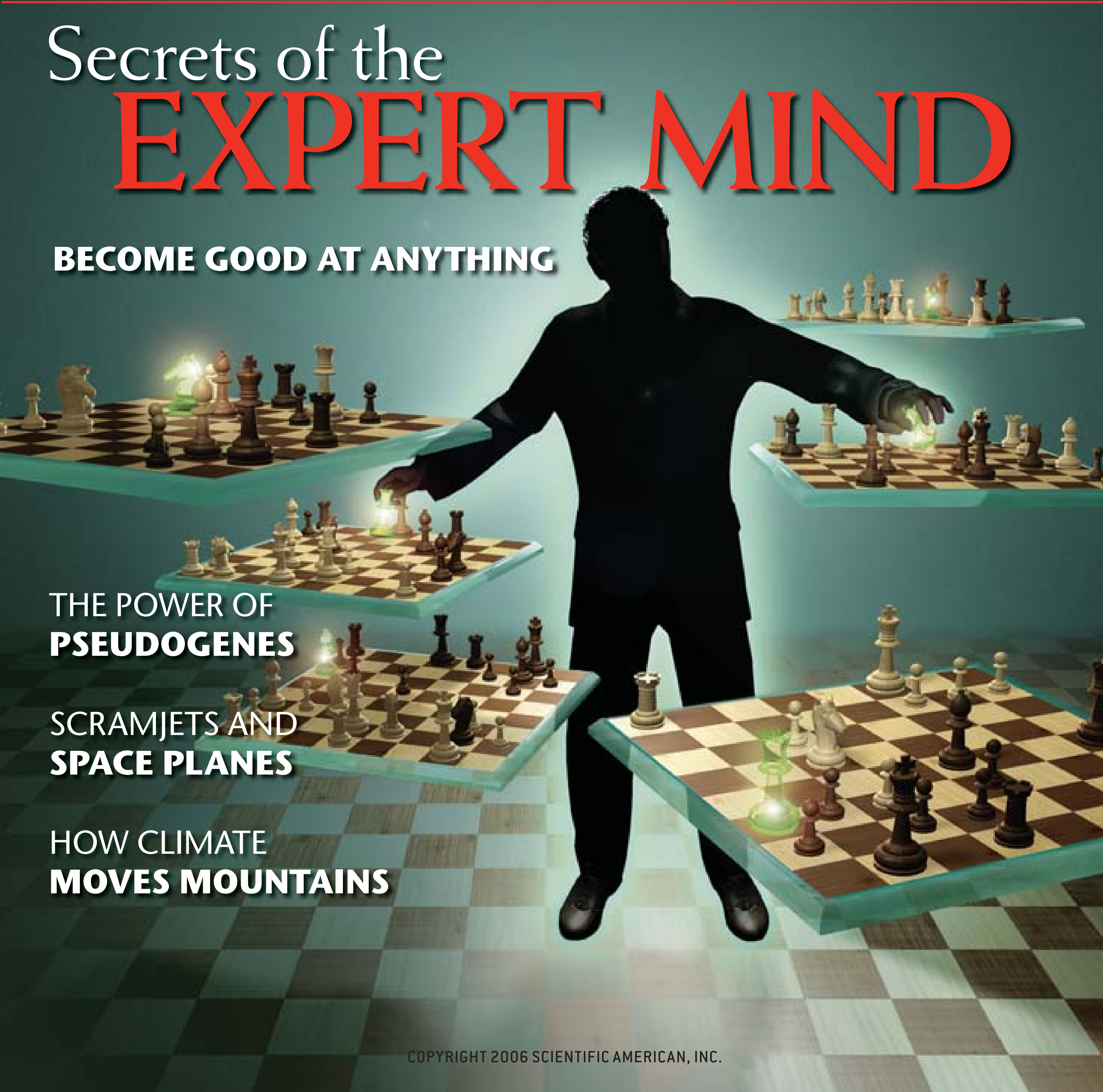
HOW SALMON FEED FORESTS • RAY-TRACED REALITY

SCIENTIFIC AMERICAN

The **Strangest** Moons

AUGUST 2006
WWW.SCIAM.COM

Secrets of the EXPERT MIND



BECOME GOOD AT ANYTHING

THE POWER OF **PSEUDOGENES**

SCRAMJETS AND **SPACE PLANES**

HOW CLIMATE **MOVES MOUNTAINS**

august 2006 contents features

SCIENTIFIC AMERICAN Volume 295 Number 2

ASTRONOMY

40 The Strangest Satellites in the Solar System

BY DAVID JEWITT, SCOTT S. SHEPPARD AND JAN KLEYNA

With peculiar orbits that often move against the grain of the rest of the solar system, an odd breed of planetary satellites is reshaping ideas about the formation of the solar system.

