and other attributes in one fell swoop [see "The Inflationary Universe," by Alan H. Guth and Paul J. Steinhard; SCIENTIFIC AMERICAN, May 1984; and "The Self-Reproducing Inflationary Universe," by Andrei Linde, November 1994]. Such stretching is predicted by a wide class of theories of elementary particles, and all available evidence bears it out. The phrase "chaotic eternal" refers to what happens on the very largest scales. Space as a whole is stretching and will continue doing so forever, but some regions of space stop stretching and form distinct bubbles, like gas pockets in a loaf of rising bread. Infinitely many such bubbles emerge. Each is an embryonic Level I multiverse: infinite in size and filled with matter deposited by the energy field that drove inflation.

Those bubbles are more than infinitely far away from Earth, in the sense that you would never get there even if you traveled at the speed of light forever. The reason is that the space between our bubble and its neighbors is expanding faster than you could travel through it. Your descendants will never see their doppelgängers elsewhere in Level II. For the same reason, if cosmic expansion is accelerating, as observations now suggest, they might not see their alter egos even in Level I.

The Level II multiverse is far more diverse than the Level I multiverse. The bubbles vary not only in their initial conditions but also in seemingly immutable aspects of nature. The prevailing view in physics today is that the dimensionality of spacetime, the qualities of elementary particles and many of the so-called physical constants are not built into physical laws but are the outcome of processes known as symmetry breaking. For instance, theorists think that the space in our universe once had nine dimensions, all on an equal footing. Early in cosmic history, three of them partook in the cosmic expansion and became the three dimensions we now observe. The other six are now unobservable, either because they have stayed microscopic with a doughnutlike topology or because all matter is confined to a three-dimensional surface (a membrane, or simply "brane") in the nine-dimensional space.

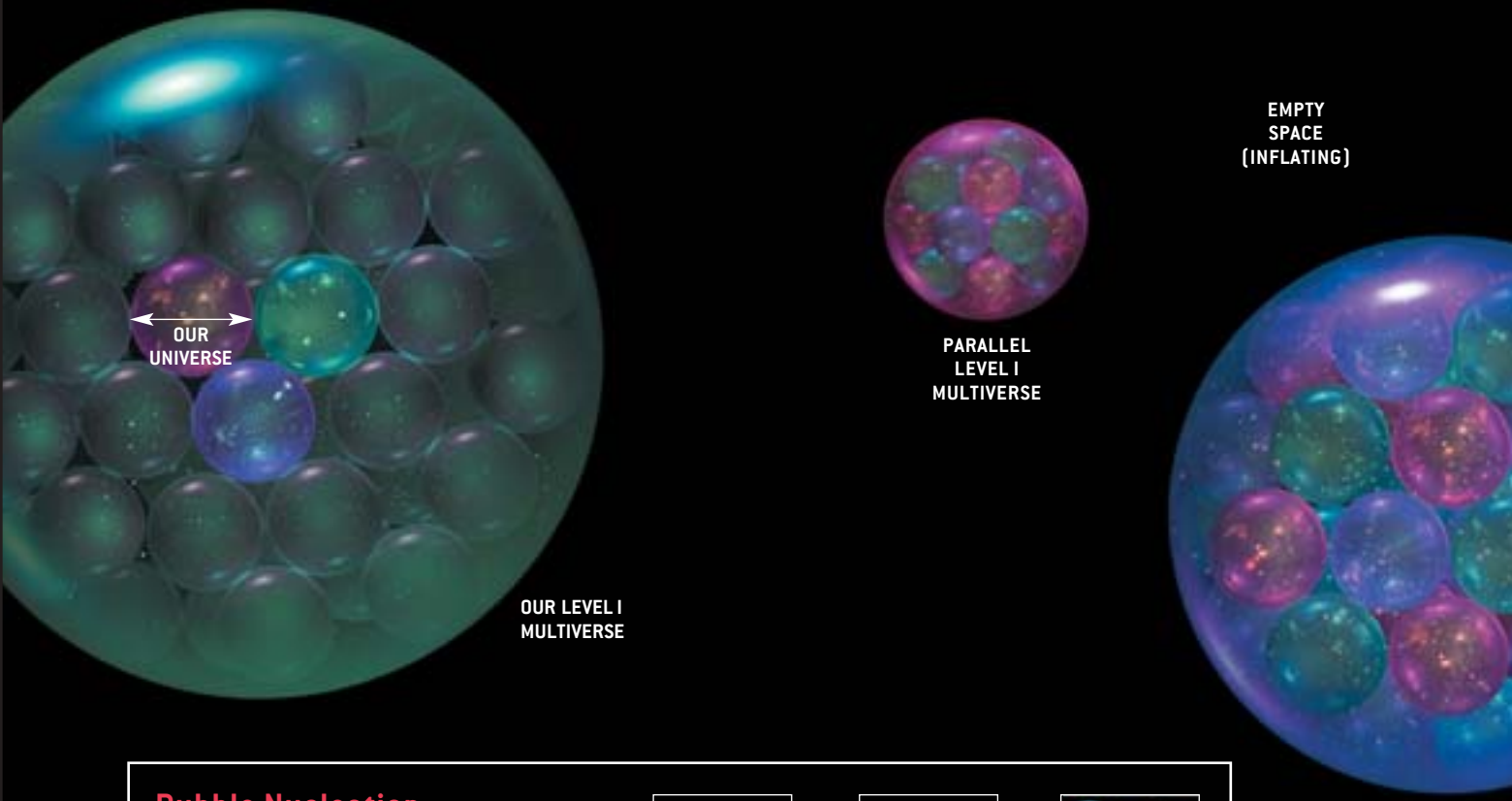
Thus, the original symmetry among the dimensions broke. The quantum fluctuations that drive chaotic inflation could cause different symmetry breaking in different bubbles. Some might become four-dimensional, others could contain only two rather than three generations of quarks, and still others might have a stronger cosmological constant than our universe does.

Another way to produce a Level II multiverse might be through a cycle of birth and destruction of universes. In a scientific context, this idea was introduced by physicist Richard C. Tolman in the 1930s and recently elaborated on by Paul J. Steinhardt of Princeton University and Neil Turok of the University of Cambridge. The Steinhardt and Turok proposal and related models involve a second three-dimensional brane that is quite literally parallel to ours, merely offset in a higher dimension [see "Been There, Done That," by George Musser; News Scan, SCIENTIFIC AMERICAN, March 2002]. This parallel universe is not

MAX TEGMARK (*sphere*); SARA CHEN (*graphs*)

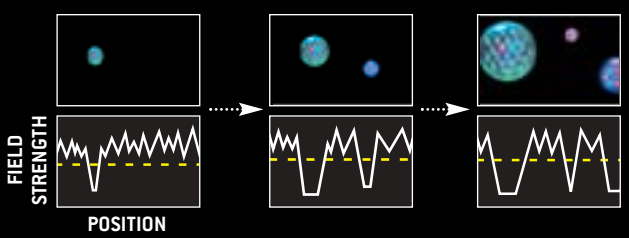
LEVEL II MULTIVERSE

A SOMEWHAT MORE ELABORATE type of parallel universe emerges from the theory of cosmological inflation. The idea is that our Level I multiverse—namely, our universe and contiguous regions of space—is a bubble embedded in an even vaster but mostly empty volume. Other bubbles exist out there, disconnected from ours. They nucleate like raindrops in a cloud. During nucleation, variations in quantum fields endow each bubble with properties that distinguish it from other bubbles.



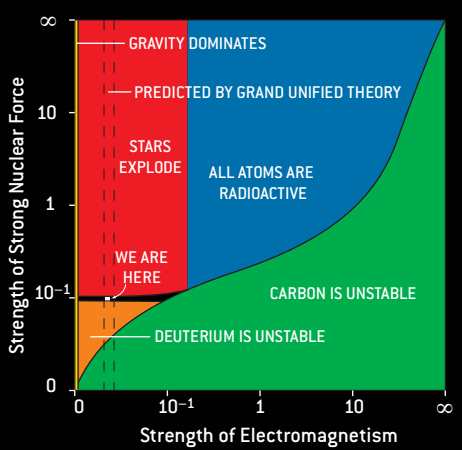
Bubble Nucleation

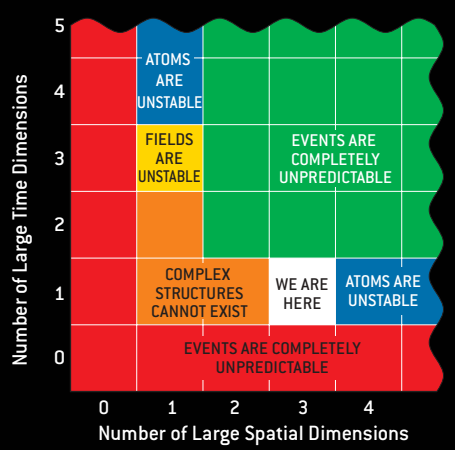
A QUANTUM FIELD known as the inflaton causes space to expand rapidly. In the bulk of space, random fluctuations prevent the field from decaying away. But in certain regions, the field loses its strength and the expansion slows down. Those regions become bubbles.



Evidence

COSMOLOGISTS INFER the presence of Level II parallel universes by scrutinizing the properties of our universe. These properties, including the strength of the forces of nature (*right*) and the number of observable space and time dimensions (*far right*), were established by random processes during the birth of our universe. Yet they have exactly the values that sustain life. That suggests the existence of other universes with other values.





ALFRED T. KAMAJIAN (*background*); CORNELIA BLIK (*top inset*); SARA CHEN (*bottom inset*)

really a separate universe, because it interacts with ours. But the ensemble of universes—past, present and future—that these branes create would form a multiverse, arguably with a diversity similar to that produced by chaotic inflation. An idea proposed by physicist Lee Smolin of the Perimeter Institute in Waterloo, Ontario, involves yet another multiverse comparable in diversity to that of Level II but mutating and sprouting new universes through black holes rather than through brane physics.

Although we cannot interact with other Level II parallel universes, cosmologists can infer their presence indirectly, because their existence can account for unexplained coincidences in our universe. To give an analogy, suppose you check into a hotel, are assigned room 1967 and note that this is the year you were born. What a coincidence, you say. After a moment of reflection, however, you conclude that this is not so surprising after all. The hotel has hundreds of rooms, and you would not have been having these thoughts in the first place if you had been assigned one with a number that meant nothing to you. The lesson is that even if you knew nothing about hotels, you could infer the existence of other hotel rooms to explain the coincidence.

As a more pertinent example, consider the mass of the sun. The mass of a star determines its luminosity, and using basic physics, one can compute that life as we know it on Earth is possible only if the sun's mass falls into the narrow range between 1.6×10^{30} and 2.4×10^{30} kilograms. Otherwise Earth's climate would be colder than that of present-day Mars or hotter than that of present-day Venus. The measured solar mass is 2.0×10^{30} kilograms. At first glance, this apparent coincidence of the habitable and observed mass values appears to be a wild stroke of luck. Stellar masses run from 10^{29} to 10^{32} kilograms, so if the sun acquired its mass at random, it had only a small chance of falling into the habitable range. But just as in the hotel example, one can explain this apparent coincidence by postulating an ensemble (in this case, a number of planetary systems) and a selection effect (the fact that we must find ourselves living on a habitable planet). Such observer-related selection effects are referred to as "anthropic," and although the "A-word" is notorious for triggering controversy, physicists broadly agree that these selection effects cannot be neglected when testing fundamental theories.

What applies to hotel rooms and planetary systems applies to parallel universes. Most, if not all, of the attributes set by symmetry breaking appear to be fine-tuned. Changing their values by modest amounts would have resulted in a qualitatively different universe—one in which we probably would not exist. If protons were 0.2 percent heavier, they could decay into neutrons, destabilizing atoms. If the electromagnetic force were 4 percent weaker, there would be no hydrogen and no normal stars. If the weak interaction were much weaker, hydrogen would not exist; if it were much stronger, supernovae would fail to seed interstellar space with heavy elements. If the cosmological constant were much larger, the universe would have blown itself apart before galaxies could form.

Although the degree of fine-tuning is still debated, these examples suggest the existence of parallel universes with other values of the physical constants [see "Exploring Our Universe and Others," by Martin Rees; SCIENTIFIC AMERICAN, December 1999]. The Level II multiverse theory predicts that physicists will never be able to determine the values of these constants from first principles. They will merely compute probability distributions for what they should expect to find, taking selection effects into account. The result should be as generic as is consistent with our existence.

Level III: Quantum Many Worlds

THE LEVEL I AND LEVEL II multiverses involve parallel worlds that are far away, beyond the domain even of astronomers. But the next level of multiverse is right around you. It arises from the famous, and famously controversial, many-worlds interpretation of quantum mechanics—the idea that random quantum processes cause the universe to branch into multiple copies, one for each possible outcome.

In the early 20th century the theory of quantum mechanics revolutionized physics by explaining the atomic realm, which does not abide by the classical rules of Newtonian mechanics. Despite the obvious successes of the theory, a heated debate rages about what it really means. The theory specifies the state of the universe not in classical terms, such as the positions and velocities of all particles, but in terms of a mathematical object called a wave function. According to the Schrödinger equation, this state evolves over time in a fashion that mathematicians term "unitary," meaning that the wave function rotates in an abstract infinite-dimensional space called Hilbert space. Although quantum mechanics is often described as inherently random and uncertain, the wave function evolves in a deterministic way. There is nothing random or uncertain about it.

The sticky part is how to connect this wave function with what we observe. Many legitimate wave functions correspond to counterintuitive situations, such as a cat being dead and alive at the same time in a so-called superposition. In the 1920s physicists explained away this weirdness by postulating that the wave function "collapsed" into some definite classical outcome whenever someone made an observation. This add-on had the virtue of explaining observations, but it turned an elegant, unitary theory into a kludgy, nonunitary one. The intrinsic randomness commonly ascribed to quantum mechanics is the result of this postulate.

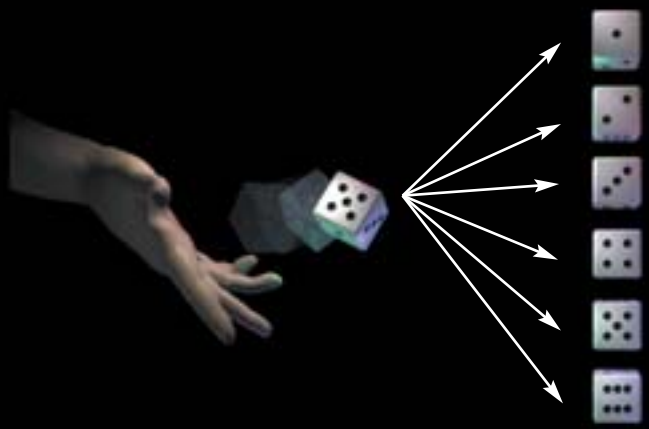
Over the years many physicists have abandoned this view in favor of one developed in 1957 by Princeton graduate student Hugh Everett III. He showed that the collapse postulate is unnecessary. Unadulterated quantum theory does not, in fact, pose any contradictions. Although it predicts that one classical reality gradually splits into superpositions of many such realities, observers subjectively experience this splitting merely as a slight randomness, with probabilities in exact agreement with those from the old collapse postulate. This superposition of classical worlds is the Level III multiverse.

Everett's many-worlds interpretation has been boggling minds inside and outside physics for more than four decades. But the theory becomes easier to grasp when one distinguishes

ALFRED T. KAMAJIAN

LEVEL III MULTIVERSE

QUANTUM MECHANICS PREDICTS a vast number of parallel universes by broadening the concept of "elsewhere." These universes are located elsewhere, not in ordinary space but in an abstract realm of all possible states. Every conceivable way that the world could be (within the scope of quantum mechanics) corresponds to a different universe. The parallel universes make their presence felt in laboratory experiments, such as wave interference and quantum computation.

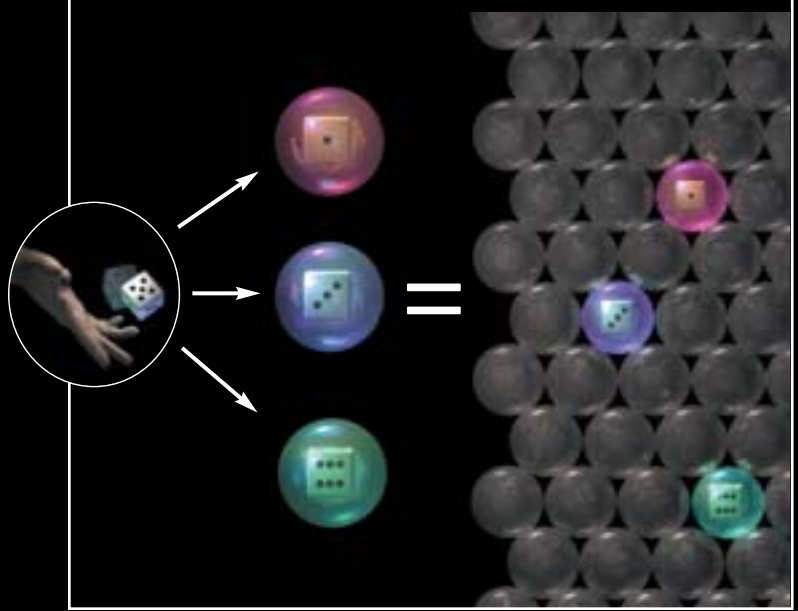


Quantum Dice

IMAGINE AN IDEAL DIE whose randomness is purely quantum. When you roll it, the die appears to land on a certain value at random. Quantum mechanics, however, predicts that it lands on all values at once. One way to reconcile these contradictory views is to conclude that the die lands on different values in different universes. In one sixth of the universes, it lands on 1; in one sixth, on 2, and so on. Trapped within one universe, we can perceive only a fraction of the full quantum reality.

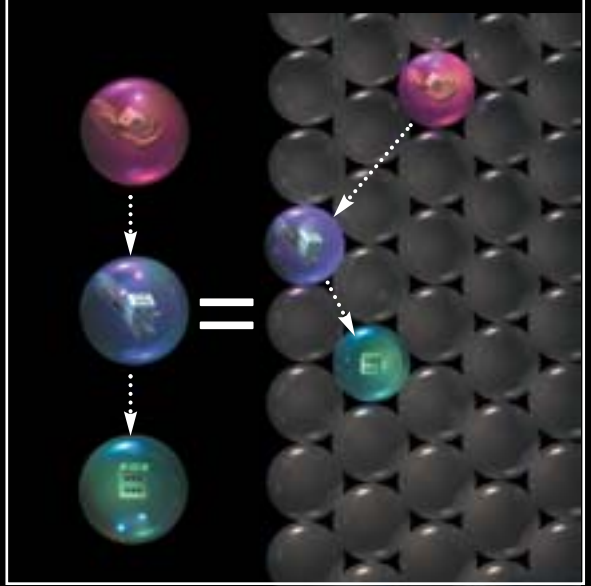
Ergodicity

ACCORDING TO THE PRINCIPLE of ergodicity, quantum parallel universes are equivalent to more prosaic types of parallel universes. A quantum universe splits over time into multiple universes (*left*). Yet those new universes are no different from parallel universes that already exist somewhere else in space—in, for example, other Level I universes (*right*). The key idea is that parallel universes, of whatever type, embody different ways that events could have unfolded.



The Nature of Time

MOST PEOPLE THINK of time as a way to describe change. At one moment, matter has a certain arrangement; a moment later, it has another (*left*). The concept of multiverses suggests an alternative view. If parallel universes contain all possible arrangements of matter (*right*), then time is simply a way to put those universes into a sequence. The universes themselves are static; change is an illusion, albeit an interesting one.



between two ways of viewing a physical theory: the outside view of a physicist studying its mathematical equations, like a bird surveying a landscape from high above it, and the inside view of an observer living in the world described by the equations, like a frog living in the landscape surveyed by the bird.

From the bird perspective, the Level III multiverse is simple. There is only one wave function. It evolves smoothly and deterministically over time without any kind of splitting or parallelism. The abstract quantum world described by this evolving wave function contains within it a vast number of parallel classical story lines, continuously splitting and merging, as well as a number of quantum phenomena that lack a classical description. From their frog perspective, observers perceive only a tiny fraction of this full reality. They can view their own Level I universe, but a process called decoherence—which mimics wave function collapse while preserving unitarity—prevents them from seeing Level III parallel copies of themselves.

Whenever observers are asked a question, make a snap decision and give an answer, quantum effects in their brains lead to a superposition of outcomes, such as "Continue reading the article" and "Put down the article." From the bird perspective, the act of making a decision causes a person to split into multiple copies: one who keeps on reading and one who doesn't. From their frog perspective, however, each of these alter egos is unaware of the others and notices the branching merely as a slight randomness: a certain probability of continuing to read or not.

As strange as this may sound, the exact same situation occurs even in the Level I multiverse. You have evidently decided to keep on reading the article, but one of your alter egos in a distant galaxy put down the magazine after the first paragraph. The only difference between Level I and Level III is where your doppelgängers reside. In Level I they live elsewhere in good old three-dimensional space. In Level III they live on another quantum branch in infinite-dimensional Hilbert space.

The existence of Level III depends on one crucial assumption: that the time evolution of the wave function is unitary. So far experimenters have encountered no departures from unitarity. In the past few decades they have confirmed unitarity for ever larger systems, including carbon 60 buckyball molecules and kilometer-long optical fibers. On the theoretical side, the case for unitarity has been bolstered by the discovery of decoherence [see "100 Years of Quantum Mysteries," by Max Tegmark and John Archibald Wheeler; Scientific American, February 2001]. Some theorists who work on quantum gravity have questioned unitarity; one concern is that evaporating black holes might destroy information, which would be a nonunitary process. But a recent breakthrough in string theory known as AdS/CFT correspondence suggests that even quantum gravity is unitary. If so, black holes do not destroy information but merely transmit it elsewhere. [*Editors' note: An upcoming article will discuss this correspondence in greater detail.*]

If physics is unitary, then the standard picture of how quantum fluctuations operated early in the big bang must change. These fluctuations did not generate initial conditions at random. Rather they generated a quantum superposition of all possible initial conditions, which coexisted simultaneously. Decoherence then caused these initial conditions to behave classically in separate quantum branches. Here is the crucial point: the distribution of outcomes on different quantum branches in a given Hubble volume (Level III) is identical to the distribution of outcomes in different Hubble volumes within a single quantum branch (Level I). This property of the quantum fluctuations is known in statistical mechanics as ergodicity.

The same reasoning applies to Level II. The process of symmetry breaking did not produce a unique outcome but rather a superposition of all outcomes, which rapidly went their separate ways. So if physical constants, spacetime dimensionality and so on can vary among parallel quantum branches at Level III, then they will also vary among parallel universes at Level II.

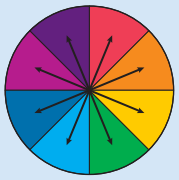
In other words, the Level III multiverse adds nothing new beyond Level I and Level II, just more indistinguishable copies of the same universes—the same old story lines playing out again and again in other quantum branches. The passionate debate about Everett's theory therefore seems to be ending in a grand anticlimax, with the discovery of less controversial multiverses (Levels I and II) that are equally large.

Needless to say, the implications are profound, and physicists are only beginning to explore them. For instance, consider the ramifications of the answer to a long-standing question: Does the number of universes exponentially increase over time? The surprising answer is no. From the bird perspective, there is of course only one quantum universe. From the frog perspective, what matters is the number of universes that are distinguishable at a given instant—that is, the number of noticeably different Hubble volumes. Imagine moving planets to random new locations, imagine having married someone else, and so on. At the quantum level, there are 10 to the 10^{118} universes with temperatures below 10^8 kelvins. That is a vast number, but a finite one.

From the frog perspective, the evolution of the wave function corresponds to a never-ending sliding from one of these 10 to the 10^{118} states to another. Now you are in universe A, the one in which you are reading this sentence. Now you are in universe B, the one in which you are reading this other sentence. Put differently, universe B has an observer identical to one in universe A, except with an extra instant of memories. All possible states exist at every instant, so the passage of time may be in the eye of the beholder—an idea explored in Greg Egan's

THE AUTHOR

MAX TEGMARK wrote a four-dimensional version of the computer game Tetris while in college. In another universe, he went on to become a highly paid software developer. In our universe, however, he wound up as professor of physics and astronomy at the University of Pennsylvania. Tegmark is an expert in analyzing the cosmic microwave background and galaxy clustering. Much of his work bears on the concept of parallel universes: evaluating evidence for infinite space and cosmological inflation; developing insights into quantum decoherence; and studying the possibility that the amplitude of microwave background fluctuations, the dimensionality of spacetime and the fundamental laws of physics can vary from place to place.



The Mystery of Probability: What Are the Odds?

AS MULTIVERSE THEORIES gain credence, the sticky issue of how to compute probabilities in physics is growing from a minor nuisance into a major embarrassment. If there are indeed many identical copies of you, the traditional notion of determinism evaporates. You could not compute your own future even if you had complete knowledge of the entire state of the multiverse, because there is no way for you to determine which of these copies is you (they all feel they are). All you can predict, therefore, are probabilities for what you would observe. If an outcome has a probability of, say, 50 percent, it means that half the observers observe that outcome.

Unfortunately, it is not an easy task to compute what fraction of the infinitely many observers perceive what. The answer depends on the order in which you count them. By analogy, the fraction of the integers that are even is 50 percent if you order them numerically (1, 2, 3, 4, ...) but approaches 100 percent if you sort them digit by digit, the way your word processor would (1, 10, 100, 1,000, ...). When observers reside in disconnected universes, there is no obviously natural way in which to order them. Instead one must sample from the different universes with some statistical weights referred to by mathematicians as a "measure."

This problem crops up in a mild and treatable manner at Level I, becomes severe at Level II, has caused much debate at Level III, and is horrendous at Level IV. At Level II, for instance, Alexander Vilenkin of Tufts University and others have published predictions for the probability distributions of various cosmological parameters. They have argued that different parallel universes that have inflated by different amounts should be given statistical weights proportional to their volume. On the other hand, any mathematician will tell you that $2 \times \infty = \infty$, so there is no objective sense in which an infinite universe that has expanded by a factor of two has gotten larger. Moreover, a finite universe with the topology of a torus is equivalent to a perfectly periodic universe with infinite volume, both from the mathematical bird perspective and from the frog perspective of an observer within it. So why should its infinitely smaller volume give it zero statistical weight? After all, even in the Level I multiverse, Hubble volumes start repeating (albeit in a random order, not periodically) after about 10 to the 10^{118} meters.

If you think that is bad, consider the problem of assigning statistical weights to different mathematical structures at Level IV. The fact that our universe seems relatively simple has led many people to suggest that the correct measure somehow involves complexity. —*M.T.*

1994 science-fiction novel *Permutation City* and developed by physicist David Deutsch of the University of Oxford, independent physicist Julian Barbour, and others. The multiverse framework may thus prove essential to understanding the nature of time.

Level IV: Other Mathematical Structures

THE INITIAL CONDITIONS and physical constants in the Level I, Level II and Level III multiverses can vary, but the fundamental laws that govern nature remain the same. Why stop there? Why not allow the laws themselves to vary? How about a universe that obeys the laws of classical physics, with no quantum effects? How about time that comes in discrete steps, as for computers, instead of being continuous? How about a universe that is simply an empty dodecahedron? In the Level IV multiverse, all these alternative realities actually exist.

A hint that such a multiverse might not be just some beer-fueled speculation is the tight correspondence between the worlds of abstract reasoning and of observed reality. Equations and, more generally, mathematical structures such as numbers, vectors and geometric objects describe the world with remarkable verisimilitude. In a famous 1959 lecture, physicist Eugene P. Wigner argued that "the enormous usefulness of mathematics in the natural sciences is something bordering on the mysterious." Conversely, mathematical structures have an eerily real feel to them. They satisfy a central criterion of objective existence: they are the same no matter who studies them. A theorem is true regardless of whether it is proved by a human, a computer or an intelligent dolphin. Contemplative alien civilizations would find the same mathematical structures as we have. Accordingly, mathematicians commonly say that they discover mathematical structures rather than create them.

There are two tenable but diametrically opposed paradigms for understanding the correspondence between mathematics and physics, a dichotomy that arguably goes as far back as Plato and Aristotle. According to the Aristotelian paradigm, physical reality is fundamental and mathematical language is merely a useful approximation. According to the Platonic paradigm, the mathematical structure is the true reality and observers perceive it imperfectly. In other words, the two paradigms disagree on which is more basic, the frog perspective of the observer or the bird perspective of the physical laws. The Aristotelian paradigm prefers the frog perspective, whereas the Platonic paradigm prefers the bird perspective.

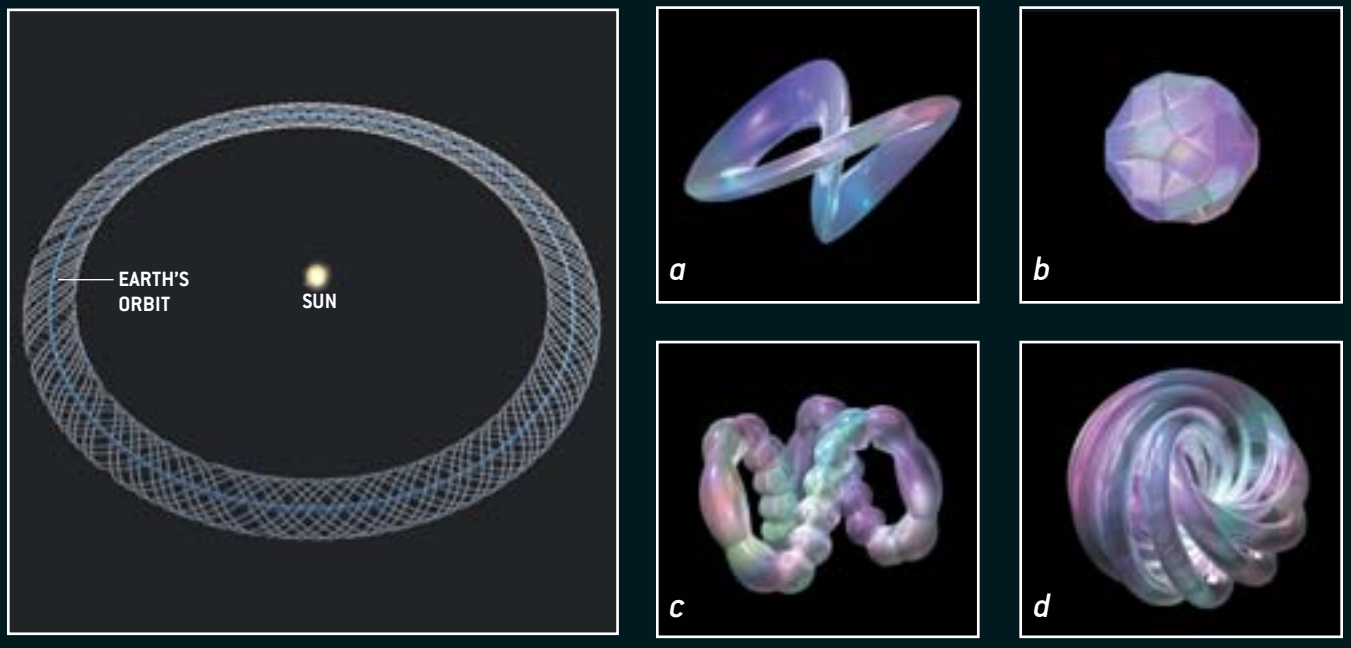
As children, long before we had even heard of mathematics, we were all indoctrinated with the Aristotelian paradigm. The Platonic view is an acquired taste. Modern theoretical physicists tend to be Platonists, suspecting that mathematics describes the universe so well because the universe is inherently mathematical. Then all of physics is ultimately a mathematics problem: a mathematician with unlimited intelligence and resources could in principle compute the frog perspective—that is, compute what self-aware observers the universe contains, what they perceive, and what languages they invent to describe their perceptions to one another.

A mathematical structure is an abstract, immutable entity existing outside of space and time. If history were a movie, the structure would correspond not to a single frame of it but to the entire videotape. Consider, for example, a world made up of pointlike particles moving around in three-dimensional space.

SARA CHEN

LEVEL IV MULTIVERSE

THE ULTIMATE TYPE of parallel universe opens up the full realm of possibility. Universes can differ not just in location, cosmological properties or quantum state but also in the laws of physics. Existing outside of space and time, they are almost impossible to visualize; the best one can do is to think of them abstractly, as static sculptures that represent the mathematical structure of the physical laws that govern them. For example, consider a simple universe: Earth, moon and sun, obeying Newton's laws. To an objective observer, this universe looks like a circular ring (Earth's orbit smeared out in time) wrapped in a braid (the moon's orbit around Earth). Other shapes embody other laws of physics (*a, b, c, d*). This paradigm solves various problems concerning the foundations of physics.



In four-dimensional spacetime—the bird perspective—these particle trajectories resemble a tangle of spaghetti. If the frog sees a particle moving with constant velocity, the bird sees a straight strand of uncooked spaghetti. If the frog sees a pair of orbiting particles, the bird sees two spaghetti strands intertwined like a double helix. To the frog, the world is described by Newton's laws of motion and gravitation. To the bird, it is described by the geometry of the pasta—a mathematical structure. The frog itself is merely a thick bundle of pasta, whose highly complex intertwining corresponds to a cluster of particles that store and process information. Our universe is far more complicated than this example, and scientists do not yet know to what, if any, mathematical structure it corresponds.

The Platonic paradigm raises the question of why the universe is the way it is. To an Aristotelian, this is a meaningless question: the universe just is. But a Platonist cannot help but wonder why it could not have been different. If the universe is inherently mathematical, then why was only one of the many mathematical structures singled out to describe a universe? A fundamental asymmetry appears to be built into the very heart of reality.

As a way out of this conundrum, I have suggested that complete mathematical symmetry holds: that all mathematical structures exist physically as well. Every mathematical structure corresponds to a parallel universe. The elements of this multiverse do not reside in the same space but exist outside of space and time. Most of them are probably devoid of observers. This hypothesis can be viewed as a form of radical Platonism, asserting that the mathematical structures in Plato's realm of ideas or the "mindscape" of mathematician Rudy Rucker of San Jose State University exist in a physical sense. It is akin to what cosmologist John D. Barrow of the University of Cambridge refers to as "π in the sky," what the late Harvard University philosopher Robert Nozick called the principle of fecundity and what the late Princeton philosopher David K. Lewis called modal realism. Level IV brings closure to the hierarchy of multiverses, because any self-consistent fundamental physical theory can be phrased as some kind of mathematical structure.

The Level IV multiverse hypothesis makes testable predictions. As with Level II, it involves an ensemble (in this case, the full range of mathematical structures) and selection effects. As mathematicians continue to categorize mathematical structures, they should find that the structure describing our world is the most generic one consistent with our observations. Similarly, our future observations should be the most generic ones that are consistent with our past observations, and our past observations should be the most generic ones that are consistent with our existence.

Quantifying what "generic" means is a severe problem, and this investigation is only now beginning. But one striking and

BRYAN CHRISTIE DESIGN (*left*); ALFRED T. KAMAJIAN (*a–d*)

encouraging feature of mathematical structures is that the symmetry and invariance properties that are responsible for the simplicity and orderliness of our universe tend to be generic, more the rule than the exception. Mathematical structures tend to have them by default, and complicated additional axioms must be added to make them go away.

What Says Occam?

THE SCIENTIFIC THEORIES of parallel universes, therefore, form a four-level hierarchy, in which universes become progressively more different from ours. They might have different initial conditions (Level I); different physical constants and particles (Level II); or different physical laws (Level IV). It is ironic that Level III is the one that has drawn the most fire in the past decades, because it is the only one that adds no qualitatively new types of universes.

In the coming decade, dramatically improved cosmological measurements of the microwave background and the large-scale matter distribution will support or refute Level I by further pinning down the curvature and topology of space. These measurements will also probe Level II by testing the theory of chaotic eternal inflation. Progress in both astrophysics and high-energy physics should also clarify the extent to which physical constants are fine-tuned, thereby weakening or strengthening the case for Level II.

If current efforts to build quantum computers succeed, they will provide further evidence for Level III, as they would, in essence, be exploiting the parallelism of the Level III multiverse for parallel computation. Experimenters are also looking for evidence of unitarity violation, which would rule out Level III. Finally, success or failure in the grand challenge of modern physics—unifying general relativity and quantum field theory—will sway opinions on Level IV. Either we will find a mathematical structure that exactly matches our universe, or we will bump up against a limit to the unreasonable effectiveness of mathematics and have to abandon that level.

So should you believe in parallel universes? The principal arguments against them are that they are wasteful and that they are weird. The first argument is that multiverse theories are vulnerable to Occam's razor because they postulate the existence of other worlds that we can never observe. Why should nature be so wasteful and indulge in such opulence as an infinity of different worlds? Yet this argument can be turned around to argue *for* a multiverse. What precisely would nature be wasting? Certainly not space, mass or atoms—the uncontroversial Level I multiverse already contains an infinite amount of all three, so who cares if nature wastes some more? The real issue here is the apparent reduction in simplicity. A skeptic worries about all the information necessary to specify all those unseen worlds.

But an entire ensemble is often much simpler than one of its members. This principle can be stated more formally using the notion of algorithmic information content. The algorithmic information content in a number is, roughly speaking, the length of the shortest computer program that will produce that number as output. For example, consider the set of all integers. Which is simpler, the whole set or just one number? Naively, you might think that a single number is simpler, but the entire set can be generated by quite a trivial computer program, whereas a single number can be hugely long. Therefore, the whole set is actually simpler.

Similarly, the set of all solutions to Einstein's field equations is simpler than a specific solution. The former is described by a few equations, whereas the latter requires the specification of vast amounts of initial data on some hypersurface. The lesson is that complexity increases when we restrict our attention to one particular element in an ensemble, thereby losing the symmetry and simplicity that were inherent in the totality of all the elements taken together.

In this sense, the higher-level multiverses are simpler. Going from our universe to the Level I multiverse eliminates the need to specify initial conditions, upgrading to Level II eliminates the need to specify physical constants, and the Level IV multiverse eliminates the need to specify anything at all. The opulence of complexity is all in the subjective perceptions of observers—the frog perspective. From the bird perspective, the multiverse could hardly be any simpler.

The complaint about weirdness is aesthetic rather than scientific, and it really makes sense only in the Aristotelian worldview. Yet what did we expect? When we ask a profound question about the nature of reality, do we not expect an answer that sounds strange? Evolution provided us with intuition for the everyday physics that had survival value for our distant ancestors, so whenever we venture beyond the everyday world, we should expect it to seem bizarre.

A common feature of all four multiverse levels is that the simplest and arguably most elegant theory involves parallel universes by default. To deny the existence of those universes, one needs to complicate the theory by adding experimentally unsupported processes and ad hoc postulates: finite space, wave function collapse and ontological asymmetry. Our judgment therefore comes down to which we find more wasteful and inelegant: many worlds or many words. Perhaps we will gradually get used to the weird ways of our cosmos and find its strangeness to be part of its charm. SA

MORE TO EXPLORE

Why Is the CMB Fluctuation Level 10^{-5}? Max Tegmark and Martin Rees in *Astrophysical Journal,* Vol. 499, No. 2, pages 526–532; June 1, 1998. Available online at **arXiv.org/abs/astro-ph/9709058**

Is "The Theory of Everything" Merely the Ultimate Ensemble Theory? Max Tegmark in *Annals of Physics,* Vol. 270, No.1, pages 1–51; November 20, 1998. Available online at **arXiv.org/abs/gr-qc/9704009**

Many Worlds in One. Jaume Garriga and Alexander Vilenkin in *Physical Review,* Vol. D64, No. 043511; July 26, 2001. Available online at **arXiv.org/abs/gr-qc/0102010**

Our Cosmic Habitat. Martin Rees. Princeton University Press, 2001.

Inflation, Quantum Cosmology and the Anthropic Principle. Andrei Linde in *Science and Ultimate Reality: From Quantum to Cosmos*. Edited by J. D. Barrow, P.C.W. Davies and C. L. Harper. Cambridge University Press, 2003. Available online at **arXiv.org/abs/hep-th/0211048**

The author's Web site has more information at **www.hep.upenn.edu/~max/multiverse.html**

Hearing Colors, Tasting Shapes

People with synesthesia—whose senses blend together—are providing valuable clues to understanding the organization and functions of the human brain

By Vilayanur S. Ramachandran and Edward M. Hubbard

When Matthew Blakeslee shapes hamburger patties with his hands, he experiences a vivid bitter taste in his mouth. Esmerelda Jones (a pseudonym) sees blue when she listens to the note C sharp played on the piano; other notes evoke different hues—so much so that the piano keys are actually color-coded, making it easier for her to remember and play musical scales. And when Jeff Coleman looks at printed black numbers, he sees them in color, each a different hue. Blakeslee, Jones and Coleman are among a handful of otherwise normal people who have synesthesia. They experience the ordinary world in extraordinary ways and seem to inhabit a mysterious no-man's-land between fantasy and reality. For them the senses—touch, taste, hearing, vision and smell—get mixed up instead of remaining separate.

Modern scientists have known about synesthesia since 1880, when Francis Galton, a cousin of Charles Darwin, published a paper in *Nature* on the phenomenon. But most have brushed it aside as fakery, an artifact of drug use (LSD and mescaline can produce similar effects) or a mere curiosity. About four years ago, however, we and others began to uncover brain processes that could account for synesthesia. Along the way, we also found new clues to some of the most mysterious aspects of the human mind, such as the emergence of abstract thought, metaphor and perhaps even language.

A common explanation of synesthesia is that the affected people are simply experiencing childhood memories and associations. Maybe a person had played with refrigerator magnets as a child and the number 5 was red and 6 was green. This theory does not answer why only some people retain such vivid sensory memories, however. You might *think* of cold when you look at a picture of an ice cube, but you probably do not feel cold, no matter how many encounters you may have had with ice and snow during your youth.

Another prevalent idea is that synesthetes are merely being metaphorical when they describe the note C flat as "red" or say that chicken tastes "pointy"—just as you and I might speak of a "loud" shirt or "sharp" cheddar cheese. Our ordinary language is replete with such sense-related metaphors, and perhaps synesthetes are just especially gifted in this regard.

We began trying to find out whether synesthesia is a genuine sensory experience in 1999. This deceptively simple question had plagued researchers in this field for decades. One natural approach is to start by asking the subjects outright: "Is this just a memory, or do you actually see the color as if it were right in front of you?" When we tried asking this question, we did not get very far. Some subjects did respond, "Oh, I see it per-

DAVID EMMITE

fectly clearly." But a more frequent reaction was, "I kind of see it, kind of don't" or "No, it is not like a memory. I see the number as being clearly red but I also *know* it isn't; it's black. So it must be a memory, I guess."

To determine whether an effect is truly perceptual, psychologists often use a simple test called pop-out or segregation. If you look at a set of tilted lines scattered amid a forest of vertical lines, the tilted lines stand out. Indeed, you can instantly segregate them from the background and group them mentally to form, for example, a separate triangular shape. Similarly, if most of a background's elements were green dots and you were told to look for red targets, the reds would pop out. On the other hand, a set of black 2's scattered among 5's of the same color almost blend in [*see illustration on page 57*]. It is hard to discern the 2's without engaging in an item-by-item inspection of numbers, even though any individual number is just as clearly different from its neighbors as a tilted line is from a straight line. We thus may conclude that only certain primitive, or elementary, features, such as color and line orientation, can provide a basis for grouping. More complex perceptual tokens, such as numbers, cannot do so.

We wondered what would happen if we showed the mixed numbers to synesthetes who experience, for instance, red when they see a 5 and green with a 2. We arranged the 2's so that they formed a triangle. If synesthesia were a genuine sensory effect, our subjects should easily *see* the triangle because for them, the numbers would look colored.

When we conducted pop-out tests with volunteers, the answer was crystal clear. Unlike normal subjects, synesthetes correctly reported the shape formed by groups of numbers up to 90 percent of the time (exactly as nonsynesthetes do when the numbers actually have different colors). This result proves that the induced colors are genuinely sensory and that synesthetes are not just making things up. It is impossible for them to fake their success. In another striking example, we asked a synesthete who sees 5 tinged red to watch a computer display. He could not tell when we surreptitiously added an actual red hue to the white number unless the red was sufficiently intense; he could instantly spot a real green added to the 5.

Visual Processing

CONFIRMATION THAT synesthesia is real brings up the question, Why do some people experience this weird phenomenon? Our experiments lead us to favor the idea that synesthetes are experiencing the result of some kind of cross wiring in the brain. This basic concept was initially proposed about 100 years ago, but we have now identified where in the brain and how such cross wiring might occur.

An understanding of the neurobiological factors at work requires some familiarity with how the brain processes visual information [*see illustration on opposite page*]. After light reflected from a scene hits the cones (color receptors) in the eye, neural signals from the retina travel to area 17, in the occipital lobe at the back of the brain. There the image is processed further within local clusters, or blobs, into such simple attributes as color, motion, form and depth. Afterward, information about these separate features is sent forward and distributed to several far-flung regions in the temporal and parietal lobes. In the case of color, the information goes to area V4 in the fusiform gyrus of the temporal lobe. From there it travels to areas that lie farther up in the hierarchy of color centers, including a region near a patch of cortex called the TPO (for the junction of the *t*emporal, *p*arietal and *o*ccipital lobes). These higher areas may be concerned with more sophisticated aspects of color processing. For example, leaves look as green at dusk as they do at midday, even though the mix of wavelengths reflected from the leaves is very different.

Numerical computation, too, seems to happen in stages. An early step also takes place in the fusiform gyrus, where the actual shapes of numbers are represented, and a later one occurs in the angular gyrus, a part of the TPO that is concerned with numerical concepts such as ordinality (sequence) and cardinality (quantity). (When the angular gyrus is damaged by a stroke or a tumor, the patient can still identify numbers but can no longer divide or subtract. Multiplication often survives because it is learned by rote.) In addition, brain-imaging studies in humans strongly hint that visually presented letters of the alphabet or numbers (graphemes) activate cells in the fusiform gyrus, whereas the sounds of the syllables (phonemes) are processed higher up, once again in the general vicinity of the TPO.

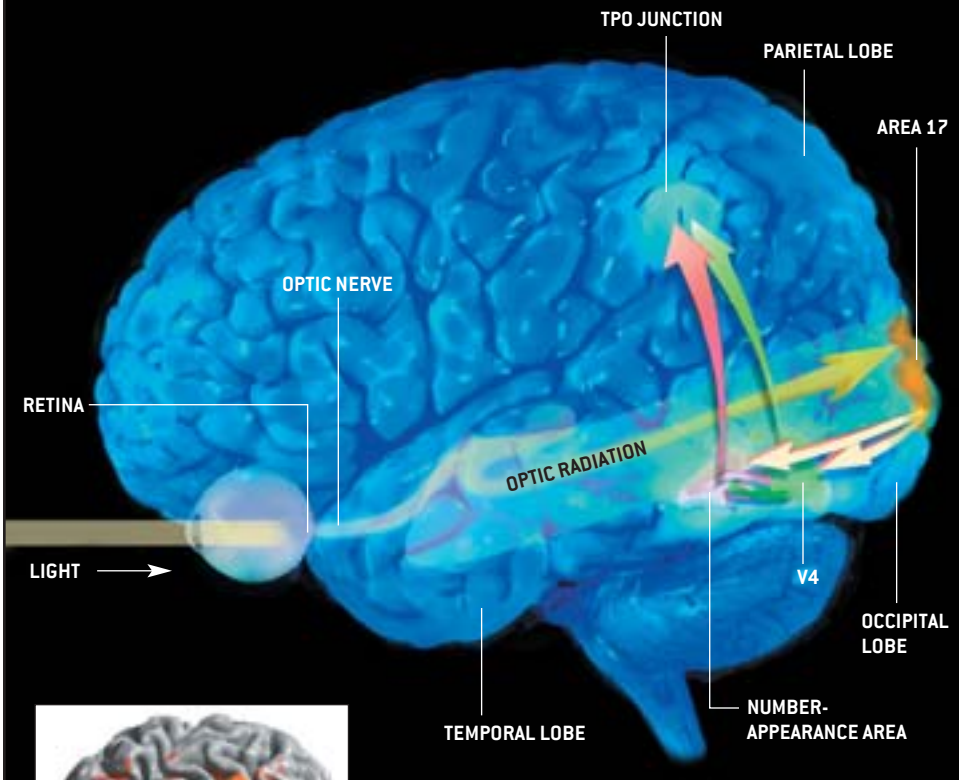
Because both colors and numbers are processed initially in the fusiform gyrus and subsequently near the angular gyrus, we suspected that number-color synesthesia might be caused by cross wiring between V4 and the number-appearance area (both within the fusiform) or between the higher color area and the number-concept area (both in the TPO). Other, more exotic forms of the condition might result from similar cross wiring of different sensory-processing regions. That the hearing center in the temporal lobes is also close to the higher brain area that receives color signals from V4 could explain sound-color synesthesia. Similarly, Matthew Blakeslee's tasting of touch might occur because of cross wiring be-

Overview/*Synesthesia*

- Synesthesia (from the Greek roots *syn,* meaning "together," and *aisthesis,* or "perception") is a condition in which otherwise normal people experience the blending of two or more senses.
- For decades, the phenomenon was often written off as fakery or simply memories, but it has recently been shown to be real. Perhaps it occurs because of cross activation, in which two normally separate areas of the brain elicit activity in each other.
- As scientists explore the mechanisms involved in synesthesia, they are also learning about how the brain in general processes sensory information and uses it to make abstract connections between seemingly unrelated inputs.

MINGLED SIGNALS

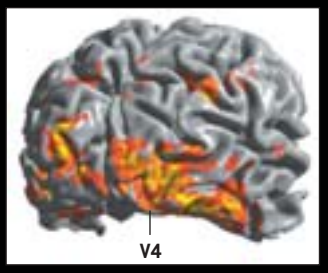
IN ONE OF THE MOST COMMON FORMS of synesthesia, looking at a number evokes a specific hue. This apparently occurs because brain areas that normally do not interact when processing numbers or colors do activate each other in synesthetes.



NEURAL SIGNALS from the retina travel via optic radiation to area 17, in the rear of the brain, where they are broken into simple attributes such as color, form, motion and depth.

Color information continues on to V4, near where the visual appearance of numbers is also represented—and thus is a site for cross-linking between the color and number areas (*short pink and green arrows*).

Ultimately, color proceeds "higher," to an area near the TPO (for temporal, *p*arietal, *o*ccipital lobes) junction, which may perform more sophisticated color processing. Similarly, a later stage of numerical computation occurs in the angular gyrus, a part of the TPO concerned with the concepts of sequence and quantity. This could explain synesthesia in people who link colors with abstract numerical sequences, like days of the week.



REAR VIEW of a synesthete's brain, made with functional magnetic resonance imaging, shows high activity (*yellow*) in the V4 color-processing area as the subject looks at white numbers on a gray background. This area is not active in people with normal perception viewing the same figures.

CAROL DONNER (*illustration*); GEOFF BOYNTON *Salk Institute for Biological Studies* AND VILAYANUR S. RAMACHANDRAN AND EDWARD M. HUBBARD (*inset*)

tween the taste cortex in a region called the insula and an adjacent cortex representing touch by the hands.