LIFE SCIENCE

48 The Real Life of Pseudogenes

BY MARK GERSTEIN AND DEYOU ZHENG

Disabled genes, once dismissed as detritus on the genomic landscape, trace the path of evolution—and may not always be entirely dead.

AEROSPACE

56 Power for a Space Plane

BY THOMAS A. JACKSON

Creating a revolutionary hypersonic jet engine that could propel a space plane to orbit affordably and routinely is a tough but seemingly achievable task.

COVER: PSYCHOLOGY AND BRAIN SCIENCE

64 The Expert Mind

BY PHILIP E. ROSS

The mental processes of chess grandmasters are unlike those of novices, a fact that illuminates the development of expertise in other fields.

EARTH SCIENCE

72 Climate and the Evolution of Mountains

BY KIP HODGES

New studies of the Himalaya and the Tibetan Plateau suggest that climate and geology can be partners in a long, slow dance.

INFORMATION TECHNOLOGY

80 A Great Leap in Graphics

BY W. WAYT GIBBS

Soon even home computers should be able to produce quick, high-quality 3-D graphics, thanks to speedier new ways to simulate the flight of light.

ECOLOGY

84 The Fish and the Forest

BY SCOTT M. GENDE AND THOMAS P. QUINN

Salmon carcasses left behind by predatory bears are unexpectedly important sources of nutrients for forests.

40 Misbegotten moon

departments

8 SA Perspectives
Keep the Internet neutral.

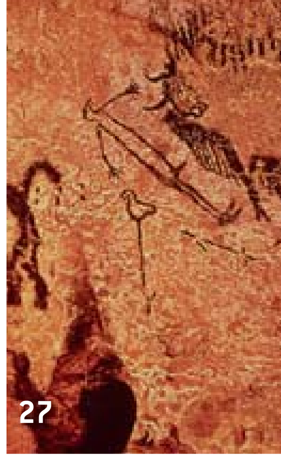
10 How to Contact Us

10 On the Web

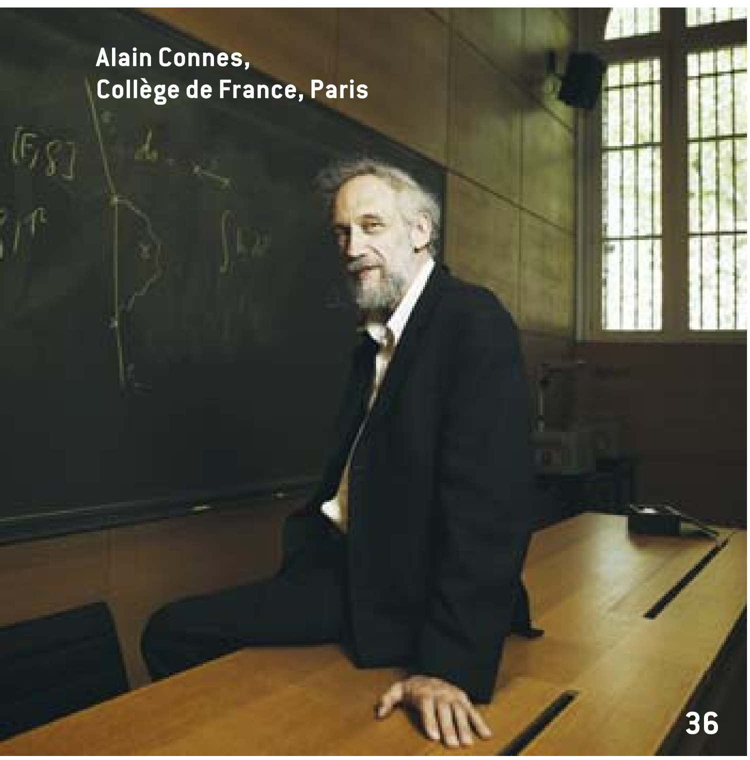
12 Letters

16 50, 100 & 150 Years Ago

18 News Scan
- For the World Cup, carbon trading goes for the gooooal!
- Fermilab finds a new focus.
- Genetic bar coding of the ocean's deep dwellers.
- A seismic space probe orbits Venus.
- Why the EPA will not regulate a harmful pesticide.
- Cave art as graffiti.
- By the Numbers: Methamphetamine use.
- Data Points: Ice-core data and global warming.


27


100


Alain Connes, Collège de France, Paris
36

36 Insights
Acclaimed mathematician Alain Connes takes a geometric approach to learn how spacetime makes particles.

39 Forum: Mihail C. Roco
Passing through four evolutionary stages, nanotechnology will become a $1-trillion industry by 2015.

90 Working Knowledge
How disk-drive makers raise storage capacities by making bits stand on end.

92 Technicalities
Here's the five-day forecast for your backyard.

96 Reviews
The play's the thing—even when theater tackles scientific concepts.

columns

34 Skeptic BY MICHAEL SHERMER
Why our intuitions about science are so often wrong.

35 Sustainable Developments
BY JEFFREY D. SACHS
The care and feeding of unstable new democracies.

98 Anti Gravity BY STEVE MIRSKY
New York, N.Y.: Nothing to see, move along.

100 Ask the Experts
How do space probes navigate such large distances accurately? What causes ringing in the ears?

Cover image by Jean-Francois Podevin; photograph at left by Nicolas Guerin, Azimuts Productions.

Scientific American (ISSN 0036-8733), published monthly by Scientific American, Inc., 415 Madison Avenue, New York, N.Y. 10017-1111. Periodicals postage paid at New York, N.Y., and at additional mailing offices. Canada Post International Publications Mail (Canadian Distribution) Sales Agreement No. 40012504. Canadian BN No. 127387652RT; QST No. Q1015332537. Publication Mail Agreement #40012504. Return undeliverable mail to Scientific American, P.O. Box 819, Stn Main, Markham, ON L3P 8A2. Subscription rates: one year $34.97, Canada $49 USD, International $55 USD. Postmaster: Send address changes to Scientific American, Box 3187, Harlan, Iowa 51537. **Reprints available:** write Reprint Department, Scientific American, Inc., 415 Madison Avenue, New York, N.Y. 10017-1111; (212) 451-8877; fax: (212) 355-0408. **Subscription inquiries: U.S. and Canada (800) 333-1199; other (515) 248-7684. Send e-mail to sacust@sciam.com** Printed in U.S.A.

Keep the Net Neutral

If the online universe has had an unofficial slogan to date, it might have been the caption to that famous cartoon by Peter Steiner: "On the Internet, nobody knows you're a dog." Not only do digital communications allow anonymity, but the underlying TCP/IP protocols that govern the flow of data are supremely egalitarian. Everybody's packets of information are treated equally by the routers. Thanks to that level playing field, entrepreneurs working out of their garages have been able to compete toe to toe with Fortune 500 companies in new businesses.

WHO WILL PAY for online video?

KRT

But with the rising popularity of streaming video and miscellaneous other services labeled "Web 2.0," some telecommunications companies are arguing that this model of "net neutrality" must change. Online video quality is relatively intolerant of even small transmission delays. AT&T, Verizon, Comcast and other companies that own the backbone lines for the Internet would like to prioritize data streams to make the traffic flow more rationally. If they have their way, the Internet's next slogan might borrow from George Orwell's *Animal Farm:* "All animals are created equal, but some animals are more equal than others."

The telcos propose "tiered service" for providers of Web content. Currently those providers pay just for the bandwidth they use, but the telcos also want to charge them a premium for guarantees that their data will get preferential treatment. The telcos argue that they will need to invest to handle the growing bandwidth demand. The only alternative to charging the content providers is to charge individual consumers more for access, which seems undesirable.