Assuming that neural cross wiring does lie at the root of synesthesia, why does it happen? We know that it runs in families, so it has a genetic component. Perhaps a mutation causes connections to emerge between brain areas that are usually segregated. Or maybe the mutation leads to defective pruning of preexisting connections between areas that are normally connected only sparsely. If the mutation were to be expressed (that is, to exert its effects) in some brain areas but not others, this patchiness might explain why some synesthetes conflate colors and numbers whereas others see colors when they hear phonemes or musical notes. People who have one type of synesthesia are more likely to have another, which adds weight to this idea.

Although we initially thought in terms of physical cross wiring, we have come to realize that the same effect could occur if the wiring—the number of connections between regions—was fine but the balance of chemicals traveling between regions was skewed. So we now speak in terms of cross activation. For instance, neighboring brain regions often inhibit one another's activity, which serves to minimize cross talk. A chemical imbalance of some kind that reduces such inhibition—for example, by blocking the action of an inhibitory neurotransmitter or failing to produce an inhibitor—would also cause activity in one area to elicit activity in a neighbor. Such cross activation could, in theory, also occur between widely separated areas, which would account for some of the less common forms of synesthesia.

Support for cross activation comes from other experiments, some of which also help to explain the varied forms synesthesia can take. One takes advantage of a visual phenomenon known as crowding [*see illustration on page 57*]. If you stare at a small plus sign in an image that also has a number 5 off to one side,

you will find that it is easy to discern that number, even though you are not looking at it directly. But if we now surround the 5 with four other numbers, such as 3's, then you can no longer identify it. It looks out of focus. Volunteers who perceive normally are no more successful at identifying this number than mere chance. That is not because things get fuzzy in the periphery of vision. After all, you could see the 5 perfectly clearly when it wasn't surrounded by 3's. You cannot identify it now because of limited attentional resources. The flanking 3's somehow distract your attention away from the central 5 and prevent you from seeing it.

A big surprise came when we gave the same test to two synesthetes. They looked at the display and made remarks like, "I cannot see the middle number. It's fuzzy but it looks red, so I guess it must be a 5." Even though the middle number did not consciously register, it seems that the brain was nonetheless processing it somewhere. Synesthetes could then use this color to deduce intellectually what the number was. If our theory is right, this finding implies that the number is processed in the fusiform gyrus and evokes the appropriate color *before* the stage at which the crowding effect occurs in the brain; paradoxically, the result is that even an "invisible" number can produce synesthesia.

Another finding we made also supports this conclusion. When we reduced the contrast between the number and the background, the synesthetic color became weaker until, at low contrast, subjects saw no color at all, even though the number was perfectly visible. Whereas the crowding experiment shows that an invisible number can elicit color, the contrast experiment conversely indicates that viewing a number does not guarantee seeing a color. Perhaps low-contrast numbers activate cells in the fusiform adequately for conscious perception of the number but not enough to cross-activate the color cells in V4.

Finally, we found that if we showed synesthetes Roman numerals, a V, say, they saw no color—which suggests that it is not the numerical *concept* of a number, in this case 5, but the grapheme's visual appearance that drives the color. This observation, too, implicates cross activation within the fusiform gyrus itself in number-color synesthesia, because that structure is mainly involved in analyzing the visual shape, not the high-level meaning of the number. One intriguing twist: Imagine an image with a large 5 made up of little 3's; you can see either the "forest" (the 5) or focus minutely on the "trees" (the 3's). Two synesthete subjects reported that they saw the color switch, depending on their focus. This test implies that even

DAVID EMMITE

THE AUTHORS

VILAYANUR S. RAMACHANDRAN and *EDWARD M. HUBBARD* collaborate on studies of synesthesia. Ramachandran directs the Center for Brain and Cognition at the University of California at San Diego and is adjunct professor at the Salk Institute for Biological Studies. He trained as a physician and later obtained a Ph.D. from Trinity College, University of Cambridge. He has received a fellowship from All Souls College, University of Oxford, the Ariens Kappers Gold Medal from the Royal Netherlands Academy, and the plenary lecture award from the American Academy of Neurology. He gave the BBC Reith Lectures for 2003. This is his fourth article for *Scientific American.* Hubbard is a fourth-year graduate student in the departments of psychology and cognitive science at U.C.S.D. His research combines psychophysics and functional magnetic resonance imaging to explore the neural basis of multisensory phenomena. A founding member of the American Synesthesia Association, he helped to organize its second annual meeting at U.C.S.D. in 2001.

though synesthesia can arise as a result of the visual appearance alone—not the high-level concept—the manner in which the visual input is categorized, based on attention, is also critical.

But as we began to recruit other volunteers, it soon became obvious that not all synesthetes who colorize their world are alike. In some, even days of the week or months of the year elicit colors. Monday might be green, Wednesday pink, and December yellow.

The only thing that days of the week, months and numbers have in common is the concept of numerical sequence, or ordinality. For certain synesthetes, perhaps it is the abstract concept of numerical sequence that drives the color, rather than the visual appearance of the number. Could it be that in these individuals, the cross wiring occurs between the angular gyrus and the higher color area near the TPO instead of between areas in the fusiform? If so, that interaction would explain why even abstract number representations, or the *idea* of the numbers elicited by days of the week or months, will strongly evoke specific colors. In other words, depending on where in the brain the mutant gene is expressed, it can result in different types of the condition—"higher" synesthesia, driven by numerical concept, or "lower" synesthesia, produced by visual appearance alone. Similarly, in some lower forms, the visual appearance of a letter might generate color, whereas in higher forms it is the *sound*, or phoneme, summoned by that letter; phonemes are represented near the TPO.

We also observed one case in which we believe cross activation enables a color-blind synesthete to see numbers tinged with hues he otherwise cannot perceive; charmingly, he refers to these as "Martian colors." Although his retinal color receptors cannot process certain wavelengths, we suggest that his brain color area is working just fine and being cross-activated when he sees numbers.

In brain-imaging experiments we are conducting with Geoff Boynton of the Salk Institute for Biological Studies in San Diego, we have obtained preliminary evidence of local activation of the color area V4 in a manner predicted by our cross-activation theory of synesthesia. (Jeffrey Gray of the Institute of Psychiatry in London and his colleagues have reported similar results.) On presenting black and white numbers to synesthetes, brain activation arose not only in the number area—as it would in normal subjects—but also in the color area. Our group also observed differences between types of synesthetes. One of our subjects with lower synesthesia showed much greater activation in earlier stages of color processing than occurred in controls. In contrast, higher synesthetes show less activation at these earlier levels.

A Way with Metaphor

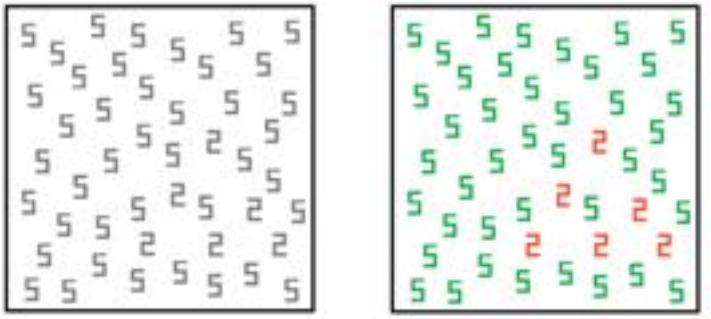
OUR INSIGHTS into the neurological basis of synesthesia could help explain some of the creativity of painters, poets and novelists. According to one study, the condition is seven times as common in creative people as in the general population.

One skill that many creative people share is a facility for using metaphor ("It is the east, and Juliet is the sun"). It is as if their brains are set up to make links between seemingly unrelated domains—such as the sun and a beautiful young woman. In other words, just as synesthesia involves making arbitrary links between seemingly unrelated perceptual entities such as colors and numbers, metaphor involves making links between seemingly unrelated conceptual realms. Perhaps this is not just a coincidence.

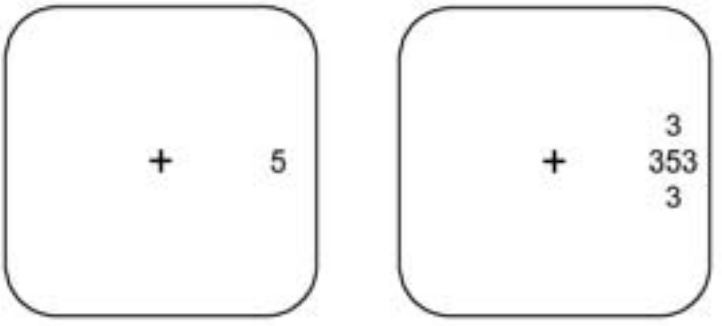
Numerous high-level concepts are probably anchored in specific brain regions, or maps. If you think about it, there is nothing more abstract than a number,

COLOR-CODED WORLD

IN A TEST of visual-segregation capabilities, synesthetes who link a specific hue with a given number can instantly see an embedded pattern in an image with black numbers scattered on a white page. Whereas a person with normal perception must undertake a digit-by-digit search to pick out, in this example, 2's amid 5's (*left*), the triangle-shaped group of 2's pops out for a synesthete (*right*).



"INVISIBLE" NUMBERS show up for synesthetes in a perceptual test. When a person stares at a central object, here a plus sign, a single digit off to one side is easy to see with peripheral vision (*left*). But if the number is surrounded by others (*right*), it appears blurry—invisible—to the average person. In contrast, a synesthete could deduce the central number by the color it evokes.



VILAYANUR S. RAMACHANDRAN

COMMON QUESTIONS

Are there different types of synesthesia?
Science counts about 50. The condition runs in families and may be more common in women and creative people; perhaps one person in 200 has synesthesia. In the most prevalent type, looking at numbers or listening to tones evokes colors. In one rare kind, each letter is associated with the male or female sex—an example of the brain's tendency to split the world into binary categories.

If a synesthete associates a color with a single letter or number, what happens if he looks at a pair of letters, such as "ea," or double digits, as in "25"?
He sees colors that correspond with the individual letters and numbers. If the letters or numbers are too close physically, however, they may cancel each other out (color disappears) or, if the two happen to elicit the same color, enhance each other.

Does it matter whether letters are uppercase or lowercase?
In general, no. But people have sometimes described seeing less saturated color in lowercase letters, or the lowercase letters may appear shiny or even patchy.

How do entire words look?
Often the color of the first letter spreads across the word; even silent letters, such as the "p" in "psalm," cause this effect.

What if the synesthete is multilingual?
One language can have colored graphemes, but a second (or additional others) may not, perhaps because separate tongues are represented in different brain regions.

What about when the person mentally pictures a letter or number?
Imagining can evoke a stronger color than looking at a real one. Perhaps that exercise activates the same brain areas as does viewing real colors—but because no competing signals from a real number are coming from the retina, the imagined one creates a stronger synesthetic color.

Does synesthesia improve memory?
It can. The late Russian neurologist Aleksandr R. Luria described a mnemonist who had remarkable recall because all five of his senses were linked. Even having two linked senses may help. *—V.S.R. and E.M.H.*

and yet it is represented, as we have seen, in a relatively small brain region, the angular gyrus. Let us say that the mutation we believe brings about synesthesia causes excess communication among different brain maps—small patches of cortex that represent specific perceptual entities, such as sharpness or curviness of shapes or, in the case of color maps, hues. Depending on where and how widely in the brain the trait was expressed, it could lead to both synesthesia and to a propensity toward linking seemingly unrelated concepts and ideas—in short, creativity. This would explain why the apparently useless synesthesia gene has survived in the population.

In addition to clarifying why artists might be prone to experiencing synesthesia, our research suggests that we all have some capacity for it and that this trait may have set the stage for the evolution of abstraction—an ability at which humans excel. The TPO (and the angular gyrus within it), which plays a part in the condition, is normally involved in cross-modal synthesis. It is the brain region where information from touch, hearing and vision is thought to flow together to enable the construction of high-level perceptions. For example, a cat is fluffy (touch), it meows and purrs (hearing), it has a certain appearance (vision) and odor (smell), all of which are derived simultaneously by the memory of a cat or the sound of the word "cat."

Could it be that the angular gyrus—which is disproportionately larger in humans compared with that in apes and monkeys—evolved originally for cross-modal associations but then became co-opted for other, more abstract functions such as metaphors? Consider two drawings, originally designed by psychologist Wolfgang Köhler. One looks like an inkblot and the other, a jagged piece of shattered glass. When we ask, "Which of these is a 'bouba,' and which is a 'kiki'?" 98 percent of people pick the inkblot as a bouba and the other one as a kiki. Perhaps that is because the gentle curves of the amoeba-like figure metaphorically mimic the gentle undulations of the sound "bouba" as represented in the hearing centers in the brain as well as the gradual inflection of the lips as they produce the curved "boo-baa" sound. In contrast, the waveform of the sound "kiki" and the sharp inflection of the tongue on the palate mimic the sudden changes in the jagged visual shape. The only thing these two kiki features have in common is the abstract property of jaggedness that is extracted somewhere in the vicinity of the TPO, probably in the angular gyrus. (We recently found that people with damage to the angular gyrus lose the bouba-kiki effect—they cannot match the shape with the correct sound.) In a sense, perhaps we are all closet synesthetes.

So the angular gyrus performs a very elementary type of abstraction—extracting the common denominator from a set of strikingly dissimilar entities. We do

Synesthesia may provide some insights about the evolution of thought and language

IMAGINE A BAND of ancestral hominids about to invent language. Clearly, they did not begin by having a leader say, "Hey, look at this—let's call it a banana. All of you say after me, *ba-na-na*." Undoubtedly, though, the group had a set of capacities that prepared the ground for systematic verbal communication. Our studies of the neurobiological basis of synesthesia suggest that a facility for metaphor—for seeing deep links between superficially dissimilar and unrelated things—provided a key seed for the eventual emergence of language.

Humans have a built-in bias to associate certain sounds with particular visual shapes, which could well have been important in getting hominids started on a shared vocabulary. In addition, specific brain areas that process visual shapes of objects, letters and numbers, and word sounds can activate each other even in nonsynesthetes, causing people to expect, say, jagged shapes to have harsh-sounding names.

Two other types of neural connections support our idea. First, the sensory areas for visual shapes and for hearing in the back of the brain can cross-activate specific motor areas in the front of the brain that participate in speech. A sharp visual inflection or a harsh sound induces the motor control area for speech to

produce an equally sudden inflection of the tongue on the palate. (Or consider the spoken words "diminutive," "teeny-weeny" and "*un peu,*" which involve pursing the lips to mimic the small size of the object.) The brain seems to possess preexisting rules for translating what we see and hear into mouth motions that reflect those inputs.

Second, a kind of spillover of signals occurs between two nearby motor areas: those that control the sequence of muscle movements required for hand gestures and those for the mouth. We call this effect "synkinesia." As Charles Darwin pointed out, when we cut paper with scissors, our jaws may clench and unclench unconsciously as if to echo the hand movements. Many linguists do not like the theory that manual gesturing could have set the stage for vocal language, but we believe that synkinesia suggests that they may be wrong.

Assume that our ancestral hominids communicated mainly through emotional grunts, groans, howls and shrieks, which are known to be produced by the right hemisphere and an area in the frontal lobes concerned with emotion. Later the hominids developed a rudimentary gestural system that became gradually more elaborate and sophisticated; it is easy to imagine how the hand movement for pulling someone toward you might have progressed to a "come hither" wave. If such gestures were translated through synkinesia into movements of the mouth and face muscles, and if emotional guttural utterances were channeled through these mouth and tongue movements, the result could have been the first spoken words.

How would we import syntax, the rules for using words and phrases in language, into this scheme? We believe that the evolution of tool use by hominids may have played an important role. For example, the tool-building sequence—first shape the hammer's head, then attach it to a handle, then chop the meat—resembles the embedding of clauses within larger sentences. Following the lead of psychologist Patricia Greenfield of the University of California at Los Angeles, we propose that frontal brain areas that evolved for subassembly in tool use may later have been co-opted for a completely novel function—joining words into phrases and sentences.

Not every subtle feature of modern language is explained by such schemes, but we suspect that these elements were critical for setting in motion the events that culminated in modern language. —*V.S.R. and E.M.H.*

IF ASKED which of the two figures below is a "bouba" and which is a "kiki," 98 percent of all respondents choose the blob as a bouba and the other as a kiki. The authors argue that the brain's ability to pick out an abstract feature in common—such as a jagged visual shape and a harsh-sounding name—could have paved the way for the development of metaphor and perhaps even a shared vocabulary.

not know how exactly it does this job. But once the ability to engage in cross-modal abstraction emerged, it might have paved the way for the more complex types of abstraction. The opportunistic takeover of one function for a different one is common in evolution. For example, bones in the ear used for hearing in mammals evolved from the back of the jawbone in reptiles. Beyond metaphor and abstract thinking, cross-modal abstraction might even have provided seeds for language [*see box above*].

When we began our research on synesthesia, we had no inkling of where it would take us. Little did we suspect that this eerie phenomenon, long regarded as a mere curiosity, might offer a window into the nature of thought. SA

VILAYANUR S. RAMACHANDRAN

MORE TO EXPLORE

The Man Who Tasted Shapes. R. E. Cytowic. MIT Press, 1993.

Synaesthesia: Classic and Contemporary Readings. S. Baron-Cohen and J. E. Harrison. Blackwell, 1997.

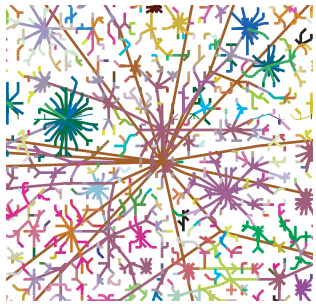
Psychophysical Investigations into the Neural Basis of Synaesthesia. V. S. Ramachandran and E. M. Hubbard in *Proceedings of the Royal Society of London,* B, Vol. 268, pages 979–983; 2001.

Synaesthesia: A Window into Perception, Thought and Language. V. S. Ramachandran and E. M. Hubbard in *Journal of Consciousness Studies,* Vol. 8, No. 12, pages 3–34; 2001.

Synaesthetic Photisms Influence Visual Perception. D. Smilek, M. J. Dixon, C. Cudahy and M. Merikle in *Journal of Cognitive Neuroscience,* Vol. 13, No. 7, pages 930–936; 2001.

Functional Magnetic Resonance Imaging of Synesthesia: Activation of V4/V8 by Spoken Words. J. A. Nunn, L. J. Gregory, M. Brammer, S.C.R. Williams, D. M. Parslow, M. J. Morgan, R. G. Morris, E. T. Bullmore, S. Baron-Cohen and J. A. Gray in *Nature Neuroscience,* Vol. 5, pages 371–375; 2002.

For more on synesthesia, visit **www.sciam.com/ontheweb**

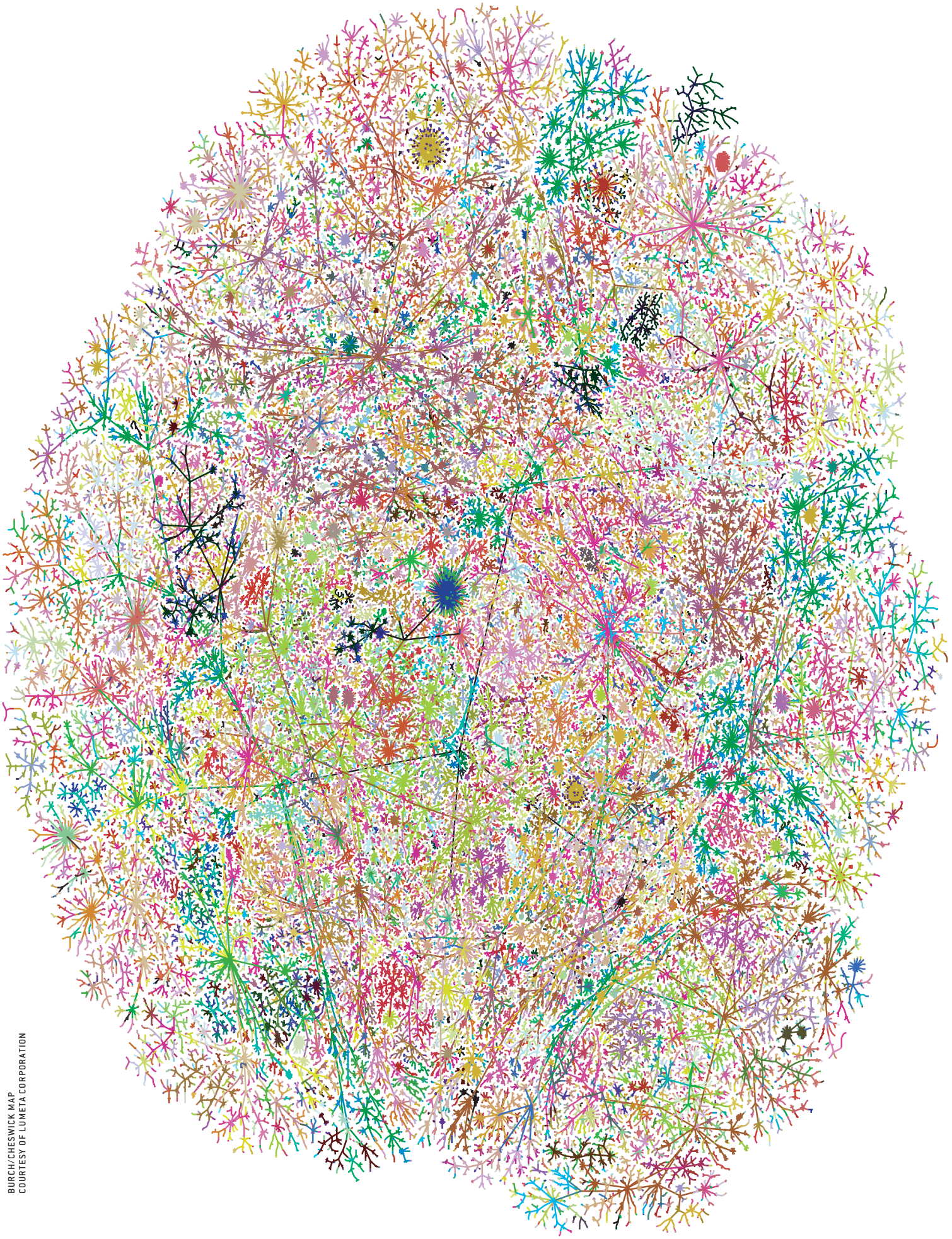


THE INTERNET, mapped on the opposite page, is a scale-free network in that some sites (*starbursts and detail above*) have a seemingly unlimited number of connections to other sites. This map, made on February 6, 2003, traces the shortest routes from a test Web site to about 100,000 others, using like colors for similar Web addresses.

Scale-Free Networks

Scientists have recently discovered that various complex systems have an underlying architecture governed by shared organizing principles. This insight has important implications for a host of applications, from drug development to Internet security

BY ALBERT-LÁSZLÓ BARABÁSI AND ERIC BONABEAU



BURCH/CHESWICK MAP
COURTESY OF LUMETA CORPORATION

Networks are everywhere.

The brain is a network of nerve cells connected by axons, and cells themselves are networks of molecules connected by biochemical reactions. Societies, too, are networks of people linked by friendships, familial relationships and professional ties. On a larger scale, food webs and ecosystems can be represented as networks of species. And networks pervade technology: the Internet, power grids and transportation systems are but a few examples. Even the language we are using to convey these thoughts to you is a network, made up of words connected by syntactic relationships.

Yet despite the importance and pervasiveness of networks, scientists have had little understanding of their structure and properties. How do the interactions of several malfunctioning nodes in a complex genetic network result in cancer? How does diffusion occur so rapidly in certain social and communications systems, leading to epidemics of diseases and computer viruses? How do some networks continue to function even after the vast majority of their nodes have failed?

Recent research has begun to answer such questions. Over the past few years, investigators from a variety of fields have discovered that many networks—from the World Wide Web to a cell's metabolic system to actors in Hollywood—are dominated by a relatively small number of nodes that are connected to many other sites. Networks containing such important nodes, or hubs, tend to be what we call "scale-free," in the sense that some hubs have a seemingly unlimited number of links and no node is typical of the others. These networks also behave in certain predictable ways; for example, they are remarkably resistant to accidental failures but extremely vulnerable to coordinated attacks.

Such discoveries have dramatically changed what we thought we knew about the complex interconnected world around us. Unexplained by previous network theories, hubs offer convincing proof that various complex systems have a strict architecture, ruled by fundamental laws—laws that appear to apply equally to cells, computers, languages and society. Furthermore, these organizing principles have significant implications for developing better drugs, defending the Internet from hackers, and halting the spread of deadly epidemics, among other applications.

Networks without Scale

FOR MORE THAN 40 YEARS, science treated all complex networks as being completely random. This paradigm has its roots in the work of two Hungarian mathematicians, the inimitable Paul Erdős and his close collaborator Alfréd Rényi. In 1959, aiming to describe networks seen in communications and the life sciences, Erdős and Rényi suggested that such systems could be effectively modeled by connecting their nodes with randomly placed links. The simplicity of their approach and the elegance of some of their related theorems revitalized graph theory, leading to the emergence of a field in mathematics that focuses on random networks.

An important prediction of random-network theory is that, despite the random placement of links, the resulting system will be deeply democratic: most nodes will have approximately the same number of links. Indeed, in a random network the nodes follow a Poisson distribution with a bell shape, and it is extremely rare to find nodes that have significantly more or fewer links than the average. Random networks are also called exponential, because the probability that a node is connected to k other sites decreases exponentially for large k.

So in 1998, when we, together with Hawoong Jeong and Réka Albert of the University of Notre Dame, embarked on a project to map the World Wide Web, we expected to find a random network. Here's why: people follow their unique interests when deciding what sites to link their Web documents to, and given the diversity of everyone's interests and the tremendous number of pages they can choose from, the resulting pattern of connections should appear fairly random.

The measurements, however, defied that expectation. Software designed for this project hopped from one Web page to another and collected all the links it could. Although this virtual robot reached only a tiny fraction of the entire Web, the map it assembled revealed something quite surprising: a few highly connected pages are essentially holding the World Wide Web together. More than 80 percent of the pages on the map had fewer than four links, but a small minority, less than 0.01 percent of all nodes, had more than 1,000. (A subsequent Web survey would uncover one document that had been referenced by more than two million other pages!)

Counting how many Web pages have exactly k links showed that the distribution followed a so-called power law: the probability that any node was connected to k other nodes was proportional to $1/k^n$. The value of n for incoming links was approximately 2, so, for instance, any node was roughly four times as likely to have just half the number of incoming links as another node. Power laws are quite different from the bell-shaped distributions that characterize random networks.

Overview/*Scale-Free Networks*

- A variety of complex systems share an important property: some nodes have a tremendous number of connections to other nodes, whereas most nodes have just a handful. The popular nodes, called hubs, can have hundreds, thousands or even millions of links. In this sense, the network appears to have no scale.
- Scale-free networks have certain important characteristics. They are, for instance, robust against accidental failures but vulnerable to coordinated attacks.
- Understanding of such characteristics could lead to new applications in many arenas. For example, computer scientists might be able to devise more effective strategies for preventing computer viruses from crippling a network such as the Internet.

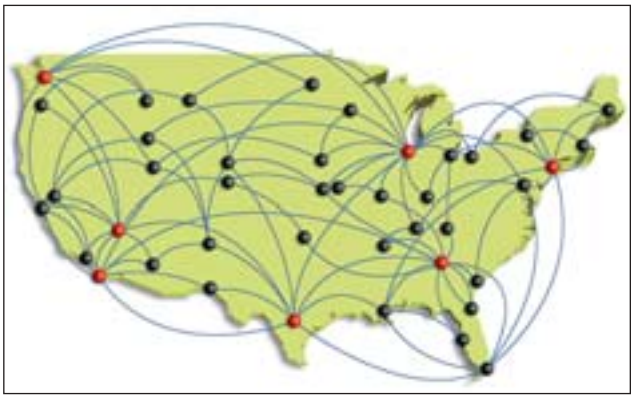
RANDOM VERSUS SCALE-FREE NETWORKS

RANDOM NETWORKS, which resemble the U.S. highway system (*simplified in left map*), consist of nodes with randomly placed connections. In such systems, a plot of the distribution of node linkages will follow a bell-shaped curve (*left graph*), with most nodes having approximately the same number of links.

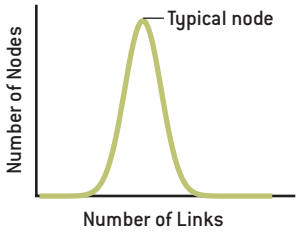
In contrast, scale-free networks, which resemble the U.S. airline system (*simplified in right map*), contain hubs (*red*)—nodes with a very high number of links. In such networks, the distribution of node linkages follows a power law (*center graph*) in that most nodes have just a few connections and some have a tremendous number of links. In that sense, the system has no "scale." The defining characteristic of such networks is that the distribution of links, if plotted on a double-logarithmic scale (*right graph*), results in a straight line.

Random Network

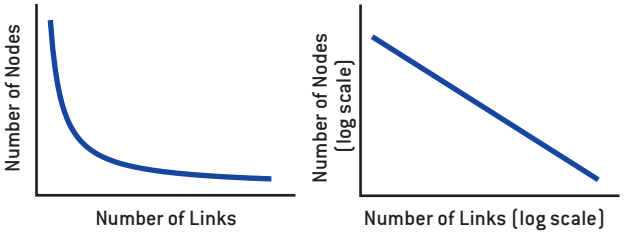
Scale-Free Network



Bell Curve Distribution of Node Linkages



Power Law Distribution of Node Linkages



Specifically, a power law does not have a peak, as a bell curve does, but is instead described by a continuously decreasing function. When plotted on a double-logarithmic scale, a power law is a straight line [*see illustration above*]. In contrast to the democratic distribution of links seen in random networks, power laws describe systems in which a few hubs, such as Yahoo and Google, dominate.

Hubs are simply forbidden in random networks. When we began to map the Web, we expected the nodes to follow a bell-shaped distribution, as do people's heights. Instead we discovered certain nodes that defied explanation, almost as if we had stumbled on a significant number of people who were 100 feet tall, thus prompting us to coin the term "scale-free."

Scale-Free Networks Abound

OVER THE PAST several years, researchers have uncovered scale-free structures in a stunning range of systems. When we studied the World Wide Web, we looked at the *virtual* network of Web pages connected to one another by hyperlinks. In contrast, Michalis Faloutsos of the University of California at Riverside, Petros Faloutsos of the University of Toronto and Christos Faloutsos of Carnegie Mellon University analyzed the *physical* structure of the Internet. These three computer-scientist brothers investigated the routers connected by optical or other communications lines and found that the topology of that network, too, is scale-free.

Researchers have also discovered that some social networks are scale-free. A collaboration between scientists from Boston University and Stockholm University, for instance, has shown that a network of sexual relationships among people in Sweden followed a power law: although most individuals had only a few sexual partners during their lifetime, a few (the hubs) had hundreds. A recent study led by Stefan Bornholdt of the University of Kiel in Germany concluded that the network of people connected by e-mail is likewise scale-free. Sidney Redner of Boston University demonstrated that the network of scientific papers, connected by citations, follows a power law as well. And Mark Newman of the University of Michigan at Ann Arbor examined collaborations among scientists in several

SLIM FILMS

disciplines, including physicians and computer scientists, and found that those networks were also scale-free, corroborating a study we conducted focusing on mathematicians and neurologists. (Interestingly, one of the largest hubs in the mathematics community is Erdős himself, who wrote more than 1,400 papers with no fewer than 500 co-authors.)

Scale-free networks can occur in business. Walter W. Powell of Stanford University, Douglas R. White of the University of California at Irvine, Kenneth W. Koput of the University of Arizona, and Jason-Owen Smith of the University of Michigan studied the formation of alliance networks in the U.S. biotechnology industry and discovered definite hubs—for instance, companies such as Genzyme, Chiron and Genentech had a disproportionately large number of partnerships with other firms. Researchers in Italy took a deeper look at that network. Using data collected by the University of Siena's Pharmaceutical Industry Database, which now provides information for around 20,100 R&D agreements among more than 7,200 organizations, they found that the hubs detected by Powell and his colleagues were actually part of a scale-free network.

Even the network of actors in Hollywood—popularized by the game Six Degrees of Kevin Bacon, in which players try to connect actors to Bacon via the movies in which they have appeared together—is scale-free. A quantitative analysis of that network showed that it, too, is dominated by hubs. Specifically, although most actors have only a few links to others, a handful of actors, including Rod Steiger and Donald Pleasence, have thousands of connections. (Incidentally, on a list of most connected actors, Bacon ranked just 876th.)

On a more serious note, scale-free networks are present in the biological realm. With Zoltán Oltvai, a cell biologist from Northwestern University, we found a scale-free structure in the cellular metabolic networks of 43 different organisms from all three domains of life, including *Archaeoglobus fulgidus* (an archaebacterium), *Escherichia coli* (a eubacterium) and *Caenorhabditis elegans* (a eukaryote). In such networks, cells burn food by splitting complex molecules to release energy. Each node is a particular molecule, and each link is a biochemical reaction. We found that most molecules participate in just one or two reactions, but a few (the hubs), such as water and adenosine triphosphate, play a role in most of them.

We discovered that the protein-interaction network of cells is scale-free as well. In such a network, two proteins are "connected" if they are known to interact with each other. When we investigated Baker's yeast, one of the simplest eukaryotic (nucleus-containing) cells, with thousands of proteins, we discovered a scale-free topology: although most proteins interact with only one or two others, a few are able to attach themselves physically to a huge number. We found a similar result in the protein-interaction network of an organism that is very different from yeast, a simple bacterium called *Helicobacter pylori.*

Indeed, the more that scientists studied networks, the more they uncovered scale-free structures. These findings raised an important question: How can systems as fundamentally different as the cell and the Internet have the same architecture and obey the same laws? Not only are these various networks scale-free, they also share an intriguing property: for reasons not yet known, the value of n in the k^n term of the power law tends to fall between 2 and 3.

The Rich Get Richer

PERHAPS A MORE BASIC question is why random-network theory fails to explain the existence of hubs. A closer examination of the work of Erdős and Rényi reveals two reasons.

In developing their model, Erdős and Rényi assumed that they had the full inventory of nodes before they placed the links. In contrast, the number of documents on the Web is anything but constant. In 1990 the Web had only one page. Now it has more than three billion. Most networks have expanded similarly. Hollywood had only a handful of actors in 1890, but as new people joined the trade, the network grew to include more than half a million, with the rookies connecting to veteran actors. The Internet had only a few routers about three decades ago, but it gradually grew to have millions, with the new routers always linking to those that were already part of the network. Thanks to the growing nature of real networks, older nodes had greater opportunities to acquire links.

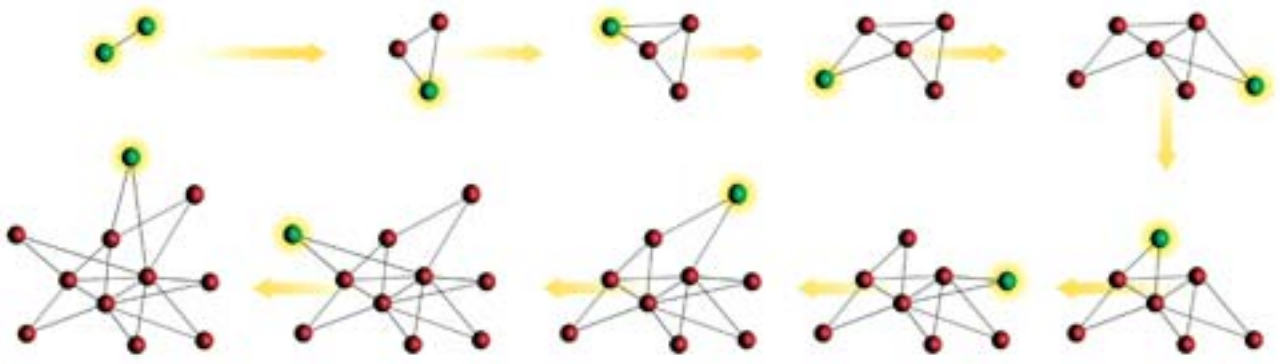
Furthermore, all nodes are not equal. When deciding where to link their Web page, people can choose from a few billion locations. Yet most of us are familiar with only a tiny fraction of the full Web, and that subset tends to include the more connected sites because they are easier to find. By simply linking to those nodes, people exercise and reinforce a bias toward them. This process of "preferential attachment" occurs elsewhere. In Hollywood the more

Examples of Scale-Free Networks

NETWORK	NODES	LINKS
Cellular metabolism	Molecules involved in burning food for energy	Participation in the same biochemical reaction
Hollywood	Actors	Appearance in the same movie
Internet	Routers	Optical and other physical connections
Protein regulatory network	Proteins that help to regulate a cell's activities	Interactions among proteins
Research collaborations	Scientists	Co-authorship of papers
Sexual relationships	People	Sexual contact
World Wide Web	Web pages	URLs

BIRTH OF A SCALE-FREE NETWORK

A SCALE-FREE NETWORK grows incrementally from two to 11 nodes in this example. When deciding where to establish a link, a new node (*green*) prefers to attach to an existing node (*red*) that already has many other connections. These two basic mechanisms—growth and preferential attachment—will eventually lead to the system's being dominated by hubs, nodes having an enormous number of links.



connected actors are more likely to be chosen for new roles. On the Internet the more connected routers, which typically have greater bandwidth, are more desirable for new users. In the U.S. biotech industry, well-established companies such as Genzyme tend to attract more alliances, which further increases their desirability for future partnerships. Likewise, the most cited articles in the scientific literature stimulate even more researchers to read and cite them, a phenomenon that noted sociologist Robert K. Merton called the Matthew effect, after a passage in the New Testament: "For unto every one that hath shall be given, and he shall have abundance."

These two mechanisms—growth and preferential attachment—help to explain the existence of hubs: as new nodes appear, they tend to connect to the more connected sites, and these popular locations thus acquire more links over time than their less connected neighbors. And this "rich get richer" process will generally favor the early nodes, which are more likely to eventually become hubs.

Along with Réka Albert, we have used computer simulations and calculations to show that a growing network with preferential attachment will indeed become scale-free, with its distribution of nodes following a power law. Although this theoretical model is simplistic and needs to be adapted to specific situations, it does appear to confirm our explanation for why scale-free networks are so ubiquitous in the real world.

Growth and preferential attachment can even help explicate the presence of scale-free networks in biological systems. Andreas Wagner of the University of New Mexico and David A. Fell of Oxford Brookes University in England have found, for instance, that the most-connected molecules in the *E. coli* metabolic network tend to have an early evolutionary history: some are believed to be remnants of the so-called RNA world (the evolutionary step before the emergence of DNA), and others are components of the most ancient metabolic pathways.

Interestingly, the mechanism of preferential attachment tends to be linear. In other words, a new node is twice as likely to link to an existing node that has twice as many connections as its neighbor. Redner and his colleagues at Boston University and elsewhere have investigated different types of preferential attachment and have learned that if the mechanism is faster than linear (for example, a new node is four times as likely to link to an existing node that has twice as many connections), one hub will tend to run away with the lion's share of connections. In such "winner take all" scenarios, the network eventually assumes a star topology with a central hub.

An Achilles' Heel

AS HUMANITY BECOMES increasingly dependent on power grids and communications webs, a much-voiced concern arises: Exactly how reliable are these types of networks? The good news is that complex systems can be amazingly resilient against accidental failures. In fact, although hundreds of routers routinely malfunction on the Internet at any moment, the network rarely suffers major disruptions. A similar degree of robustness characterizes living systems: people rarely notice the consequences of thousands of errors in their cells, ranging from mutations to misfolded proteins. What is the origin of this robustness?

Intuition tells us that the breakdown of a substantial number of nodes will result in a network's inevitable fragmenta-

THE AUTHORS

ALBERT-LÁSZLÓ BARABÁSI and *ERIC BONABEAU* study the behavior and characteristics of myriad complex systems, ranging from the Internet to insect colonies. Barabási is Emil T. Hofman Professor of Physics at the University of Notre Dame, where he directs research on complex networks. He is author of *Linked: The New Science of Networks.* Bonabeau is chief scientist at Icosystem, a consulting firm based in Cambridge, Mass., that applies the tools of complexity science to the discovery of business opportunities. He is co-author of *Swarm Intelligence: From Natural to Artificial Systems.* This is Bonabeau's second article for *Scientific American.*

SLIM FILMS

tion. This is certainly true for random networks: if a critical fraction of nodes is removed, these systems break into tiny, noncommunicating islands. Yet simulations of scale-free networks tell a different story: as many as 80 percent of randomly selected Internet routers can fail and the remaining ones will still form a compact cluster in which there will still be a path between any two nodes. It is equally difficult to disrupt a cell's protein-interaction network: our measurements indicate that even after a high level of random mutations are introduced, the unaffected proteins will continue to work together.

In general, scale-free networks display an amazing robustness against accidental failures, a property that is rooted in their inhomogeneous topology. The random removal of nodes will take out mainly the small ones because they are much more plentiful than hubs. And the elimination of small nodes will not disrupt the network topology significantly, because they contain few links compared with the hubs, which connect to nearly everything. But a reliance on hubs has a serious drawback: vulnerability to attacks.

In a series of simulations, we found that the removal of just a few key hubs from the Internet splintered the system into tiny groups of hopelessly isolated routers. Similarly, knockout experiments in yeast have shown that the removal of the more highly connected proteins has a significantly greater chance of killing the organism than does the deletion of other nodes. These hubs are crucial—if mutations make them dysfunctional, the cell will most likely die.

A reliance on hubs can be advantageous or not, depending on the system. Certainly, resistance to random breakdown is good news for both the Internet and the cell. In addition, the cell's reliance on hubs provides pharmaceutical researchers with new strategies for selecting drug targets, potentially leading to cures that would kill only harmful cells or bacteria by selectively targeting their hubs, while leaving healthy tissue unaffected. But the ability of a small group of well-informed hackers to crash the entire communications infrastructure by targeting its hubs is a major reason for concern.

The Achilles' heel of scale-free networks raises a compelling question: How many hubs are essential? Recent research suggests that, generally speaking, the simultaneous elimination of as few as 5 to 15 percent of all hubs can crash a system. For the Internet, our experiments imply that a highly coordinated attack—first removing the largest hub, then the next largest, and so on—could cause significant disruptions after the elimination of just several hubs. Therefore, protecting the hubs is perhaps the most effective way to avoid large-scale disruptions caused by malicious cyber-attacks. But much more work is required to determine just how fragile specific networks are. For instance, could the failure of several hubs like Genzyme and Genentech lead to the collapse of the entire U.S. biotech industry?

Scale-Free Epidemics

KNOWLEDGE ABOUT scale-free networks has implications for understanding the spread of computer viruses, diseases and fads. Diffusion theories, intensively studied for decades by both epidemiologists and marketing experts, predict a critical threshold for the propagation of a contagion throughout a population. Any virus, disease or fad that is less infectious than that well-defined threshold will inevitably die out, whereas those above the threshold will multiply exponentially, eventually penetrating the entire system.

Recently, though, Romualdo Pastor-Satorras of the Polytechnic University of Catalonia in Barcelona and Alessandro Vespigniani of the International Center for Theoretical Physics in Trieste, Italy, reached a disturbing conclusion. They found that in a scale-free network the threshold is zero. That is, all viruses, even those that are weakly contagious, will spread and persist in the system. This result explains why Love Bug, the most damaging computer virus thus far (it shut down the British Parliament in 2000), was still one of the most pervasive viruses a year after its supposed eradication.

Because hubs are connected to many other nodes, at least one hub will tend to be infected by any corrupted node. And once a hub has been infected, it will pass the virus to numerous other sites, eventually compromising other hubs, which will

The Potential Implications of Scale-Free Networks for...

Computing

- Computer networks with scale-free architectures, such as the World Wide Web, are highly resistant to accidental failures. But they are very vulnerable to deliberate attacks and sabotage.
- Eradicating viruses, even known ones, from the Internet will be effectively impossible.

Medicine

- Vaccination campaigns against serious viruses, such as smallpox, might be most effective if they concentrate on treating hubs—people who have many connections to others. But identifying such individuals can be difficult.
- Mapping out the networks within the human cell could aid researchers in uncovering and controlling the side effects of drugs. Furthermore, identifying the hub molecules involved in certain diseases could lead to new drugs that would target those hubs.

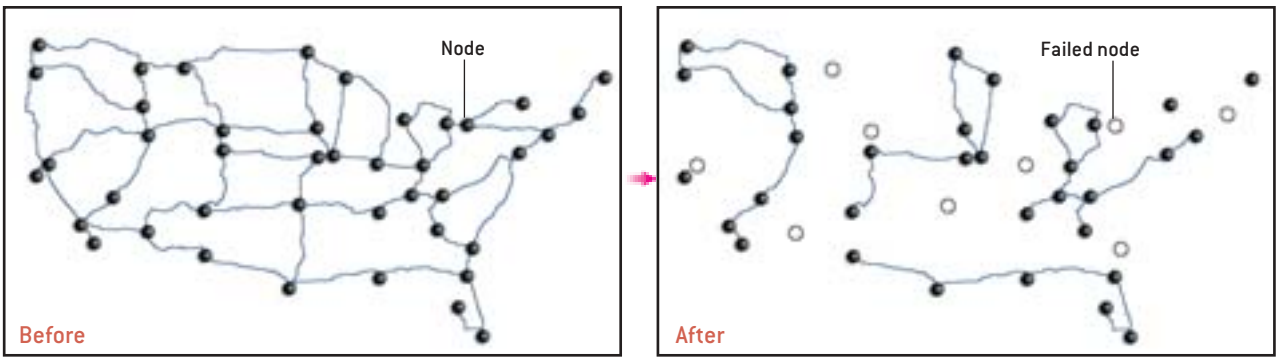
Business

- Understanding how companies, industries and economies are interlinked could help researchers monitor and avoid cascading financial failures.
- Studying the spread of a contagion on a scale-free network could offer new ways for marketers to propagate consumer buzz about their products.

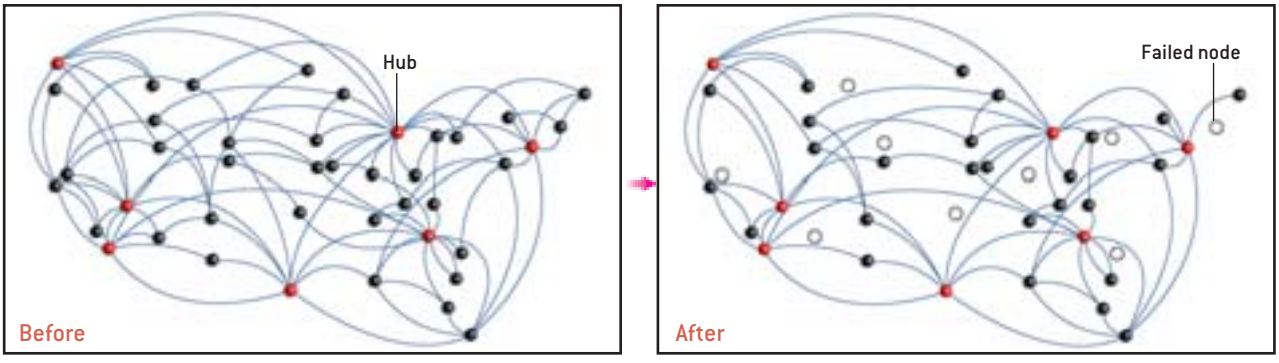
HOW ROBUST ARE RANDOM AND SCALE-FREE NETWORKS?

THE ACCIDENTAL FAILURE of a number of nodes in a random network (*top panels*) can fracture the system into non-communicating islands. In contrast, scale-free networks are more robust in the face of such failures (*middle panels*). But they are highly vulnerable to a coordinated attack against their hubs (*bottom panels*).

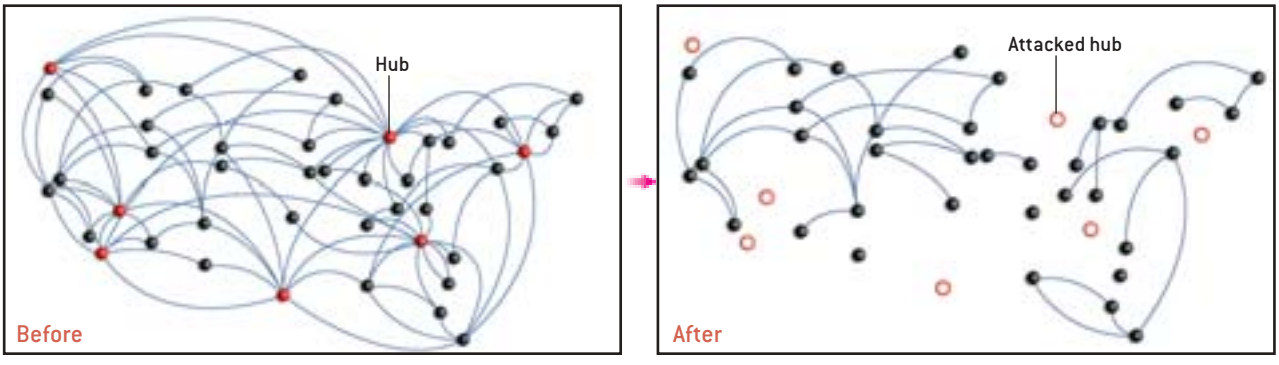
Random Network, Accidental Node Failure



Scale-Free Network, Accidental Node Failure



Scale-Free Network, Attack on Hubs



then spread the virus throughout the entire system.

The fact that biological viruses spread in social networks, which in many cases appear to be scale-free, suggests that scientists should take a second look at the volumes of research written on the interplay of network topology and epidemics. Specifically, in a scale-free network, the traditional public health approach of random immunization could easily fail because it would very likely neglect a number of the hubs. In fact, nearly everyone would have to be treated to ensure that the hubs were not missed. A vaccination for measles, for instance, must reach 90 percent of the population to be effective.

Instead of random immunizations, though, what if doctors targeted the hubs, or the most connected individuals? Research in scale-free networks indicates that this alternative approach could be effective even if the immunizations reached only a small fraction of the overall population, provided that the fraction contained the hubs.

But identifying the hubs in a social

SLIM FILMS

network is much more difficult than detecting them in other types of systems. Nevertheless, Reuven Cohen and Shlomo Havlin of Bar-Ilan University in Israel, together with Daniel ben-Avraham of Clarkson University, have proposed a clever solution: immunize a small fraction of the random acquaintances of arbitrarily selected individuals, a procedure that selects hubs with a high probability because they are linked to many people. That approach, though, leads to a number of ethical dilemmas. For instance, even if the hubs could be identified, should they have priority for immunizations and cures? Such issues notwithstanding, targeting hubs could be the most pragmatic solution for the future distribution of AIDS or smallpox vaccines in countries and regions that do not have the resources to treat everyone.

In many business contexts, people want to start, not stop, epidemics. Viral marketing campaigns, for instance, often specifically try to target hubs to speed the adoption of a product. Obviously, such a strategy is not new. Back in the 1950s, a study funded by pharmaceutical giant Pfizer discovered the important role that hubs play in how quickly a community of doctors begins using a new drug. Indeed, marketers have intuitively known for some time that certain customers outshine others in spreading promotional buzz about products and fads. But recent work in scale-free networks provides the scientific framework and mathematical tools to probe that phenomenon more rigorously.