Critics see a catch. Companies that sign with the telcos, or the content arms of the telcos themselves, could have a huge advantage over their rivals—an antimeritocratic arrangement that would distort competition and handicap start-ups. In the most abusive situations, some Web sites would become virtually unusable. And of course, the expense of those extra fees will eventually get passed along to consumers anyway in higher costs for content.

On balance, those favoring net neutrality make the better case. A system for prioritizing data traffic might well be necessary someday, yet one might hope that it would be based on the needs of the transmissions rather than the deal making and caprices of the cable owners. Moreover, personal blogs and other Web pages are increasingly patchworks of media components from various sources. Tiered service would stultify that trend. If the costs for video are not to be universally shared, perhaps it will ultimately be fairer and more practical for individuals to pay for the valued data they receive.

Ending net neutrality might feel safer if the telcos did not often enjoy local monopolies on broadband service. Almost half of all Americans have limited or no choice if they want high-speed connections. That dearth of competition lowers incentives for the telcos to keep overall network service high.

In June the House of Representatives dealt net neutrality a blow by passing an overhaul of the Telecommunications Act of 1996 that specifically omitted any protections for it. The Senate is drafting its own sweeping telecommunications reform bill. To express your preferences on this important issue, contact your congressional representatives and consider signing one of the dueling petitions organized at SavetheInternet.org (favoring net neutrality) and HandsOff.org (against 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
212-451-8200
fax: 212-755-1976
or
editors@sciam.com

SUBSCRIPTIONS

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

REPRINTS

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

PERMISSIONS

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

ADVERTISING

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

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

Los Angeles
310-234-2699
fax: 310-234-2670

San Francisco
415-503-3630
fax: 415-437-2892

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

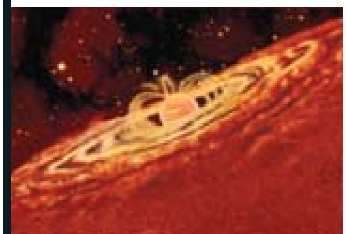
On the Web

WWW.SCIAM.COM

UPDATED EVERY WEEKDAY

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

NEWS: "Planemos" May Give Rise to Planets, Moons



WWW.JONLOMBERG.COM

The textbook account of where planetary systems come from—namely, the disks of dust and gas that encircle stars—may require an addendum. Researchers have found that comparable disks girdle distant exoplanets far less massive than our sun. The discovery raises the possibility that these planetary mass objects, or planemos, are orbited by other planets and moons—like our own solar system but on a smaller scale.

BLOG: Salvos Exchanged in Battle over Hobbit Brain

In a paper published online recently by *Science,* Robert D. Martin of the Field Museum and his colleagues make the case that the diminutive human skeleton from Flores popularly known as "the Hobbit" is not a new species of human but a modern human who suffered from a disorder that left it with an abnormally small brain—a condition known as microcephaly.

PODCAST

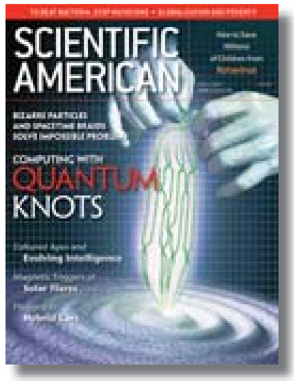
Visit www.sciam.com/podcast to download our free weekly audio program, *Science Talk*. Recent guests include Nobel Prize–winning neurobiologist Eric Kandel and Ed Cutrell of Microsoft Research's Adaptive Systems and Interaction Group.

Ask the Experts

How do batteries store and discharge electricity?
Kenneth Buckle, a visiting scientist at the Center for Integrated Manufacturing Studies at the Rochester Institute of Technology, enlightens.

Letters

EDITORS@SCIAM.COM



APRIL'S ISSUE TANGLED WITH quantum braids in "Computing with Quantum Knots," by Graham P. Collins. Readers also learned about new vaccines coming to market that promise to conquer childhood diarrhea caused by rotavirus, a frequent killer of young children in the developing world, in "New Hope for Defeating Rotavirus," by Roger I. Glass.