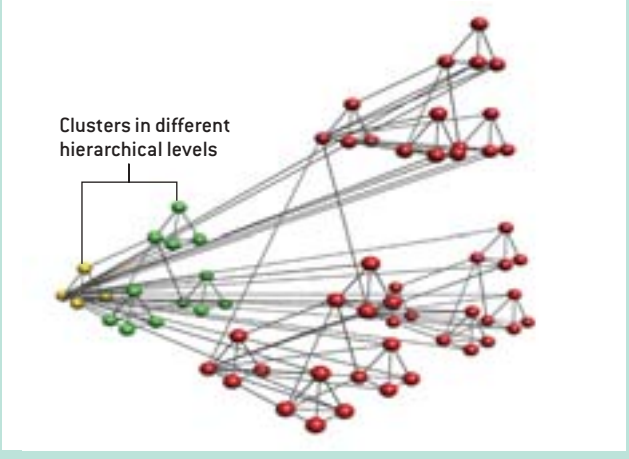
From Theory to Practice

ALTHOUGH SCALE-FREE networks are pervasive, numerous prominent exceptions exist. For example, the highway system and power grid in the U.S. are not scale-free. Neither are most networks

It's a Small World, After All

IN 1967 STANLEY MILGRAM, a social psychologist at Harvard University, sent hundreds of letters to people in Nebraska, asking them to forward the correspondence to acquaintances who might be able to shepherd it closer to a target recipient: a stockbroker in Boston. To track each of the different paths, Milgram asked the participants to mail a postcard back to him when they passed the letter to someone else. Milgram found that the letters that eventually arrived at the final destination had passed through an average of six individuals—the basis of the popular notion of "six degrees of separation" between everyone.

Although Milgram's work was hardly conclusive—most of the letters never made their way to the stockbroker—scientists have recently learned that other networks exhibit this "small world" property. We have found, for instance, that a path of just three reactions will connect almost any pair of chemicals in a cell. And on the World Wide Web, which contains more than three billion documents, Web pages are typically 19 clicks from one another.



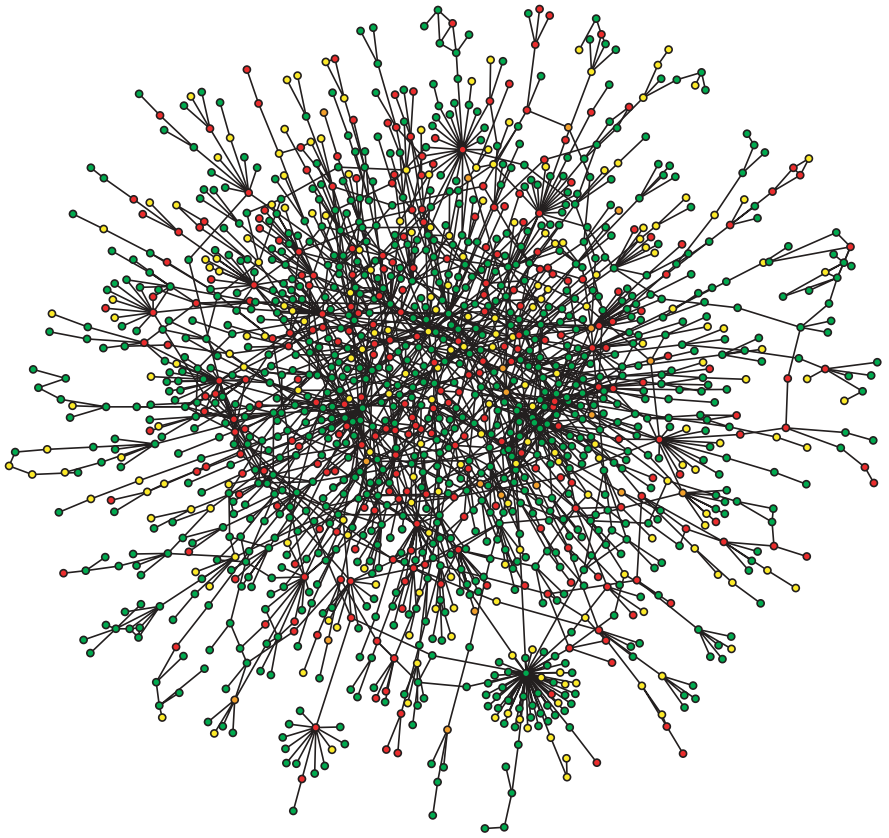
HIERARCHICAL CLUSTERS, shown schematically, could include, say, Web pages on the Frank Lloyd Wright home Fallingwater (*yellow*), which could be linked to other clusters (*green*) focusing on Wright, famous homes or Pennsylvania's attractions. Those sites, in turn, could be connected to clusters (*red*) on famous architects or architecture in general.

The small-world property does not necessarily indicate the presence of any magic organizing principle. Even a large network with purely random connections will be a small world. Consider that you might have about 1,000 acquaintances. If each of those individuals also knows another 1,000, then a million people will be just two handshakes away from you, a billion will be just three away, and the earth's entire population will be well within four. Given that fact, the notion that any two strangers in the world are separated by an average of six degrees seems almost trivial. But further investigation reveals some deeper insights.

Our simple calculation assumes that the people you know are all strangers to one another. In reality, there is much overlap. Indeed, society is fragmented into clusters of individuals having similar characteristics (such as income or interests), a feature that has been widely discussed in the sociology literature following the seminal work in the 1970s of Mark Granovetter, then a graduate student at Harvard. Clustering is also a general property of many other types of networks. In 1998 Duncan Watts and Steven Strogatz, then both at Cornell University, found significant clustering in a variety of systems, from the U.S. power grid to the neural network of the *Caenorhabditis elegans* worm.

At first glance, isolated clusters of highly interconnected nodes appear to run counter to the topology of scale-free networks, in which a number of hubs radiate throughout the system, linking everything. Recently, however, we have shown that the two properties are compatible: a network can be both highly clustered and scale-free when small, tightly interlinked clusters of nodes are connected into larger, less cohesive groups (*left*). This type of hierarchy appears to exist in a number of systems, from the World Wide Web (in which clusters are groupings of Web pages devoted to the same topic) to a cell (in which clusters are teams of molecules responsible for a specific function). *—A.-L.B. and E.B.*

SLIM FILMS



MAP OF INTERACTING PROTEINS in yeast highlights the discovery that highly linked, or hub, proteins tend to be crucial for a cell's survival. Red denotes essential proteins (their removal will cause the cell to die). Orange represents proteins of some importance (their removal will slow cell growth). Green and yellow represent proteins of lesser or unknown significance, respectively.

COURTESY OF HAWOONG JEONG

seen in materials science. In a crystal lattice, for instance, atoms have the same number of links to their neighbors. With other networks, the data are inconclusive. The relatively small size of food webs, which show predator-prey relationships, has prevented scientists from reaching a clear conclusion regarding that network's type. And the absence of large-scale connectivity maps of the brain has kept researchers from knowing the nature of that important network as well.

Determining whether a network is scale-free is important in understanding the system's behavior, but other significant parameters merit attention, too. One such characteristic is the diameter, or path length, of a network: the largest number of hops required to get from one node to another by following the shortest route possible [*see box on opposite page*].

Finally, knowledge of a network's general topology is just part of the story in understanding the overall characteristics and behavior of such systems. There might be steep costs, for instance, with the addition of each link to a given node that could prevent certain networks (such as the U.S. highway system) from becoming scale-free. In food chains, some prey are easier to catch than others, and that fact has a profound effect on the overall ecosystem. With social networks, ties among household members are much stronger than connections to casual acquaintances, so diseases (and information) are more likely to spread through such linkages. For transportation, transmission and communications systems (such as the Internet), congestion along specific links is a major consideration: too much traffic on a particular link can cause it to break down, leading to the potential failure of other links that must then handle the spillover. And the nodes themselves might not be homogeneous—certain Web pages have more interesting content, for instance—which could greatly alter the preferential-attachment mechanism.

Because of these and other factors, scientists have only begun to uncover the behavior of scale-free systems. Immunizing hubs, for instance, might not be sufficient to stop the spread of a disease; a more effective solution might be found by considering not just the number of connections a person has but also the frequency and duration of contact for those links.

In essence, we have studied complex networks first by ignoring the details of their individual links and nodes. By distancing ourselves from those particulars, we have been able to better glimpse some of the organizing principles behind these seemingly incomprehensible systems. At the very least, knowledge from this endeavor has led to the rethinking of many basic assumptions. In the past, for example, researchers modeled the Internet as a random network to test how a new routing protocol might affect system congestion. But we now know that the Internet is a scale-free system with behavior that is dramatically different from a random network's. Consequently, investigators such as John W. Byers and his colleagues at Boston University are revamping the computer models they have been using to simulate the Internet. Similarly, knowledge of the properties of scale-free networks will be valuable in a number of other fields, especially as we move beyond network topologies to probe the intricate and often subtle dynamics taking place within those complex systems. SA

MORE TO EXPLORE

All the World's a Net. David Cohen in *New Scientist,* Vol. 174, No. 2338, pages 24–29; April 13, 2002.

Statistical Mechanics of Complex Networks. Réka Albert and Albert-László Barabási in *Reviews of Modern Physics,* Vol. 74, pages 47–97; January 2002.

Linked: The New Science of Networks. Albert-László Barabási. Perseus Publishing, 2002.

Evolution of Networks: From Biological Nets to the Internet and WWW. J.F.F. Mendes and Sergei N. Dorogovtsev. Oxford University Press, 2003.

Find links to papers on scale-free networks at **www.nd.edu/~networks**

The Iceman Reconsidered

By James H. Dickson, Klaus Oeggl and Linda L. Handley

On a clear day in September 1991 a couple hiking along a high ridge in the Alps came upon a corpse melting out of the ice. When they returned to the mountain hut where they were staying, they alerted the authorities, who assumed the body was one of the missing climbers lost every year in the crevasses that crisscross the glaciers of the region. But after the remains were delivered to nearby Innsbruck, Austria, Konrad Spindler, an archaeologist from the university there, ascertained that the corpse was prehistoric. The victim, a male, had died several thousand years ago. Spindler and other scientists deduced that his body and belongings had been preserved in the ice until a fall of dust from the Sahara and an unusually warm spell combined to melt the ice, exposing his head, back and shoulders.

No well-preserved bodies had ever been

LANDESGENDARMERIEKOMMANDO FÜR TIROL/AUSTRIA

THE ICEMAN was discovered in a rocky hollow high in the Alps, in the zone of perennial snow and ice. Pressure from the overlying ice had removed a piece of the scalp. His corpse lay draped over a boulder. Contrary to earlier assumptions, evidence indicates it had floated into that position during previous thaws.

Where was the Iceman's home and what was he doing at the high mountain pass where he died? Painstaking research—especially of plant remains found with the body—contradicts many of the initial speculations

Ötzi had been WARMLY DRESSED in leggings, loincloth and jacket made of the hide of deer and goat, and a cape made of grass and bast.

found in Europe from this period, the Neolithic, or New Stone Age. The Iceman is much older than the Iron Age men from the Danish peat bogs and older even than the Egyptian royal mummies. Almost as astounding was the presence of a complete set of clothes and a variety of gear.

In the ensuing excitement over the discovery, the press and researchers offered many speculations about the ancient man. Spindler hypothesized an elaborate disaster theory. He proposed that the man had fled to safety in the mountains after being injured in a fight at his home village. It was autumn, Spindler went on, and the man was a shepherd who sought refuge in the high pastures where he took his herds in summer. Hurt and in a state of exhaustion, he fell asleep and died on the boulder on which he was found five millennia later. The beautiful preservation of the body, according to this account, was the result of a fall of snow that protected the corpse from scavengers, followed by rapid freeze-drying.

Because the uniqueness of the discovery had not been immediately evident, the corpse was torn from the ice in a way that destroyed much archaeological information and damaged the body itself. A more thorough archaeological excavation of the site took place in the summer of 1992 and produced much valuable evidence, including an abundance of organic material (seeds, leaves, wood, mosses). This material added greatly to the plant remains, especially mosses, already washed from the clothes during the conservation process. Now, after a decade of labor-intensive research by us and other scientists on these plant remains and on samples taken from the Iceman's intestines, some hard facts are revising those first, sketchily formed impressions and replacing them with a more substantiated story.

Who Was He?

THE HIKERS HAD DISCOVERED the body at 3,210 meters above sea level in the Ötztal Alps, which led to the popular humanizing nickname Ötzi. A mere 92 meters south of the Austrian-Italian border, the shallow, rocky hollow that sheltered the body is near the pass called Hauslabjoch between Italy's Schnalstal (Val Senales in Italian) and the Ventertal in Austria [*see map on opposite page*]. Ötzi lay in an awkward position, draped prone over a boulder, his left arm sticking out to the right, and his right hand trapped under a large stone. His gear and clothing, also frozen or partially frozen in the ice, were scattered around him, some items as far as several meters away. Radiocarbon dates from three different laboratories made both on plant remains found with the body and on samples of Ötzi's tissues and gear all confirm that he lived about 5,300 years ago.

Certain other features of Ötzi were relatively easy to discover as well. At 159 centimeters (5′2.5″), he was a small man, as many men in the Schnalstal vicinity are today. Bone studies show he was 46 years old, an advanced age for people of his time. DNA analysis indicates his origin in central-northern Europe, which may seem obvious, but it differentiates him from Mediterranean people, whose lands lie not too far distant to the south.

In an unusual congenital anomaly, his 12th ribs are missing. His seventh and eighth left ribs had been broken and had healed in his lifetime. According to Peter Vanezis of the University of Glasgow, his right rib cage is deformed and there are possible fractures of the third and fourth ribs. These changes happened after he died, as did a fracture of the left arm. That these breakages occurred after death is among the considerable evidence that casts doubt on the early disaster theory. So does the finding that an area of missing scalp was caused by pressure, not by a blow or decay.

Holding aside the unanswered questions concerning Ötzi's death and whether it was violent or not, several sound reasons suggest that he had not been in the best of health when he died. Although most of his epidermis (the outer layer of the skin), hair and fingernails are gone, probably having decayed as a result of exposure to water during occasional thaws, his remains still offer something of a health record for modern investigators. Examination of the only one of his fingernails to have been found revealed three Beau's lines, which develop when the nails stop growing and then start again. These lines show that he had been very ill

Overview/*A New Look at an Ancient Man*

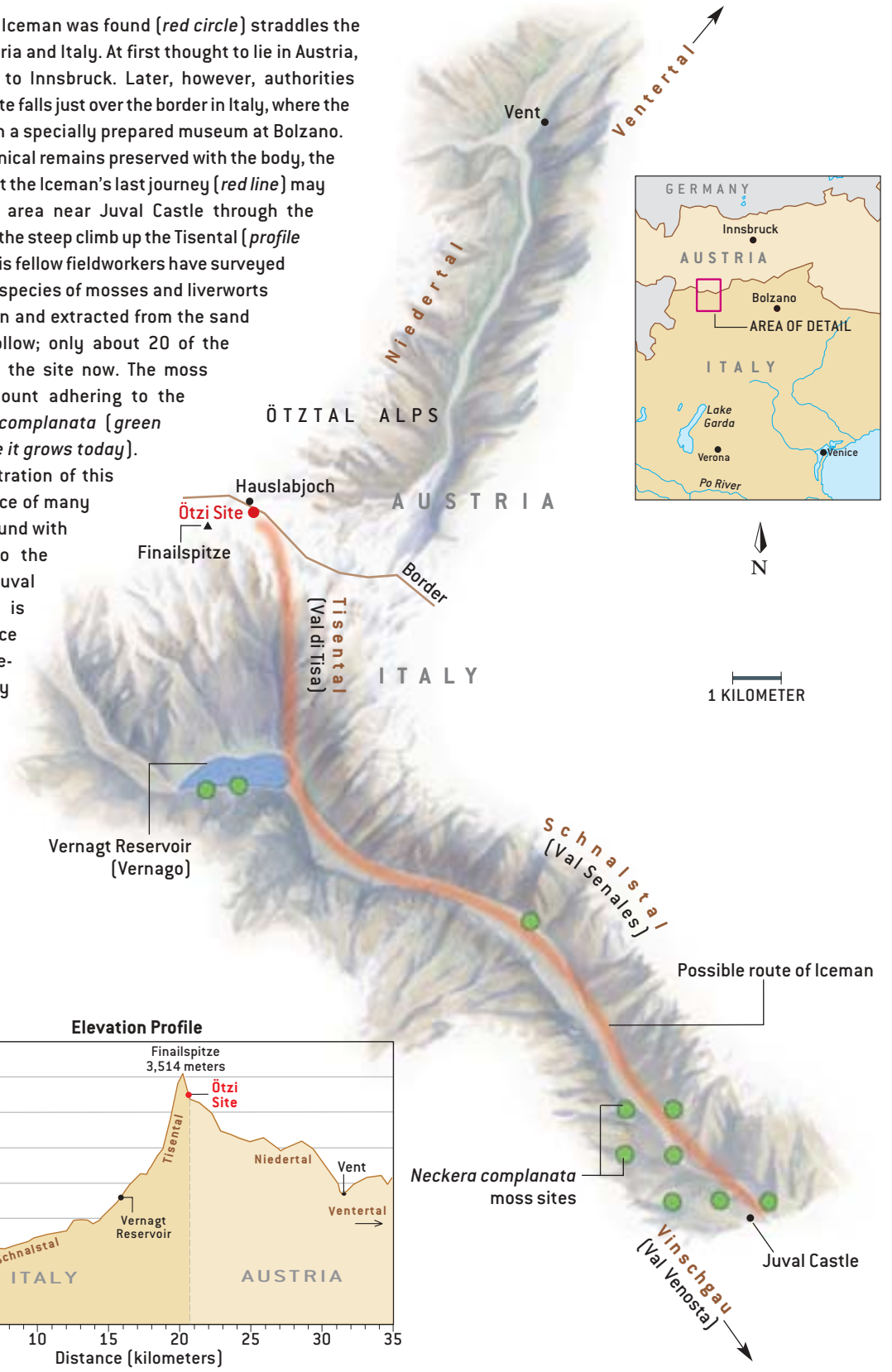
The most current research indicates that the Iceman:

- May have lived near where Juval Castle now stands in southern Tyrol (Italy)
- Ate a varied diet of primitive wheat, other plants and meat
- Was 46 years old and had not been in the best of health
- Died in the spring, not in the autumn as previously thought
- May have been killed by being shot in the back with an arrow
- Did not expire on the boulder where he was found, as was believed, but floated into position there during occasional thaws

SOUTH TYROL MUSEUM OF ARCHAEOLOGY, ITALY (www.iceman.it)

THE ROUTE THE ICEMAN MAY HAVE TAKEN

THE AREA WHERE the Iceman was found (*red circle*) straddles the frontier between Austria and Italy. At first thought to lie in Austria, the body was taken to Innsbruck. Later, however, authorities determined that the site falls just over the border in Italy, where the Iceman now resides in a specially prepared museum at Bolzano. Based mainly on botanical remains preserved with the body, the authors speculate that the Iceman's last journey (*red line*) may have been from the area near Juval Castle through the Schnalstal and finally the steep climb up the Tisental (*profile below*). Dickson and his fellow fieldworkers have surveyed this region for the 80 species of mosses and liverworts found with the Iceman and extracted from the sand and gravel in the hollow; only about 20 of the species grow around the site now. The moss found in largest amount adhering to the clothing is *Neckera complanata* (*green circles indicate where it grows today*). The greatest concentration of this moss and the presence of many of the other plants found with the Iceman occur to the south of the site, at Juval Castle, where there is archaeological evidence of a prehistoric settlement. This spot may have been his home.

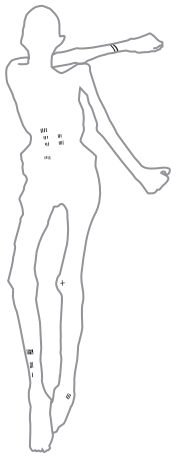


ROB WOOD *Wood Ransaville Harlin, Inc. (topographic map)*; MICHAEL SHAND *University of Glasgow (elevation profile)*

WHAT THE ICEMAN AND HIS BELONGINGS TELL US

Ötzi was small, standing only 159 centimeters (5′2.5″). Desiccation after he died shriveled the body both externally and internally. Ice pressure deformed his upper lip, nose and ears.

He was not bald in life, and he probably had a beard. The epidermis has come off, and all the hair and nails have fallen out. Some of his hair, up to nine centimeters long, was found. Analysis of the hair indicates that he ate a mixed diet of plants and animals.



Inconspicuous tattoos, most visible on the back of the body, were simple lines and crosses that may have been intended as therapy.

The isotopic composition of the tooth enamel suggests he had lived in at least two different areas.

The cap was sewn of brown bearskin.

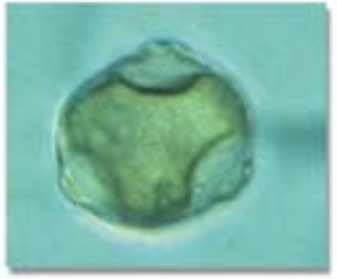
The handle of the dagger is ash wood; the flint tip may have been broken in antiquity or during the excavation.

Natural processes after death caused the fingers to clench. One fingernail (*below*) was recovered; the lines (*arrows*) reveal that he had been very ill three times in the months before he died.

A primitive wheat called einkorn has been identified in his gut, ground so minutely that it may well have been used to make bread. Tiny charcoal particles (*fuzzy, dark shapes*) suggest that the bread had been baked on an open fire.

SOUTH TYROL MUSEUM OF ARCHAEOLOGY, ITALY (www.iceman.it) (*iceman, ax, bearskin cap*); CHRISTIN BEECK, TAKEN FROM "DIE GLETSCHERMUMIE VOM ENDE DER STEINZEIT AUS DEN ÖTZTALER ALPEN VORBERICHT," RÖMISCH-GERMANISCHES ZENTRALMUSEUM, MAINZ (RGZM) (*shoes, pouch, dagger*); TAKEN FROM "DIE GLETSCHERMUMIE," RGZM (*iceman outline*).

Contents of the gut confirm that his diet was omnivorous and disclose details of his last meals (red deer, wild goat, plants and ground grains), his environment, his domicile and even his last journey.



In the gut samples, pollen grains (*left*) of the hop hornbeam tree (*below*) indicate that the Iceman died in late spring, when this small tree blooms.

The oldest intact ax ever found, it has a copper head secured to the yew handle by birch glue and bound with strips of hide.

The shoes were carefully stitched from hide and insulated with grass but were in a poorly preserved state, perhaps partly as result of wear and tear during the Iceman's journey.

Leaves of the moss *Neckera complanata,* also in the gut samples, suggest that he had wrapped his food in the moss. At top is the moss growing on a shady rock; below is one leafy stem washed from the clothes.

Radiocarbon dates of body tissues and gear (as well as plants) agree that he lived 5,300 years ago.

A pouch that probably fastened around his waist, although it was no longer in place, contained a fire-making kit, including true tinder fungus (*lower left*) and small flints (*middle*). A flint-sharpening tool is at the lower right.

TAKEN FROM "THE MAN IN THE ICE," VOL. 2, BY K. SPINDLER ET AL. (HRSG), DER MANN IM EIS. NEUE FUNDE UND ERGEBNISSE (SPRINGER, VIENNA-NEW YORK, 1995) (*fingernail*); KLAUS OEGGL (*pollen grain, einkorn*); JAMES H. DICKSON (*hop hornbeam tree, moss, leafy stem*)

On hide thongs, he carried two pierced pieces of BIRCH BRACKET FUNGUS, known to contain pharmacologically active compounds.

three times in the last six months of his life and that the final episode, about two months before his death, was the most serious and lasted at least two weeks. Horst Aspöck of the University of Vienna found that he had an infestation of the intestinal parasite whipworm, which can cause debilitating diarrhea and even dysentery, although we do not know how bad his infestation was.

Moreover, many simple, charcoal-dust tattoos are visible on the layer of skin under the missing epidermis. These marks were certainly not decorative and were probably therapeutic. Several are on or close to Chinese acupuncture points and at places where he could have suffered from arthritis—the lower spine, right knee and ankle. This coincidence has led to claims of treatment by acupuncture. Yet, according to Vanezis and Franco Tagliaro of the University of Rome, x-rays show little if any sign of arthritis.

The little toe of his left foot reveals evidence of frostbite. Ötzi's teeth are very worn, a reflection of his age and diet. Remains of two human fleas were found in his clothes. No lice were seen, but because his epidermis had been shed, any lice may have been lost.

What Was His Gear Like?

TURNING TO ÖTZI'S clothing and gear, scientists have learned not only about Ötzi himself but about the community in which he lived. The items are a testament to how intimately his people knew the rocks, fungi, plants and animals in their immediate surroundings. And we can see that they also knew how to obtain resources from farther afield, such as flint and copper ore. This knowledge ensured that Ötzi was extremely well equipped, each object fashioned from the material best suited to its purpose.

He had been warmly dressed in three layers of clothing—leggings, loincloth and jacket made of the hide of deer and goat, and a cape made of grass and bast, the long, tough fibers from the bark of the linden tree. His hat was bearskin, and his shoes, which were insulated with grass, had bearskin soles and goatskin uppers.

He had carried a copper ax and a dagger of flint from near Lake Garda, about 150 kilometers to the south. The handle of the dagger was ash wood, a material still used for handles today because it does not splinter easily. His unfinished longbow was carved from yew, the best wood for such a purpose because of its great tensile strength. The famous English longbows used to defeat the French at Agincourt some 4,000 years later were made of yew. A hide quiver contained 14 arrows, only two of which had feathers and flint arrowheads attached, but these two were broken. Thirteen of the arrow shafts were made of wayfaring tree, which produces long, straight, rigid stems of suitable diameter; one was partly of wayfaring tree and partly of dogwood.

A belted pouch contained a tinder kit, which held a bracket fungus that grows on trees, known as the true tinder fungus, and iron pyrites and flints for making sparks. A small tool for sharpening the flints was also found with the body. On hide thongs, Ötzi carried two pierced pieces of birch bracket fungus; it is known to contain pharmacologically active compounds (triterpens) and so may have been used medicinally. There were also the fragments of a net, the frame of a backpack, and two containers made of birch bark; one held both charcoal and leaves of Norway maple—perhaps it originally transported embers wrapped in the leaves.

Where Was He From?

IN THIS PART OF THE ALPS, the valleys run north and south between towering ranges of mountains. Thus, the question of Ötzi's homeland resolves itself into north versus south rather than east versus west. The botanical evidence points to the south. A Neolithic site has been discovered at Juval, a medieval castle at the southern end of the Schnalstal, more than 2,000 meters lower but only 15 kilometers from the hollow as the crow flies. Archaeologists have not excavated the site in modern times, and there has been no radiocarbon dating, but Juval is the nearest place to the hollow where a number of the flowering plants and mosses associated with Ötzi now grow. We have no reason to suppose that they did not grow there in prehistoric times, and so perhaps that is the very place where Ötzi lived.

When his clothes were conserved, the washing revealed many plant fragments, including a mass of the large woodland moss *Neckera complanata*. This moss and others he had carried grow to the north and to the south of where he was found, but the southern sources are much closer. *N. complanata* grows in some abundance near Juval. Wolfgang Hofbauer of the Fraunhofer Institute for Building Physics in Valley, Germany, has discovered that this moss grows, in more moderate amounts, at Vernagt (Vernago), just 1,450 meters lower than the site and only five kilometers away.

CHRISTIN BEECK, TAKEN FROM "DIE GLETSCHERMUMIE," RGZM

THE AUTHORS

JIM DICKSON, KLAUS OEGGL and *LINDA HANDLEY* share an interest in the plants that the Tyrolean Iceman may have used in his daily life. Dickson, professor of archaeobotany and plant systematics at the University of Glasgow, is recipient of the Neill Medal of the Royal Society of Edinburgh. He has written more than 150 papers and five books, including *Plants and People in Ancient Scotland* (Tempus Publishing, 2000), which he co-authored with his late wife, Camilla. Oeggl is professor of botany at the University of Innsbruck in Austria. He is an expert in archaeobotany and co-editor of the book *The Iceman and His Natural Environment* (Springer-Verlag, 2000). Handley, an ecophysiologist at the Scottish Crop Research Institute in Invergowrie, near Dundee, Scotland, specializes in the study of stable isotopes of carbon and nitrogen in plants and soils.

And most recently, Alexandra Schmidl of the University of Innsbruck Botanical Institute discovered small leaf fragments of the moss *Anomodon viticulosus* in samples taken from the stomach. This woodland moss grows with *N. complanata* in lowermost Schnalstal.

If Juval was not his home, signs of Neolithic occupation at other locations in the immediately adjacent Vinschgau (Val Venosta), the valley of the River Etsch (Adige), offer other possibilities. In contrast, to the north, the nearest known Stone Age settlements are many tens of kilometers away, and we are not aware of any Neolithic settlements in the Ventertal or elsewhere in the Ötztal. If Ötzi's home was indeed in lowermost Schnalstal or in Vinschgau, then his community lived in a region of mild, short, largely snow-free winters, especially so if the climate was then slightly warmer.

Investigations by Wolfgang Müller of the Australian National University of the isotopic composition of the Iceman's tooth enamel suggest that he had grown up in one area but spent the last several decades of his life in a different place. Investigating stable isotopes and trace elements, Jurian Hoogewerff of the Institute of Food Research in Norwich, England, and other researchers have claimed that Ötzi probably spent most of his final years in the Ventertal or nearby valleys to the north. If these deductions can be substantiated, they are intriguing developments.

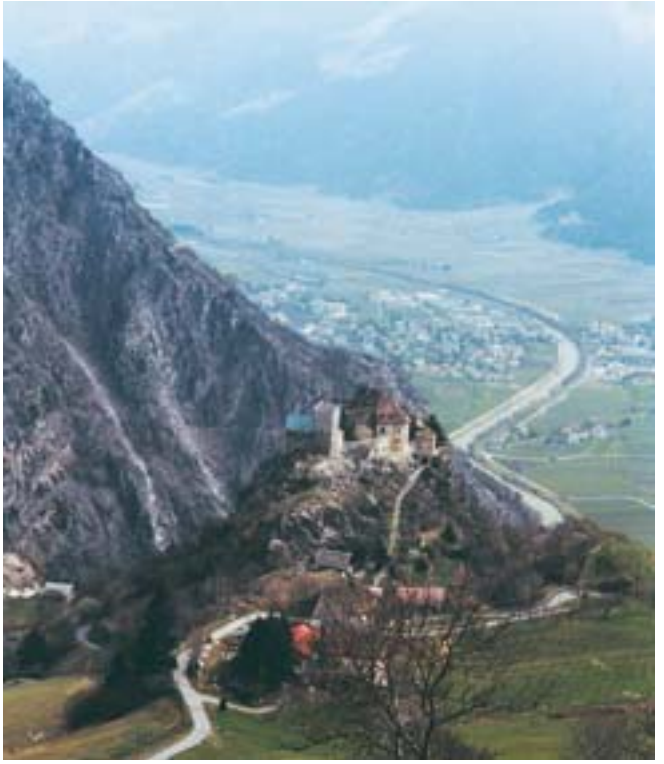
What Did He Eat?

THE ONGOING STUDIES of the plant remains in samples taken from the digestive tract provide direct evidence of some of Ötzi's last meals. One of us (Oeggl) has detected bran of the primitive wheat called einkorn, so fine that it may well have been ground into flour for baking bread rather than having been made into a gruel. Microscopic debris of as yet unidentified types shows that he had eaten other plants as well. And Franco Rollo and his team at the University of Camerino in Italy, in their DNA studies of food residues in the intestines, have recognized both red deer and alpine ibex (wild goat). Splinters of ibex neck bones were also discovered close to Ötzi's body. A solitary but whole sloe lay near the corpse as well. Sloes are small, bitter, plumlike fruit, and Ötzi may have been carrying dried sloes as provisions.

Several types of moss were recovered from the digestive tract. There is virtually no evidence that humans have ever eaten mosses, certainly not as a staple of their diet. But 5,000 and more years ago no materials were manufactured for wrapping, packing, stuffing or wiping. Mosses were highly convenient for such purposes, as many archaeological discoveries across Europe have revealed: various mosses in Viking and medieval cesspits were clearly used as toilet paper. Had Ötzi's provisions been wrapped in moss, that would neatly explain, as an accidental ingestion, the several leaves and leaf fragments of *N. complanata* recovered from the samples taken from the gut.

Analyzing archaeological remains of bone and hair for their abundances of the stable isotopes of carbon and nitrogen

ROCK-STREWN, ICY SITE where the Iceman was found has been marked by a commemorative monument. The three men are standing on the very spot where the body lay. Typically, the snow and ice did not melt enough in the summer of 2000, when the photograph was taken, to reveal the boulders and sandy gravel on the bottom of the hollow.



JUVAL CASTLE, in lowermost Schnalstal, is the place nearest to the site where many of the low-altitude plants preserved with the Iceman grow today.

JAMES H. DICKSON

(carbon 13 and nitrogen 15) can provide information about a person's diet. Nitrogen 15 can reveal the extent to which the individual relied on animal or plant protein. Carbon 13 can indicate the type of food plant the person ate and whether seafood or terrestrial carbon was an important part of the diet.

The isotopic data agree with the other evidence that Ötzi ate a mixed diet of plants and animals. He obtained about 30 percent of his dietary nitrogen from animal protein and the rest from plants. This value is consistent with those found in hunter-gatherer tribes living today. The data also indicate that seafood was probably not a component of his diet, a finding that makes sense because of the great distance to the sea.

What Was He Doing There?

TO THIS DAY, in what may be an ancient custom, shepherds take their flocks from the Schnalstal up to high pastures in the Ötztal in June and bring them down again in September. The body was found near one of the traditional routes, which is why early theories held that he was a shepherd. Nothing about his clothing or equipment, however, proves that he had done such work. No wool was on or around his person, no dead collie by his feet, no crook in his hand. Some support for the shepherd hypothesis comes from the grass and bast cape, which has modern parallels in garments worn by shepherds in the Balkans, but that alone is not conclusive; for all we know, it was standard dress for travelers at that time.

Analysis of the few strands of Ötzi's hair that survived reveals very high values of both arsenic and copper. The published explanation (also given independently on television) was that he had taken part in the smelting of copper. But Geoffrey Grime of the University of Surrey in England now considers that these exceptional levels may have resulted from the action of metal-fixing bacteria after Ötzi died and that the copper was *on*, not *in*, the hair. Further support for the possibility of copper having attached itself to the hair after death comes from the presence of the moss *Mielichhoferia elongata*, called copper moss, which spreads preferentially on copper-bearing rocks. It has been found growing at the site by one of us (Dickson) and, independently, by Ronald D. Porley of the U.K. government agency English Nature.

Another hypothesis is that Ötzi was a hunter of alpine ibex; the longbow and quiver of arrows may support this notion. If, however, he had been actively engaged in hunting at the time of his death, why is the bow unfinished and unstrung and all but two of the arrows without heads and feathers and those two broken?

Other early ideas about Ötzi are that he was an outlaw, a trader of flint, a shaman or a warrior. None of these has any solid basis, unless the pieces of bracket fungus he was carrying had medicinal or spiritual use for a shaman.

How Did He Die?

IN JULY 2001 Paul Gostner and Eduard Egarter Vigl of the Regional Hospital of Bolzano in Italy announced that x-rays had revealed an arrowhead in Ötzi's back under the left shoulder. This assertion has led to numerous statements in the media that Ötzi was murdered and to claims from Gostner and Egarter Vigl that

THE ICEMAN was well attired in three layers of clothing fashioned from skins, grass, and bark fibers. These reconstructions were made by the archaeologists in the Roman-Germanic Museum in Mainz, Germany, where all of the Iceman's clothing and gear were restored.

JULIA RIBBECK, TAKEN FROM "DIE GLETSCHERMUMIE," RGZM

it is "now proven that Ötzi did not die a natural death, nor due to exhaustion or frostbite alone." Although three-dimensional reconstructions of the object, which is 27 millimeters long and 18 millimeters wide, exist, requests by Vanezis and Tagliaro for the object to be removed to show convincingly that it is an arrowhead are still unanswered. Furthermore, it must be removed in a way that makes clear what fatal damage it might have done.

The arrowhead need not have caused death. Many people stay alive after foreign objects such as bullets have entered their bodies. A notable archaeological example is the Cascade spear point in the right pelvis of the famous Kennewick Man in North America; it had been there long enough for the bone to begin healing around it.

Even more recently, in a statement to the media, Egarter Vigl has reported that Ötzi's right hand reveals a deep stab wound. No scientific publication of this finding has been made yet.

THE ICEMAN now lies in a specially built chamber at the South Tyrol Museum of Archaeology in Bolzano, Italy, that keeps him at –6 degrees Celsius and roughly 99 percent humidity.

At What Time of Year?

INITIAL REPORTS PLACED the season of death in autumn. The presence of the sloe, which ripens in late summer, near the body and small pieces of grain in Ötzi's clothing, presumed to have lodged there during harvest threshing, formed the basis for these reports. But strong botanical evidence now indicates that Ötzi died in late spring or early summer. Studies by Oeggl of a tiny sample of food residue from Ötzi's colon have revealed the presence of the pollen of a small tree called hop hornbeam. Strikingly, much of that pollen has retained its cellular contents, which normally decay swiftly. This means that Ötzi might have ingested airborne pollen or drunk water containing freshly shed pollen shortly before he died. The hop hornbeam, which grows up to about 1,200 meters above sea level in the Schnalstal, flowers only in late spring and early summer.

As for the sloe found near his body, if Ötzi had been carrying sloes dried like prunes, the drying could have taken place some time before his journey. Small bits of grain also keep indefinitely, and a few scraps could have been carried inadvertently in his clothes for a long period.

What We Know

MORE THAN 10 YEARS after the discovery of the oldest, best-preserved human body, interpretations about who he was and how he came to rest in a rocky hollow high in the Alps have changed greatly. Just as important, we see that much careful research still needs to be done. The studies of the plant remains—the pollen, seeds, mosses and fungi found both inside and outside the body—have already disclosed a surprising number of Ötzi's secrets. We are aware of his omnivorous diet, his intimate knowledge of his surroundings, his southern domicile, his age and state of health, the season of his death, and something of his environment. Perhaps one of the most surprising reinterpretations is that Ötzi did not die on the boulder on which he was found. Rather he had floated there during one of the temporary thaws known to have occurred over the past 5,000 years. The positioning of the body, with the left arm stuck out awkwardly to the right and the right hand trapped under a stone, and the missing epidermis both suggest this conclusion. So does the fact that some of his belongings lay several meters distant, as if they had floated away from the body.

But we do not know and may never know what reason Ötzi had for being at a great altitude in the Alps. And we may never understand exactly how he died. An autopsy would be too destructive to be carried out. In the absence of this kind of proof, we cannot completely exclude the possibility that perhaps Ötzi died elsewhere and was carried to the hollow where the hikers found him 5,000 years later. SA

MORE TO EXPLORE

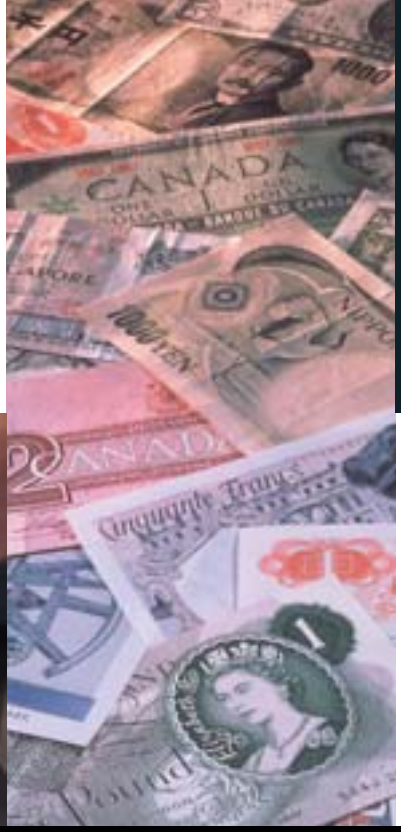
The Omnivorous Tyrolean Iceman: Colon Contents (Meat, Cereals, Pollen, Moss and Whipworm) and Stable Isotope Analyses. James H. Dickson et al. in *Philosophical Transactions of the Royal Society of London, Series B,* Vol. 355, pages 1843–1849; December 29, 2000.

INSIGHT: Report of Radiological-Forensic Findings on the Iceman. Paul Gostner and Eduard Egarter Vigl in *Journal of Archaeological Science,* Vol. 29, No. 3, pages 323–326; March 2002.

Ötzi's Last Meals: DNA Analysis of the Intestinal Content of the Neolithic Glacier Mummy from the Alps. Franco Rollo et al. in *Proceedings of the National Academy of Sciences USA,* Vol. 99, No. 20, pages 12594–12599; October 1, 2002.

SOUTH TYROL MUSEUM OF ARCHAEOLOGY, ITALY (www.iceman.it)

LAURA McGINNIS (*below*) relies on an orphan drug called Cystagon to treat her cystinosis, a rare disorder in which the amino acid cystine can build up to organ-destroying levels. Cystagon, which can cost more than $7,000 a year, can also be mixed with food. McGinnis receives her dose through a stomach tube.



DANIELLE BARCKETT, age seven (*left* and *above*), has hereditary tyrosinemia type 1, an enzyme deficiency that can kill patients by destroying the liver and the kidneys. An orphan drug called Orfadin—which can cost up to $88,000 a year—has allowed her to thrive, although she must still eat a special diet.

The Orphan Drug Backlash

The Orphan Drug Act of 1983 was supposed to provide incentives for private industry to develop needed, but unprofitable, drugs to treat rare diseases. It has done so, but not without eliciting controversy

By Thomas Maeder

IN JUNE 1989 AMGEN, A SMALL BIOTECHNOLOGY COMPANY IN THOUSAND OAKS, CALIF., gained U.S. Food and Drug Administration approval to market its first product, epoetin alfa (Epogen), to treat the anemia that accompanies end-stage kidney failure. Because the number of patients with this condition was not large—only around 78,000 at the time—it seemed unlikely that Amgen could make a profit or even recover the development costs of the drug. But it sought FDA approval for the agent anyway, in part because of incentives it was deemed eligible to receive under a law called the Orphan Drug Act. The act, which encompasses a set of laws that went into effect in 1983, provides benefits to encourage private industry to develop treatments for rare diseases. Treatments with such modest markets would otherwise remain "orphans," with no one to sponsor them through FDA scrutiny.

Once on the market, Epogen proved useful for other, more common purposes: restoring red blood cells in people suffering from bone marrow suppression as a result of taking AIDS drugs or cancer chemotherapy, and reducing the need for transfusions

in surgery patients. It wasn't long before the company started to earn tremendous profits on the drug. Outraged legislators and consumer groups cried foul, accusing Amgen and a few other companies of parlaying government largesse into private fortunes. By 2001, Epogen and Procrit, the latter a version of epoetin alfa made by Ortho Biotech in Raritan, N.J., were the sixth and seventh best-selling drugs in America, respectively, together generating more than $5 billion in revenues a year. Based on the drugs' success, people began to ask: Had the Orphan Drug Act been co-opted as a Biotechnology Promotion Act? And if so, shouldn't something be done to curb potential abuses in the future?

ACTORS JACK KLUGMAN and Robert Ito testify before a fake Congress in an episode of the television drama *Quincy, M.E.,* that focused on the dearth of treatments for rare disorders. Five days after the show aired, Klugman appeared before the real Congress and offered testimony that aided the passage of the 1983 Orphan Drug Act, which offers incentives to drugmakers.

A Look Back

THESE ARE REASONABLE questions, best answered by first recalling some of the act's history. Before 1983, even when treatments for rare diseases had been discovered, drug companies often did not want to make them. Although rare diseases collectively affect 25 million Americans—a huge constituency in aggregate—they are an unattractive market because they are subdivided into more than 6,000 subpopulations ranging in size from a handful of patients to a couple hundred thousand. Development costs for drugs are incredibly steep (the pharmaceutical industry routinely claims that it costs $800 million to get a new drug to market), so large pharmaceutical companies have historically concentrated on top-selling products, especially those to treat relatively common, chronic disorders such as hypertension, depression and arthritis.

In the late 1970s individual patients and volunteer health organizations began agitating for legislation to encourage the development of needed therapies that basic research had already identified but that were unavailable commercially. A company scrapped plans to make pimozide for Tourette syndrome, for instance, even though it was the only drug that helped many sufferers. Penicillamine for Wilson disease, 5-hydroxytryptophan for myoclonus, gamma-hydroxybutyrate for narcolepsy, sodium valproate for certain forms of epilepsy, and cysteamine to treat children with cystinosis [*see glossary on opposite page*] were also quietly dropped when they proved more costly than their meager sales projections warranted. Patients had to smuggle supplies from abroad, concoct illegal home brews or simply do without. The challenge was how to motivate the powerful drug industry to respond to desperate needs—how to make the unprofitable profitable. Orphan drugs "are like children who have no parents," says Representative Henry A. Waxman of California, "and they require special effort."

Early orphan drug legislation failed because it lacked the necessary economic incentives, yet it focused industry, government and popular attention on the problem and galvanized patient advocacy groups into action. One of the most influential events was a curious case of life imitating art imitating life: actor Jack Klugman, who played a strident medical examiner on the television show *Quincy, M.E.*, was inspired to create an episode

Overview/*Orphan Drugs*

- "Orphan" drugs are those for which the markets are so small that they are unlikely to be produced by a for-profit drug company. The U.S. Food and Drug Administration defines an orphan drug as one that is anticipated to treat fewer than 200,000 people.
- The 1983 Orphan Drug Act offers incentives for pharmaceutical and biotechnology companies to develop drugs for rare disorders. The act provides tax credits and seven years of market exclusivity to a company willing to make an orphan drug.
- Several orphan drugs—notably epoetin alfa, which builds up red blood cells—have now become blockbusters, leading critics to question whether drug companies are abusing the Orphan Drug Act.
- But advocates say the act has worked well: 229 orphan drugs that together treat 11 million patients, most with serious or life-threatening diseases, are now on the market.

CLOCKWISE FROM TOP LEFT: MARY ANN CHASTAIN *AP Photo*; JOHN GILLMOURE *Corbis*; ALIAS; PETER MURPHY; CORBIS; FRANK FILECCIA; PETER MURPHY *(page 80)*; EVERETT COLLECTION *(above)*

focusing on rare diseases after reading a newspaper article about the plight of a Tourette patient and of someone with myoclonus. In the show, which aired in 1981, Quincy testified before Congress about the shameful lack of drugs for rare diseases. Later, Klugman reenacted his fictitious TV testimony in front of the real Congress. Finally, in January 1983, President Ronald Reagan signed the Orphan Drug Act into law. It was subsequently modified several times and now allows companies to take a 50 percent tax credit on all clinical trial costs, exempts them from paying the so-called user fee (currently $533,400) that the FDA usually charges drug sponsors, and bars other firms from obtaining FDA approval for the same drug for seven years.

Cornering the Market

THE SEVEN-YEAR MARKET exclusivity clause has been key to the effectiveness of the act. The FDA can approve the same drug made by a prospective competitor only if it is "clinically superior"—if the product is safer, more effective or easier to take. The original act defined orphan drugs as those that could not reasonably be expected to recover their cost of development through sales in the U.S. But the complexity of making such projected economic analyses at first deterred companies with orphan drug candidates, and the initial industry response was discouraging. A 1984 amendment clarified matters by stipulating that drugs for conditions with fewer than 200,000 American sufferers would be presumed to be unprofitable and would therefore automatically qualify for orphan drug designation.

The Orphan Drug Act was needed not only to provide financial incentives to companies but also to allow drugmakers more leeway in designing studies to prove that a drug candidate is safe and effective. The standard set of human clinical trials can take years and involve thousands of patients at multiple sites. The entire patient population with a rare disease, on the other hand, may be smaller than the number of subjects in most ordinary trials, so testing cannot follow the usual protocols. Only 12 children in the U.S. suffered from adenosine deaminase (ADA) deficiency, a cause of severe combined immunodeficiency disease (SCID), for example, when a company was developing a drug for it. Similarly, sacrosidase, for a congenital enzyme disorder (sucrase-isomaltase deficiency), was approved on the basis of two trials that had a grand total of 41 patients.

Because of such special circumstances, interactions between the FDA and companies working on orphan drugs tend to be more collegial than the usual arm's-length relationships the agency typically maintains with drug sponsors. Marlene E. Haffner—director of the Office of Orphan Products Development at the FDA, and the self-styled "mother of orphan drugs"—is adamant that the role of her office is both to regulate new drugs for rare disorders and to help get safe and effective ones on the market. To this end, her agency administers a grants program to assist researchers working on drugs eligible for orphan status. Since 1983 the FDA has awarded 370 such grants, totaling more than $150 million. It also assists orphan drug sponsors in designing statistically meaningful clinical trials, a tricky undertaking for rare disorders.

Grateful Orphans

BUT RESEARCH into rare disorders has sometimes yielded disproportionately fruitful results. This effect, which has much to recommend it, also occasionally leads to windfalls for drugmakers and, in turn, to some seriously raised eyebrows.

Rare diseases often result from a specific genetic defect, such as a single mutation, so their symptoms may reveal the normal function of a particular gene. Alpha-1-antitrypsin deficiency, for instance, produces emphysema in young people—a tragedy for those with the underlying mutation but an opportunity for researchers to study the causes of the dis-

GLOSSARY OF SELECTED RARE DISORDERS

CYSTINOSIS: A buildup of the amino acid cystine, a constituent of proteins, causing organ damage (particularly in the kidneys and eyes). Affects roughly 400 people in the U.S.

GAUCHER DISEASE: An accumulation of a fatty compound, especially in the bone marrow, spleen, lungs and liver, resulting from a deficiency in glucocerebrosidase. Symptoms, afflicting fewer than 10,000 people worldwide, include enlarged liver and spleen, anemia, low levels of blood-clotting platelets, and skeletal abnormalities.

HEREDITARY TYROSINEMIA TYPE 1: A deficiency in the enzyme that normally breaks down the amino acid tyrosine. It can result in severe liver and kidney disease among the 100 individuals who have it in the U.S.

MYOCLONUS: A neurological disorder characterized by sudden, involuntary muscle contractions and relaxations. Strikes roughly nine of every 100,000 individuals.

SUCRASE-ISOMALTASE DEFICIENCY: The lack of the enzymes sucrase and isomaltase, leading to the inability to digest sugars and starches properly. Experienced by 0.2 percent of North Americans.

THROMBOTIC THROMBOCYTOPENIC PURPURA: Abnormally low platelet counts and shortened red blood cell survival time, resulting in a tendency to bleed excessively into the skin or mucous membranes. There are 15,000 to 22,000 U.S. sufferers.

TOURETTE SYNDROME: Muscle and vocal tics that can take the form of involuntary movements of the extremities and face, accompanied by uncontrollable sounds or socially inappropriate words. Roughly 100,000 Americans have the disorder.

WILSON DISEASE: A buildup of copper in various body tissues, particularly in the liver, brain and corneas. Can lead to liver failure and central nervous system dysfunction in the 30,000 people affected worldwide.

ease without the confounding effects of smoking and age. In the case of thrombotic thrombocytopenic purpura, blood clots caused by the absence of an enzyme that normally cleaves a blood protein may highlight a possible contributor to heart attack and stroke.

Indeed, rare diseases and orphan drugs have been boons to applied pharmaceutical research, although no one anticipated that the Orphan Drug Act would end up kick-starting the nascent biotech industry. Unlike conventional pharmaceutical companies, which mostly manufacture drugs composed of small molecules, biotech firms at the time tended to be start-ups and to focus on producing proteins needed to replace those that were defective or missing in unlucky people. And biotech firms made the proteins using recombinant DNA technology, which was first introduced in the 1970s. They would isolate, or clone, the gene encoding a human protein and splice it into bacteria or mammalian cells grown in laboratory culture dishes to produce the protein in quantity.

Many rare diseases, it turned out, were ideally suited to treatment with biotech products. Serious or life-threatening hereditary disorders, such as enzyme deficiencies stemming from a single defect in the enzyme's genetic blueprint, were both rare and potentially treatable with recombinantly produced replacement proteins—molecules that were too difficult and expensive for conventional pharmaceutical companies to manufacture and often costly and hard to extract from human or animal tissues. When the act went into effect, it motivated small biotech companies to develop these recombinant proteins into drugs even if they couldn't patent them—a particularly acute worry in the early days of biotechnology, when the U.S. Patent and Trademark Office was still struggling to figure out how best to provide patent protection to naturally occurring molecules.

A recent study by the Tufts Center for the Study of Drug Development found that from 1983 to 1992 the biotech industry secured 19 percent of all orphan drug approvals; 76 percent of such approvals went to pharmaceutical companies. By 2001 biotech's share had grown to 41 percent. Of the 10 best-selling biotech drugs worldwide in 2001, five were originally approved as orphan drugs, and three more were approved for orphan indications in addition to their original use, which afforded their developers seven years of marketing exclusivity [*see table below*]. In-

TOP 10 BIOTECH DRUGS AND THEIR ORPHAN DRUG STATUS

RANK	TRADE NAME	GENERIC NAME	MAJOR INDICATION	2001 SALES WORLDWIDE (in U.S. millions)	ORIGINAL U.S. DEVELOPER	ORIGINAL U.S. APPROVAL DATE
1	Epogen Procrit Eprex	Epoetin alfa	Anemia	$5,588	Amgen	June 1989
2	Intron A PEG-Intron Rebetron	Interferon-alpha 2b	Hepatitis C	$1,447	Schering-Plough	November 1988
3	Neupogen	Filgrastim	Neutropenia (low white blood cell count)	$1,300	Amgen	February 1991
4	Humulin	Human insulin	Diabetes	$1,061	Genentech	October 1982
5	Avonex	Interferon-beta 1a	Multiple sclerosis	$972	Biogen	May 1996
6	Rituxan	Rituximab	Non-Hodgkin's lymphoma	$819	IDEC Pharmaceuticals	November 1997
7	Protropin Nutropin Genotropin Humatrope	Somatropin	Growth disorders	$771	Genentech	October 1985
8	Enbrel	Etanercept	Arthritis	$762	Amgen (formerly Immunex)	November 1998
9	Remicade	Infliximab	Crohn's disease	$721	Centocor	August 1998
10	Synagis	Palivizumab	Pediatric respiratory disease	$516	MedImmune	June 1998

Originally approved as an orphan drug

Not originally approved as an orphan drug but now granted orphan drug status for one or more subsets of disorders

Never designated an orphan drug

SOURCES: Nature Reviews: Drug Discovery, *Vol. 1, No. 11, page 846; November 2002; FDA*

People began to ask: Had the Orphan Drug Act been co-opted as a BIOTECHNOLOGY PROMOTION ACT?

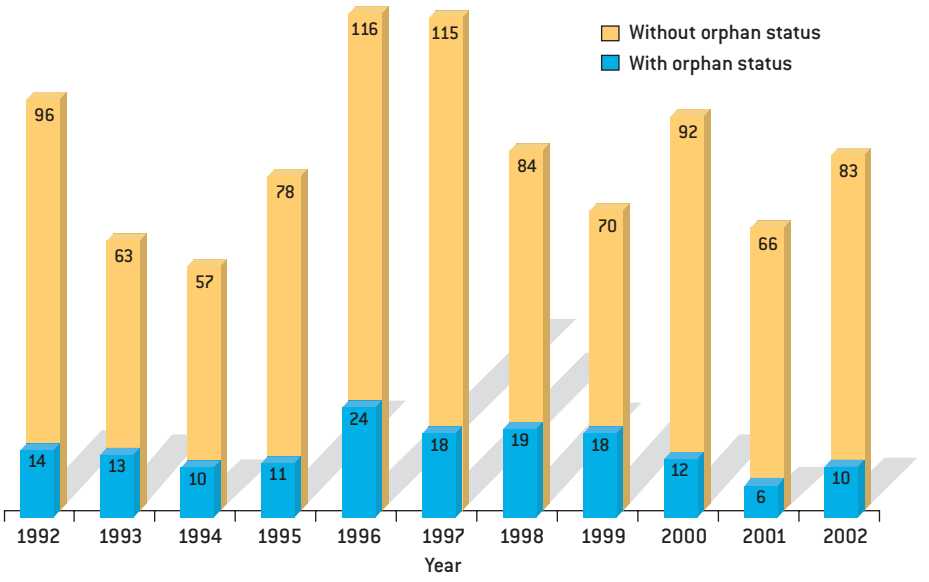
deed, the biggest moneymaking orphan products helped to launch some of the major players in the biotech industry, including Amgen and Genentech.

A striking example of an orphan drug technology that blossomed into more widespread use is pegylation, the process of adding to a protein a waxy substance called polyethylene glycol (PEG), which slows the drug's clearance from the bloodstream and masks it from attack by the immune system. Pegylation debuted in 1990 in Enzon's orphan drug Adagen for the treatment of ADA-SCID. Although only a few dozen children worldwide suffer from this condition, pegylation technology is now used in PEG-Intron, part of a combination treatment for hepatitis C, and has tremendous potential in other therapeutic applications.

Even aside from serendipitous blockbusters, one can make money on orphan drugs. In 1988 Lars-Uno Larsson, a former Bristol-Myers Squibb executive, founded Swedish Orphan International in Stockholm on the belief that the Orphan Drug Act made it possible to earn modest but sufficient returns on drugs for rare diseases. Now with affiliates around the world, Swedish Orphan has developed a number of products and inspired others to establish similar companies. "The financial markets look on smaller products and say, 'How can you make money with a $10-million product?' " observes John Bullion, a former venture capitalist and now the CEO of Orphan Medical, a Minnesota-based company with half a dozen approved orphan products. "Well, you can make very good money with a $10-million product, but you need several of them," he says.

People with rare diseases are usually treated by a handful of doctors who have experience in the disease and often join patient education or advocacy groups to share information and to lobby for more research on their disorder. This combination makes finding patients to participate in clinical trials and to buy the drugs relatively easy and cost-effective. According to a recent industry analysis, it costs about one fourth as much to develop a drug for a rare disease as one for high blood pressure, and annual marketing costs are one seventh as high.

DRUGS AND BIOLOGICS APPROVED BY FDA



ORPHAN DRUGS account for 17 percent of all drugs and biologics (therapeutics derived from living sources) approved for sale in the U.S. over the past 10 years.

JOHNNY JOHNSON; SOURCE: FDA

Rich Orphans

BUT SOME OF THE ABUSES of the Orphan Drug Act have been glaring—and it's not only Amgen's epoetin alfa that has hit the jackpot. When they were first introduced, Genentech's human growth hormone (hGH) and GlaxoSmithKline's AZT were approved for rare disorders (hGH deficiency and AIDS, respectively), but they subsequently earned billions when physicians began to prescribe hGH for many short-statured children and when the AIDS epidemic ballooned. Some critics suggest that orphan drugs that make profits no longer need help and should forfeit the act's benefits. The European Union recently enacted a law that would strip orphan status from a drug that becomes "extraordinarily profitable" after five years, but similar measures have been rejected repeatedly by American legislators under the lobbying pressure of the U.S. pharmaceutical industry. It is too early to tell whether Europe's law will prevent drug companies from abusing orphan drug designation; national health plans that exert controls on drug prices will also make the law's effects hard to assess.

A notable case of pricing and profits, according to some, is Genzyme's imiglucerase (Cerezyme). The drug—an enzyme-replacement therapy for Gaucher disease, which afflicts 2,000 Americans—is the world's most expensive medicine. Genzyme reportedly earns close to half a billion dollars a year from this treatment by charging patients between $100,000

THE AUTHOR

THOMAS MAEDER is senior adviser for scientific strategies at the Georgetown University Medical Center. He is the author of numerous books, including *Adverse Reactions,* about modern drug regulation as seen through the tragic story of chloramphenicol, which caused a fatal blood disorder. He has done graduate work in neurobiology at the University of Pennsylvania and was previously senior science writer at *Red Herring* magazine.

Good Things for Small Populations

The Humanitarian Device Exemption is for medical devices what the Orphan Drug Act is for drugs—almost

MEDICAL PRODUCTS regulated by the FDA include not only drugs but a dizzying variety of *things.* The FDA's Center for Devices and Radiological Health (CDRH) regulates thousands of medical devices ranging from tongue depressors and surgical drapes to artificial hearts and CT scan machines.

The CDRH has a "custom device exemption" that waives many requirements for devices that are sold to meet the specific needs of a small number of patients or the idiosyncratic requirements of a particular physician or dentist. But that exemption has traditionally been limited to half a dozen or so devices a year. So when demand began to grow for a special tube—called a stent—that is the only way to treat bladder obstruction in fetuses, the FDA didn't know what to do.

Bladder obstructions afflict roughly 200 fetuses in the U.S. every year and can cause death unless cleared to allow the fetus's urine to pass into the amniotic fluid. For decades, expectant mothers had to endure having doctors use long needles to draw urine from the fetus's bladder on a regular basis. In 1982 medical supplier Cook Urological in Spencer, Ind., began to sell a stent that can be inserted to allow urine to drain past the obstruction. At first, the company made the stent as a custom device sold at cost for individual patients. But when the number of stents sold reached 680, the CDRH could no longer consider them custom items. To cover devices such as the Cook Harrison fetal bladder stent, the agency created the Humanitarian Device Exemption in 1996.

The exemption is limited to devices for the diagnosis or treatment of conditions affecting fewer than 4,000 people in the U.S., and—unlike the Orphan Drug Act—it explicitly bars manufacturers from making a profit. To offset this clause, companies are required to prove only that their devices are safe, not that they work, and are allowed to sell them while conducting clinical trials of their efficacy. But just 28 devices have been approved under the exemption in six years, causing some to question whether the profit ban should be repealed to offer companies more incentive to bring such products to market. *—T.M.*

FETAL BLADDER STENT (*shown enlarged*) is inserted into the urinary tract of a fetus to treat bladder obstruction. Its success prompted the U.S. Food and Drug Administration to draft the Humanitarian Device Exemption to spur companies to make such devices.

COOK UROLOGICAL

and $400,000 a year for it, depending on whether the patient is a child or an adult. And the company did not lower the price when it switched from extracting the substance from human placentas to using the less expensive recombinant method of production. Patients, some of whom must spend their way into poverty to qualify for Medicaid to afford Cerezyme, are angry about the price but grateful for their lives.

Abbey S. Meyers, president of the National Organization for Rare Disorders (NORD)—the fiercest defender of those with rare diseases—is weary of answering the question of whether companies should be barred from making unreasonable profits on orphan drugs. "To know whether something is excessively profitable, you have to be able to look at their books, and they won't let you do that," she states. "And how would you define 'unreasonably profitable' anyway?"

Companies defend their crushingly expensive prices by citing their need to survive and to fund future research. And they point out that in some cases, they could charge much more than they do. Rare Disease Therapeutics, a Tennessee affiliate of Swedish Orphan, priced Orfadin, its recently approved treatment for hereditary tyrosinemia type 1, by making it $80,000 cheaper than a liver transplant, the only other therapy for the disorder. (Company officials say they arrived at the price by computing their costs.) The firm—which is not publicly held and therefore not under profit pressure from Wall Street—could easily have gotten twice as much for the only drug that keeps these children alive.

That's small comfort to patients. Happily, though, patient-advocate Meyers can think of no one with a rare disease who has been unable to obtain an FDA-approved drug because he or she could not pay for it. Most companies that produce orphan drugs have formal or informal programs for providing drugs free to indigent patients, but they hardly ever make public the number of patients they accommodate or what such patients must do to qualify. State and federal plans also exist for helping those with various rare disorders, especially hemophiliacs, some

"To know whether [an orphan drug] is excessively PROFITABLE, you have to look at [the company's] books. And how would you define 'unreasonably profitable'?"

of whom must spend $100,000 or more each year for a blood-clotting protein called Factor VIII.

But only selected disorders are covered by such corporate and government programs, a policy that "leaves the rest of us out in the cold," Meyers comments. And many families have sold their houses or spent their life savings to qualify for the support. Even those with insurance are unprotected when treatment costs either exceed their plan's lifetime cap or drive premiums unaffordably high, she explains: One man who worked as an accountant for a small company resigned because payments for his son's human growth hormone drove his employer's insurance rates through the roof and affected fellow employees. He went to work as a manual laborer in a factory with a workforce large enough to distribute the cost.

No one has been able to offer an acceptable solution to the problem of unseemly profits. Haffner accepts it as the way things are in a free-market society. Meyers, the most obvious person to protest, is grateful for what populations with rare diseases have won. "Pricing is something we can't do anything about. As long as we can make sure that patients have access, it's great. If they don't, *then* it should be on the front page of the newspapers," she asserts. She worries that efforts to fine-tune the delicate structure might bring the whole system down. "We use all of our energy trying to keep the Orphan Drug Act the same," she says. "When the drug companies want to fight, they can hire a lobbyist for every member of Congress, while we have a bunch of mothers wandering around with sick babies."

In the end, the high cost of orphan drugs probably has to be addressed as part of the bigger problem of high drug costs in general. The FDA itself is in no position to insist that costs come down for orphan drugs or any others. It has no authority over pricing, only over which, if any, uses can be claimed for a given drug. All it can do to influence profits is to decide whether a compound said to target a rare disease is truly likely to have an extremely limited market and thus deserves the protection of the act.

Salami Slicing

BESIDES ALLOWING companies to retain orphan status for drugs that end up with an unusually large market, the act has been used to favor corporate bottom lines in another way that critics are protesting. The FDA permits companies to parse diseases into "medically plausible subsets," a term that it does not define clearly. The agency's orphan products office has tried to fight such "salami slicing," but some firms still try to define one stage or manifestation of a disease as a distinct entity entitled to orphan drug benefits if it affects fewer than 200,000 people.

This issue will inevitably grow more complex with the advent of pharmacogenomics and personalized medicine—the possibility of targeting treatments more accurately, and thus more safely and effectively, at subpopulations and even at individuals by determining through genetic profiles who would respond best and who would be least likely to suffer adverse reactions to a given drug. This prospect could benefit patients, but it is also expected to challenge the pharmaceutical industry, which has hitherto relied on mass sales of identical drugs to large, poorly differentiated patient populations. In the next decade a multitude of diseases may be broken down into diagnostically and therapeutically distinct populations that meet the threshold requirement of medically plausible subsets. Then society may find that many or all drugs are orphans, and policymakers will need to revisit the question of how to stimulate research for truly rare diseases.

Twenty years after the passage of the Orphan Drug Act, many believe that it has exceeded its original expectations. During the decade before its appearance, 34 orphan products went on the market, 10 of them developed by the pharmaceutical industry and the other 24 by federally funded efforts. In the two decades since, 229 orphan drugs that together treat 11 million patients, most with serious or life-threatening diseases, have entered the market, and the FDA has granted orphan status to nearly 1,000 other drugs. "The Orphan Drug Act works fabulously well," Meyers concludes. "We have treatments we never imagined we would." Existing pharmaceutical and biotech companies have been induced to develop orphan products, and new companies have been founded for the exclusive purpose of addressing unmet needs.

In a more subjective measure of the act's success, it has been copied almost verbatim by the European Union, Australia, Japan and several other Asian countries. And the U.S. Department of Homeland Security is even considering it as a model for how to steer efforts toward developing vaccines and antidotes for possible biological warfare agents.

The act certainly has it warts, but in a free-market economy, it is the best model devised so far to ensure that those with rare diseases can get the treatments they so desperately need. SA

MORE TO EXPLORE

Two Decades of Orphan Product Development. Marlene E. Haffner, Janet Whitley and Marie Moses in *Nature Reviews: Drug Discovery,* Vol. 1, No. 10, pages 821–825; October 2002.

The U.S. Food and Drug Administration Office of Orphan Products Development Web site includes complete information on the Orphan Drug Act and amendments, procedural guidelines, and lists of all orphan drug designations, approvals and grants: **www.fda.gov/orphan**

The National Organization for Rare Disorders (NORD) is a source of comprehensive information on rare diseases, support groups and news: **www.rarediseases.org**

Orphanet is a multilingual European orphan drug/disease database: **orphanet.infobiogen.fr**

ANTENNAS

Catch a Wave

Despite the availability of cable television and Web radio, millions of people rely on antennas for TV and radio reception. And with the popularity of satellite TV, portable radios, cell phones and wireless gadgets, more antennas are in use today than ever.

But why all the shapes and sizes? Antenna span, for one, is dictated by the type of broadcast it captures; an incoming radio-frequency wave resonates best when an antenna's length is one half that of the wavelength. Because waves can vary from 18 feet for channel 2 television to three inches for cellular telephony, proper antenna length is key to reception.

Correct orientation can help. The electric part of television and FM broadcasts is horizontally polarized, so an antenna lying in a horizontal plane works best. AM broadcasts are vertically polarized. Yet most modern transmitters provide both polarizations by propagating waves in a corkscrew pattern, so antennas in various planes can receive the signal nearly as well as those in the optimum plane. Certain configurations, such as TV rabbit ears, also allow consumers to rotate or pitch antenna elements to reduce interference.

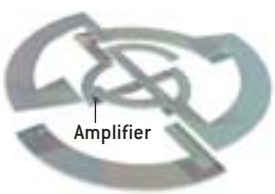
Beyond antenna length and orientation, radio-frequency waves aren't fussy. A wave simply induces a voltage in an antenna, so many of the odd shapes and contraptions touted by marketers add nothing electrically. "They're done for window dressing," says Kent Britain of RDI Electronics in Grand Prairie, Tex. Certain styles, such as flat TV-top antennas, sell "because they're more appealing cosmetically," notes Len Davi of Recoton/Jensen in New York City.

Examples abound. For a time, it was trendy to embed antennas in car windshields; they didn't work as well as antennas out in the open, but the vehicles looked sleeker. And what about the coils or fins in some TV-top antennas? They're useless, Britain says.

Amplifiers in an antenna base can boost the millivolt signals, yet these may be oversold as well. "You only need to amplify by 12 to 14 decibels in most applications," Britain points out. "More than that doesn't help and can actually overload the electronics. A 30-decibel amplifier offers no advantage over a 20-decibel amplifier." —*Mark Fischetti*

AM
antenna is usually a coil of wire around an iron ferrite rod, which captures the magnetic part of the 1,000-foot AM waves; it is less efficient than a dipole but more convenient than a 500-foot-long antenna.

FM
antenna used commonly indoors is a simple dipole wire that intercepts the electric part of a radio wave; it is 4.5 to 5.5 feet long to cover the half-wavelengths of most FM transmissions.



SOLID-STATE TV
antenna uses small amplifiers powered by a slight direct current, coupled with several angular elements inside a disk-shaped housing. The antenna is small and light and is often added to satellite TV dishes, which don't pick up local stations.

JOHN MacNEILL

DID YOU KNOW . . .

➤ CLEAR ADVICE: Which antenna design and orientation will optimize television reception at your home? Enter any U.S. address at www.antennaweb.org, and a Consumer Electronics Association database will respond. For a host of tutorials and tricks, consult the RadioShack book *Antennas,* by Alvis Evans and Kent Britain (1998).

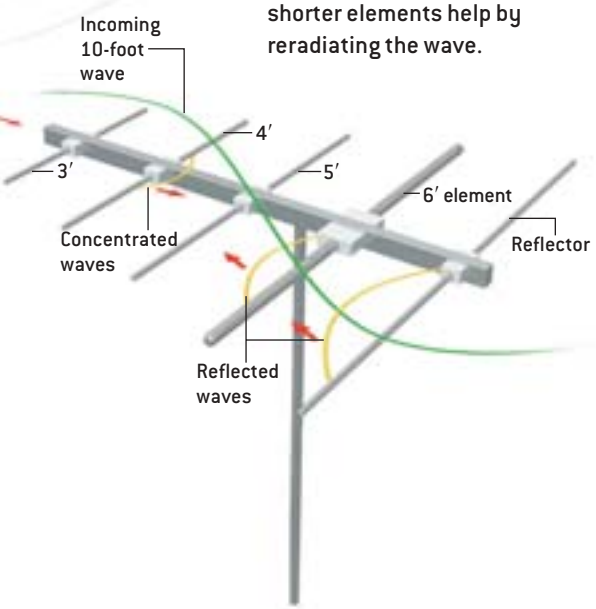
➤ FREQUENCY HOPPING: Certain high-performance gear, from cordless phones to military radars, use a spread-spectrum technique of sending and receiving signals in a set pattern to minimize interference—say, at 64 frequencies close to 900 megahertz. Depending on the application, the antenna can be a three-inch rod or a drum two feet in diameter.

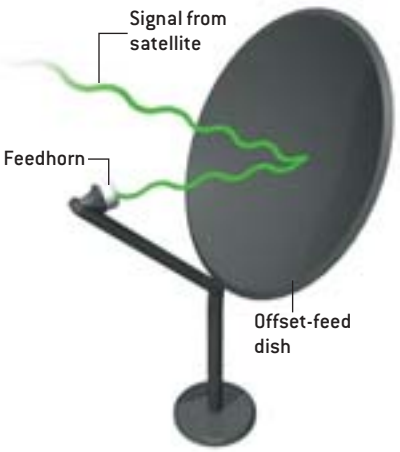
➤ HEARING AID: Telephone transmitters send sound to cell phones by modulating pulses on and off. Electronic noise caused by the action is imperceptible to the ear but is amplified by hearing aids, which cause annoying buzzing. Myers Johnson, Inc., offers a small antenna that clips to a phone, senses the clatter and creates opposing waveforms that cancel the noise before it reaches the hearing aid.

➤ WET WAVES: Ground waves transmitted from TV and radio towers travel farthest over water or flat, moist soil, both of which provide a greater index of refraction than uneven or dry terrain. Very moist air hovering close to the ground after widespread, heavy rain can channel waves for hundreds of miles, bringing in stations from far away.

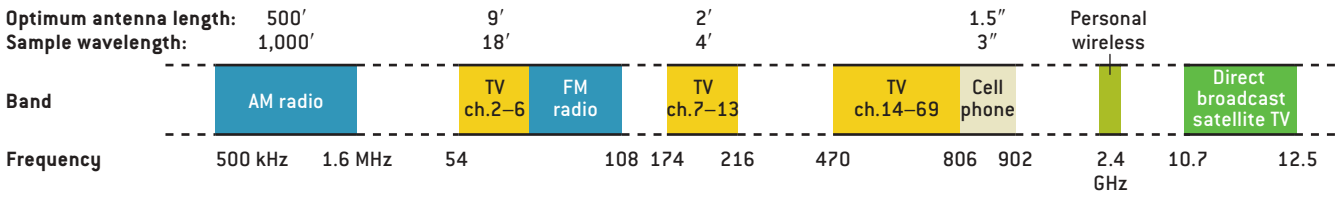
RABBIT EARS
can be extended or retracted to match full, half- or quarter-wavelengths for each TV channel. The V shape offers two elements that can be rotated to face different broadcast towers and angled to minimize interference from waves that arrive out of phase after hitting obstructions.

ROOFTOP TV
antenna optimizes reception of various channels with different elements. A five-foot element will best pick up a 10-foot wave. Longer elements add to its gain by reflecting the wave, and shorter elements help by reradiating the wave.





SATELLITE DISH
concentrates TV waves onto a low-noise block converter inside the feedhorn. In older, parabolic designs, the feedhorn and struts block some of the waves and cause interference, reducing reception efficiency to 50 to 55 percent. In new, offset designs, the parts lie below the dish, raising efficiency to about 80 percent.



RADIO-FREQUENCY SPECTRUM
in the U.S. is allocated by the Federal Communications Commission. Generally, the optimum antenna length for an application is one half the wavelength at a given frequency. A quarter-wavelength antenna is effective when mounted over a large section of reflective metal—such as a car. These lengths resonate most efficiently, transferring more wave energy to the receiver.

Desert Metropolis

NAMIBIA'S ENDLESS ARID EXPANSES ARE HOME TO A MENAGERIE OF CREATURES THAT LIVE NOWHERE ELSE **BY GARY STIX**

Before us stand the skeletons of camelthorn acacia trees that flourished before Vasco da Gama made his way around the Cape of Good Hope at the end of the 15th century. The cracked whitish basin, or pan, on which the acacias are still rooted is appropriately named Dead Vlei. On three sides, reddish-orange sand mountains rise as high as 300 meters. Our group is in the midst of Namibia's dune sea, more than a 400-kilometer drive southeast from the coastal city of Swakopmund. The Namib Desert extends in a strip 2,000 kilometers long and up to 150 kilometers wide down the southwestern African coast.

CORPSES of 600-year-old acacia trees populate the pan known as Dead Vlei, which is surrounded by looming reddish-orange sand dunes—a hallmark of the Namib Desert.

The sand heaps that tower immediately over us are star dunes. The wind blowing from multiple directions gives them right-angled ridges. When viewed from an airplane or a balloon, they assume the namesake shape. If you've seen one dune, you haven't seen them all. Namib cousins of star dunes bear names like parabolic, transverse and Barchan (crescent-shaped). According to comparative dunologist Nicholas Lancaster, a research professor at the Desert Research Institute in Reno, Nev., satellite imagery reveals that the same types of dunes can be found in Saudi Arabia and southern California, in addition to Namibia.

The surreality of Dead Vlei's parched flat is a magnet for international film crews; a bottled-water commercial was shot on the pan. We walk over paw prints, perhaps of a brown hyena that trod here sometime in the past several years when a summer rain turned the surface to mud. Suzie Van Der Walt, the resilient Afrikaner who is the guide on our 12-day cross-country camping safari, which took place last August (the Namibian winter), tells me that the dune at Dead Vlei's southern end is known as Crazy Dune.

"Why?" I ask.

"Because anyone who wants to climb it has got to be crazy," she says.

Hundreds of species that have never ventured beyond the desert's borders call the Namib home. Walking to Dead Vlei, we had paused to observe a bluish-green six-legged creature that moved with astonishing speed through a valley, an indentation formed by the impress of a tourist's footprint. A tenebrionid beetle, *Onymacris rugatipennis,* it buried itself one moment and then rushed forth again, as if skating along the surface of infinite sand grains. Its agitated movements occurred so quickly that I almost didn't have time to snap a picture.

The beetle sports a waxy coating on its elytra, or wing coverings, which keeps its bodily fluids from drying out in temperatures that exceed 40 degrees Celsius in January and other summer months. Namibia is full of discoveries that have never ranged outside of southwestern Africa, including the dancing white lady spider (*Leucorchestris arenicola*), Grant's golden mole (*Eremitalpa granti*) and the *Welwitschia mirabilis* plant, which can live more than 1,000 years. Some months before our journey, working on Namib-

MANFRED DELPHO

STARS OF THE NAMIB include the *Welwitschia* plant (*above*), the !Nara bush (*top right*), which is a delicacy for the oryx, and Grant's golden mole (*bottom right*).

PETER JOHNSON *Corbis* (Welwitschia); JOE McDONALD *Corbis* (*!Nara and oryx*); MICHAEL AND PATRICIA FOGDEN *Corbis* (*Grant's golden mole*)

ia's highest peak, several hundred kilometers to the north, an international team of researchers found a living member of the first new insect order discovered since 1915 [see "Gladiators: A New Order of Insect," by Joachim Adis, Oliver Zompro, Esther Moombolah-Goagoses and Eugène Marais; SCIENTIFIC AMERICAN, November 2002].

Another of the 650 species of tenebrionids that inhabit Namibia, *Onymacris unguicularis,* has become an international novelty because of its ingenuity. The same fogs that have caused shipwrecks for centuries, giving a section of the Namibian shoreline the name of Skeleton Coast, have shaped insect morphology and behavior. When the mist arrives, *O. unguicularis* tilts its body forward into the wind, and moisture rolls down its fused elytra into its mouth. In a 2001 *Nature* article British researchers described the structure of water-attracting bumps on Namib tenebrionid beetles that could inspire the design of water-collecting tents.

Not far from where *O. rugatipennis* appeared resided another icon of the Namib. This spiky green thornbush, a member of the cucumber family, holds a central place in the lives of the insects, animals and humans who inhabit the desert. The !Nara (the exclamation mark denotes a clicking sound in the Nama language) bush grows a long taproot that stretches down many meters to the water table. Its fruit, the !Nara melon, is a staple of Namib species at every rung of the food chain.

After Dead Vlei, we head north. A few days later, in Etosha National Park, about 560 kilometers northeast of Swakopmund, we stare at two male giraffes standing 15 or 20 meters from our Toyota Land Cruiser. (In the Lozi language the vehicle name, "Mutosi," means "aardwolf," a type of hyena that eats termites.) "I've got $5 on the one on the right," my son, Benjamin, exclaims. My family and the other eight members of the safari are spectators of a boxing match between the roughly five-meter-tall ungulates. As we watch, the slightly smaller animal bends his endless neck and swings his head up toward the torso of his companion. The assaulted animal just stands there. "I think the bigger one is challenged by this pipsqueak he's ignoring," remarks guide Van Der Walt.

These skirmishes are important events in the life of a giraffe. The victor of such a battle is the one that gets the girl. The little one continues his provocations, sometimes attacking the front of his competitor, sometimes bending so far that he comes up underneath his opponent, aiming for a sensitive part. "That's gotta hurt," Benjamin observes. The big one remains unfazed, standing his ground and occasionally making a distracted swipe to show who's boss.

Etosha is a world-renowned game reserve; it and the dune sea of the Namib are Namibia's two most famous attractions. Etosha, meaning "great white place," is a reference to Etosha Pan, the enormous planar saline desert that occupies more than a fifth of the park. Other areas of the reserve are a mosaic of grassland and bush. The park is home to 114 species of mammals, 340 kinds of birds, 111 types of reptiles, 18 species of amphibians and several kinds of fish. Among this lot are cheetahs, lions, leopards, rhinoceroses, spotted hyenas and elephants.

A visitor can spend hours watching the goings-on at the Okaukuejo Rest Camp, one of three overnight places within the park. Okaukuejo's water hole can be observed from benches behind a wall front-

ed by a sloping wire fence. The happenings at the floodlit pool resemble a scripted performance. One evening at about 10 P.M., a black rhinoceros grabs the stage, lumbering back and forth in an aimless march in front of the diminutive drinking spot. Except for the occasional jackal flitting by, the rhino has the pond to himself. After 20 minutes or so, the animal exits stage left. In the distance, we can just make out the dark silhouettes of a line of Burchell's zebras, which soon arrive in single file. Once sated, the zebras leave as a herd, making way for the giraffes. One animal approaches the water, and its front legs splay to each side as it executes the difficult maneuver of getting its mouth from firmament down to hoof level. The giraffes drink and depart. Then, emerging from the blackness, immense shadows. The elephants are here. A herd of about 10 lumbers forth, including a baby fiddling its useless trunk as it tries to imitate the way its mother hoses water into her mouth.

Our family and safari group sit with tourists who have come to Namibia from all over the world, with Etosha billed as a highlight of their visit. Etosha encompasses about 22,900 square kilometers, larger than the state of Israel. But to some of those who have known Etosha for decades, the scene at Okaukuejo has lost its allure. It is more natural than a zoo. Predators still seek out and kill prey. In fact, a tourist who slept out next to the wall was eaten by lions that scaled the barrier. But water is pumped into the hole throughout the year, attracting animals when it might otherwise be dry, and overgrazing has substantially depleted the surrounding veld vegetation.

Hu Berry, formerly Etosha's chief biologist, thinks the deterioration is unhealthy. It could promote disease, such as outbreaks of anthrax, which is endemic to the area. (The U.S. Navy has even used the park to perform testing on detectors that probe for the bacterium lodged in the soil or diagnose the disease in animals.) To Berry's eye, the parade of animals has the appearance of a circus act. "It's almost like a mechanical rhino is replaced by the next rhino," he says.

Namibia's Ministry of Environment and Tourism has undertaken a review to determine what limits should be placed on visitor growth. "The tourist-carrying capacity is something we're very aware of and would like to control to some extent, because it's the only way to ensure long-term sustainability," says Pauline Lindeque, the ministry's director of scientific services. She notes, however, that the water holes at the camps will remain open—the animals won't be rotated to different drinking spots—because they are critical to drawing tourists. Okaukuejo's denuded vegetation, she argues, remains relatively localized around the water hole.

According to Berry, who serves as a guide for high-end tourists, the water and road infrastructures of Etosha were never built for the volume of visitors that the park now receives. Some 320,000 holiday visitors came to Namibia in 2001, a significant increase from a decade earlier, when the country had just won independence and the memories of fighting between the South West African People's Organization and the South African Defense Force were still fresh. "Namibia is a very harsh country, as you know," Berry says. "But the soil and the vegetation are very fragile. You've got to be careful. If you destroy it, it'll take a lifetime, if you're lucky, to recover." It is most likely a problem that the country will have to live with. Its stark beauty—and now stable political environment—means that visitors will continue to throng to this enticing corner of southern Africa.

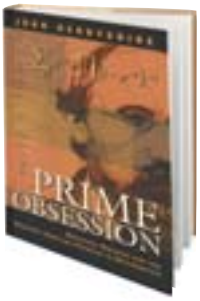
For those who wish to visit Namibia, Africa Tourist Info, based in the capital city of Windhoek, specializes in customized tours at all price ranges throughout the country and the rest of southern Africa (from the U.S., call 011-264-61-228717 or check out www.infotour-africa.com/office.htm). U.S.-based firms that offer Namibian and African tours include Adventure Center (800-228-8747; www.adventure-center.com) and Premier Tours (800-545-1910; www.premiertours. com). SA

OKAUKUEJO WATER HOLE presents magnificent game viewing. Visitors include elephants, giraffes and rhinoceroses. But the surrounding area has been overgrazed.

MARCELLO CALADRINI *Corbis*

Math's Most Wanted

A TRIO OF BOOKS TRACES THE QUEST TO PROVE THE RIEMANN HYPOTHESIS **BY KRISTIN LEUTWYLER**



PRIME OBSESSION: BERNHARD RIEMANN AND THE GREATEST UNSOLVED PROBLEM IN MATHEMATICS
by John Derbyshire
Joseph Henry Press, Washington, D.C., 2003 ($27.95)

THE RIEMANN HYPOTHESIS: THE GREATEST UNSOLVED PROBLEM IN MATHEMATICS
by Karl Sabbagh
Farrar, Straus and Giroux, New York, 2003 ($25)

THE MUSIC OF THE PRIMES: SEARCHING TO SOLVE THE GREATEST MYSTERY IN MATHEMATICS
by Marcus du Sautoy
HarperCollins, New York, 2003 ($24.95)

The unpredictable drip from a leaky faucet can drive almost anyone mad. Prime numbers, those divisible only by one and themselves, present a numerical equivalent. For centuries, mathematicians have tried to find a simple formula to describe where these numbers fall along the number line. But their spacing—1, 2, 3, drip, 5, drip, 7, drip, drip, drip, 11, drip, and so forth—seems to defy prediction. In 1859 German mathematician Bernhard Riemann uncovered an apparent key to unlocking the pattern, but he couldn't verify it. Many great minds have become obsessed with proving his guess, referred to as the Riemann Hypothesis (RH), ever since.

Three books published in April chronicle this quest. The books cover much of the same ground, but each has a different strength. The text with the simplest title, *The Riemann Hypothesis,* by science writer Karl Sabbagh, provides ample hand-holding for anyone who pales at the sight of symbols or can't quite distinguish an asymptote from a hole in the graph. In *Prime Obsession,* by John Derbyshire, a mathematically trained banker and novelist, Riemann and his colleagues come to life as real characters and not just adjectives for conjectures and theorems. And in *The Music of the Primes,* written by University of Oxford mathematics professor Marcus du Sautoy, the meaning of Riemann's work unfolds by way of rich musical analogies.

Why three books on the same difficult subject now? One obvious answer is that the notoriety of the RH only recently spread to circles beyond math-faculty common rooms. In 2000 the Clay Mathematics Institute (CMI), a private research organization funded by Boston banker cum math fan Landon T. Clay, offered a $1-million prize for the solution. The move won Riemann almost as many posthumous headlines as Fermat. CMI offers the one-buck bounty on seven outstanding mathematical mysteries. These so-called millennium problems are a 21st-century follow-up to German mathematician David Hilbert's famous stumpers, presented in 1900 to the Second International Congress of Mathematicians in Paris.

The Riemann Hypothesis is the only problem to make both lists, a century apart—and with good reason: it is exceedingly complex, and a mounting number of results require that it be true. Timing, too, has played a part. At the end of the 18th century Carl Friedrich Gauss, one of Riemann's mentors, produced what was then the best approximation for the number of primes less than some number N—namely, $N/\log N$. This value is sometimes too big and sometimes too small, but Gauss predicted that the error would shrink for larger Ns. By the end of the 19th century Jacques Hadamard and Charles de la Vallée Poussin proved this suggestion, called the prime number theorem (PNT). The RH was the next obvious mark.

Riemann's original wording does not mention prime numbers at all but instead addresses the so-called zeta function, $\zeta(s) = 1 + 1/2^s + 1/3^s + 1/4^s + \ldots 1/n^s$. For $s = 1$, this function is the familiar harmonic series. For inputs greater than one, however, zeta becomes more exotic. Swiss mathematician Leonhard Euler discovered in the 1700s that for $s = 2$, zeta converges on the square of pi divided by six. It was a startling find. The decimal expansion of pi is unpredictable, and yet by way of the zeta function, it could be summed from an infinite series of neat fractions. Euler's break was the first such "zeta bridge" be-

tween seeming randomness and order.

Riemann forged the next by feeding the zeta function complex numbers, those of the form $a + bi$, having both real and imaginary parts. These numbers were a new invention at the time. Riemann had learned about them in Paris and brought them back to Göttingen, where he studied under Lejeune Dirichlet, Gauss's successor. The older man was well acquainted with the zeta function, which he had invoked to prove one of Fermat's prime-number assertions. For Riemann, then, it was a small leap to try the new numbers in the old function.

To sum up what these books take 300-plus pages to explain, Riemann homed in on points for which the zeta function fed with imaginary numbers equaled zero and viewed these "zeros" as waves—much as Euler had produced sine waves corresponding to musical notes from plugging imaginary numbers into the exponential function 100 years before. Riemann further made a connection between these waves and his own refinement of Gauss's PNT, dubbed R(N): by adding R(N) to the height of each wave above N, he could generate the exact number of primes less than N. The location of the zeros, therefore, led to that of the primes, and Riemann asserted that the zeros followed a simple pattern. They all had a real part of $\frac{1}{2}$. In other words, were you to graph zeta, the zeros would fall along a single line.

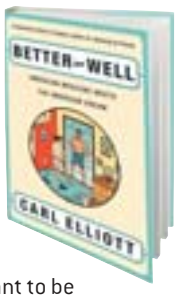
Each of the books satisfactorily presents Riemann's math—as much as it is possible to do so for a general audience—but they offer very different reading experiences. *The Music of the Primes* made me feel as if I were sitting through a gracefully worded lecture. *The Riemann Hypothesis* is more journalistic, relying on quotes from working mathematicians to tell the story. Parts of *Prime Obsession* read almost like a novel, others like a mathematical text. Its author, Derbyshire, segmented the book so that most of the math falls into odd chapters and the history and biographical material in even ones, but the math is as interesting as the rest. When will the RH be solved? None of the books dares to predict. Hilbert, one of the greatest mathematicians of all time, forecasted that it would happen within his lifetime. He died in 1943. In other words, it's still anyone's guess. SA

Kristin Leutwyler turned from the study of mathematics to journalism, serving until recently as editor of Scientific American*'s Web site. Now a freelance writer, she is the author of the forthcoming book* The Moons of Jupiter *(W. W. Norton, 2003).*

THE EDITORS **RECOMMEND**

BETTER THAN WELL: AMERICAN MEDICINE MEETS THE AMERICAN DREAM

by Carl Elliott. W. W. Norton, New York, 2003 ($26.95)



Elliott's springboard is enhancement technologies—Prozac, liposuction, cosmetic surgery and other drugs and procedures that aim not at curing illness but at making healthy people better endowed in some way. He uses the springboard to reach a larger subject: "the nature of identity, at least as we have come to understand identity in the contemporary West." The connection is the ambivalence of attitudes toward enhancement technologies. Users see them as a means to self-improvement. Others worry that people will use the technologies to conform to a narrow cultural ideal: "Everyone will want to be young, white, thin, smart, athletic, and good-looking in a very conventional, Hollywood sort of way." And what is the appeal of enhancement? "We are compelled to pursue fulfillment through enhancement technologies not in order to get ahead of others, but to make sure that we have lived our lives to the fullest." Elliott, professor of philosophy and bioethics at the University of Minnesota, grips the reader's attention all the way.

THE PATH: A ONE-MILE WALK THROUGH THE UNIVERSE

by Chet Raymo. Walker & Company, New York, 2003 ($23)

"For thirty-seven years I have walked the same path back and forth each day from my home in the village of North Easton, Massachusetts, to my place of work, Stonehill College. The path takes me along a street of century-old houses, through woods and fields, across a stream, along a water meadow, and through an old orchard and community gardens." Raymo, professor emeritus of physics and astronomy at Stonehill College and a science writer at the *Boston Globe,* walks with an observant eye and a ruminative mind. The stream, which in the 19th century powered the machines of the Ames Shovel Company, leads him into a discussion of gravity. Similarly prompted by what he sees, Raymo discusses engagingly such topics as photosynthesis, geology and evolution. The path so intimately familiar to him runs for barely more than a mile, "but the territory it traverses is as big as the universe."

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

Bounded Regrets

BY DENNIS E. SHASHA

Suppose you've taken up the game of squash and want to play at a health club. The club has two payment plans: $400 for a yearly membership or $20 for each use of the squash court. You'd like to play several times a week, but you're prone to injury and know you won't be able to play anymore if you get hurt. An oracle would be able to predict if and when you get injured and thus determine whether it would be more cost-effective to buy the yearly membership or pay for each use of the court. You are not clairvoyant, however, so you wish to devise a strategy for minimizing your "regret ratio"—the amount you spend divided by the amount that the oracle would have spent.

If you decide to buy a yearly membership right from the start and then get injured on your first day on the court, your regret ratio would be 20—the $400 you spent divided by the $20 that the oracle would have spent. If you decide to pay for each use for an entire year and you play 100 times and then get injured, your regret ratio would be 5—the $2,000 you spent divided by the $400 that the oracle would have spent. Is there a way to keep your regret ratio below 2 no matter when you get hurt? The answer to this warm-up puzzle is at the bottom of the page.

In mathematicians' lingo, this kind of problem is called competitive analysis. Let's consider another example: Say you have 90 tickets that you can trade for cash. There is a ticket exchange booth in which a man has a pile of $1 and $5 bills. When you approach the booth, the man inside will offer the bill at the top of the pile—either $1 or $5—for each of your tickets. You have the option of rejecting a $1 offer in the hope that the man will later proffer $5 for the ticket. If you do so, you keep the ticket and the man puts the $1 bill aside, never to be seen again. But the man has the right to halt the trading at any time, and after that point the tickets are worth nothing.

An oracle would know in advance what the man will offer and when the exchange will end, but you don't. Can you find a strategy that will guarantee a regret ratio (in this case, the oracle's winnings divided by your winnings) that is no more than 1.8? And what would be your strategy if the two possible offers are $1 and $1 million? Does the regret ratio improve or worsen? SA

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

Answer to Last Month's Puzzle

In the first problem, BEF derives directly from AB and EF; DEF from DH and EF; ADE from AB and DEF; ACD from ADE and CG; ACF from ACD and EF; ADG from ACD and CG; and BEG from BEF and CG. FGH requires three progenitors.

In the second problem, four alternative source species are BE, DF, AC and GH.

Web Solution

For a fuller explanation of last month's solution and a peek at the answer to this month's problem, visit www.sciam.com

ANSWER TO WARM-UP PUZZLE: Pay for each use 19 times, then buy a yearly membership. If you injure yourself during the first 19 outings, your regret ratio will be 1. If you injure yourself afterward, your regret ratio will be 1.95.

MATT COLLINS

Doing What Comes Unnaturally

FROM SHEEP TO SHEEPSKINS IN THE FIELD OF GENES **BY STEVE MIRSKY**

Dolly is dead. (For now, anyway.) Dolly, of course, was the world's most famous sheep, the first mammal to be cloned from an adult animal, the creature that could have answered poet William Blake's previously rhetorical question "Little lamb, who made thee?" by pointing a hoof at her creator, scientist Ian Wilmut.

The six-year-old Dolly was put down in February because of a lung infection (a common enough ailment among sheep as to make her clonehood possibly not a factor). News stories referred to Dolly as middle-aged, but the average life span of sheep is still a subject of debate, compounded by a reality that Dolly was blissfully unaware of: "Nine months and then we eat them," noted Wilmut to a reporter.

Press coverage also had scientists saying variations of "no one thought cloning was possible before Dolly," which is just silly: Wilmut clearly thought it was possible, as did the writers of *Jurassic Park,* in which cloned dinosaurs run amok, and the even earlier *Boys from Brazil,* in which cloned Hitlers run amok. Dolly, on the other hand, battled obesity throughout her life; real clones thus seem more likely to stroll amok, amble amok or just lie around in their own muck amok.

Meanwhile Ted Williams is *still* dead. The great Red Sox slugger remains frozen in a cryogenics facility, put there by his son, John Henry, in the hopes that the old man's genetic material might one day be harvested for cloning. And you do want every last gene: John Henry has half of Ted's genes, but in a short minor-league career he wasn't close to being half the ballplayer. Note to John Henry: when Hall of Famer Ernie Banks famously said, "Let's play two," he meant "Let's play a doubleheader," not "Let's play two identical guys in left field and right field."

The Williams case, which stinks on ice, was back in the news in February because a friend of Ted's was able to finagle a visit to the cryogenics lab in question, located in Scottsdale, Ariz., where

baseball players who are still alive were gathering for spring training. The friend was disturbed to discover that Ted actually shares a tank with three other bodies as well as five heads, not exactly the starting nine Williams was used to. Nevertheless, maybe one day a spanking new Ted Williams clone, perhaps with additional genetic engineering and bionic implants, will play left field for the Red Sox. I was going to say that he might even help them win the World Series, but I don't want to strain credulity.

What's so unnatural about cloning, of course, is that the next "generation" is genetically identical to its predecessor. Genetic variation followed by selection—better known as evolution—is the norm. This idea really riles creationists, who invest a great deal of energy in trying to drive evolution education from classrooms in favor of their nonscientific creed. Part of the effort involves swaying public opinion by circulating antievolution statements signed by a few dozen holders of Ph.D.s in various sciences, giving the impression that the scientific community is deeply divided about the fact of evolution. In February the National Center for Science Education (NCSE), a nonprofit based in Oakland, Calif., wisecracked back with a statement in support of evolution education, which was signed by more than 200 holders of science doctorates—and, in a clear case of intelligent design, if not cloning, every signatory was named Steve.

The name Steve was chosen as an homage to the late Stephen Jay Gould, who would have been the first to point out that scientific truth is not decided by petition. But with Steves making up about 1 percent of the population and with the number of signatories up to 287 at last count, a reasonable extrapolation is that at least 28,700 scientists, almost enough to fill Fenway Park, would support the NCSE statement. And that number should strike out the notion that the scientific community is divided over evolution, or my name's not, well, Steve. SA

ILLUSTRATION BY MATT COLLINS; PHOTOGRAPH BY FRANK VERONSKY

ASK THE EXPERTS

Q Why do computers crash?

—R. L. Feigenbaum, Croton-on-Hudson, N.Y.

Clay Shields, assistant professor of computer science at Georgetown University, explains:

The short answer is: for many reasons. Computers crash because of errors in the operating system (OS) software or the machine's hardware. Software glitches are probably more common, but those in hardware can be devastating.

The OS does more than allow the user to operate the computer. It provides an interface between applications and the hardware and directs the sharing of system resources among different programs. Any of these tasks can go awry. Perhaps the most common problem occurs when, because of a programming flaw, the OS tries to access an incorrect memory address. In some versions of Microsoft Windows, users might see a general protection fault (GPF) error message; the solution is to restart the program or reboot the computer. Other programming mistakes can drive the OS into an infinite loop, in which it executes the same instructions over and over. The computer appears to lock up and must be reset. Another way things can go amiss: when a programming bug allows information to be written into a memory buffer that is too small to accept it. The information "overflows" out of the buffer and overwrites data in memory, corrupting the OS.

NINA FINKEL

Application programs can also cause difficulties. Newer operating systems (such as Windows NT and Macintosh OS X) have built-in safeguards, but application bugs can affect older ones. Software drivers, which are added to the OS to run devices such as printers, may stir up trouble. That's why most modern operating systems have a special boot mode that lets users load drivers one at a time, so they can determine which is to blame.

Hardware components must also function correctly for a computer to work. As these components age, their performance degrades. Because the resulting defects are often transient, they are hard to diagnose. For example, a computer's power supply normally converts alternating current to direct current. If it starts to fail and generates a noisy signal, the computer can crash.

The random-access memory (RAM) can err intermittently, particularly if it gets overheated, and that can corrupt the values the RAM stores at unpredictable times and cause crashes. Excessive heat can crash the central processing unit (CPU). Fans, which blow cooling air into the computer's case, may fail, making components susceptible to overheating. And they push dirt and dust inside, which can lead to intermittent short circuits; compressed air or a vacuum cleaner easily gets rid of such dirt. Still other hardware problems, including a failed video or network card, are trickier to identify, requiring software tests or the sequential replacement of components.

Errors on a computer's hard drive are the most intractable. Hard disks store information in units called sectors. If sectors go bad, the data stored on them go, too. If these sectors hold system information, the computer can seize up. Bad sectors also can result from an earlier crash. The system information becomes corrupted, making the computer unstable; ultimately the OS must be reinstalled. Last and worst, a computer can fail completely and permanently if the machine gets jarred and the head that reads information makes contact with the disk surface.

Q What causes thunder?

—Tom Blighes, San Antonio, Tex.

Richard C. Brill, professor at Honolulu Community College, offers this answer:

Thunder is caused by lightning, which is essentially a stream of electrons flowing between or within clouds or between a cloud and the ground. The air surrounding the electron stream becomes so hot—up to 50,000 degrees Fahrenheit—that it forms a resonating tube of partial vacuum surrounding the lightning's path. The nearby air rapidly expands and contracts, making the column vibrate like a tubular drumhead and producing a tremendous *crack*. As the vibrations gradually die out, the sound echoes and reverberates, generating the rumbling we call thunder. We can hear the booms from great distances, 10 or more miles from the lightning that caused them. SA

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