Most interesting were reader responses to another health problem, depicted by Madhusree Mukerjee, about public health scientist Smarajit Jana's work in organizing sex workers to fight HIV in India ["The Prostitutes' Union," Insights]. Some letters demonstrated that there are those who prefer to draw a sharply defined line between science and medicine, culture and politics—an impossible luxury when scientific knowledge is applied to ameliorating human ills. Robert L. Teeter e-mailed: "Do you really think that articles on prostitution are appropriate for what used to be a dignified and respected *scientific* magazine?" More opinions were closer to Andy Benton's, who wrote from Flourtown, Pa.: "Thank you for sharing Jana's story. He is truly a hero of our times."

CARPING ABOUT CONGRESS

Steve Mirsky's application of "tiny backbone living in corrosive swamp" in "Short Takes" [Anti Gravity] to describe both a kind of carp and any member of the House ethics committee surely deserves at the least a tiny Pulitzer Prize for its succinctness, completeness and brevity.

Harry Ison
Bellingham, Wash.

UNION BROTHELS

"The Prostitutes' Union," by Madhusree Mukerjee [Insights], betrays an irony: India, a country with one of the oldest and deepest-rooted organized religions in history, is perhaps the first to successfully apply a purely secular solution to a problem that may be older than religion itself. Imagine if the U.S. surgeon general came up with this idea. I think the firestorm of revulsion from the religious right would overwhelm all the public health considerations. And yet in India, religious leaders became part of the solution. How sad that we have not made life better for our own sex workers because of belief systems infected with pride. I guess that is the contradiction of "scientific" America.

Carey McConnell
Palm Harbor, Fla.

AERODYNAMIC AIR OUT

As magazine editors dedicated to educating the public about scientific and technological issues, you should not perpetuate misleading descriptions, as was the case in "Big Squeeze," by Mark Fischetti [Working Knowledge]. The article described how an airplane wing produces lift thus: "... because the wing top is curved, air streaming over it must travel farther and thus faster than air passing underneath the flat bottom." If so, how could a plane fly upside down with the flat surface on top and the curved surface on the bottom?

In fact, if a winglike object, such as a flat plate, is inclined with respect to the airflow (relative wind), the air will travel much faster over the surface away from the wind and slower over the other surface, giving rise to a pressure differential. This difference is dependent on the wing's angle of attack (angle of the wing with respect to the relative wind) but not on one surface being more curved than the other.

A better explanation of lift is: as a wing angled up with respect to the relative wind moves through a volume of air, it deflects downward a large amount of air above and below the wing. Thus, the wing exerts a downward force to accelerate this mass of air downward (Newton's second law). The air exerts an equal but opposite force upward on the wing (Newton's third law). This upward force by the deflected volume of air is the lift. (This force also produces drag.)

Klaus Fritsch
Department of Physics
John Carroll University
University Heights, Ohio

FISCHETTI REPLIES: Numerous readers wrote to correct a common but faulty explanation of how an airplane wing creates lift, noting that it has somehow persisted for years, even in textbooks. We wrote, "... because the wing top is curved, air streaming over it must travel farther and thus faster than air passing underneath the flat bottom. According to Bernoulli's principle, the slower air below exerts more force on the wing than the faster air above, thereby lifting the plane."

Or not. As Fritsch points out, the key factor is the wing's angle of attack, not its shape. As for the topside curvature of many wings, some readers noted that Bernoulli's principle can add a small amount of additional lift. Others pointed out that stunt planes and certain fighter aircraft have wings that are flat on top and bottom (or have equivalently curved surfaces) so they can better fly upside down. And yet the "third law" explanation is not the full story either: according to NASA, the complex "turning" of airflow, both below and above the wing, is the real driver. For a vetting of both the Newtonian and Bernoullian explanations, see www.grc.nasa.gov/WWW/K-12/airplane/bernnew.html

ROUTING ROTAVIRUS

In "New Hope for Defeating Rotavirus," Roger I. Glass indicated that two rotavirus vaccines, RotaTeq and Rotarix, have recently proved highly effective in clinical trials and are soon to be marketed in a number of countries. The RotaShield rotavirus vaccine has also been shown to be highly effective in clinical trials; his discussion on RotaShield is related only to its withdrawal from the market in 1999.

Now that new scientific evidence has demonstrated that RotaShield can be safely used in infants, BIOVIRx intends to bring RotaShield back to the market, subject to appropriate regulatory approvals. Our goal is to make it affordable, because we believe that is necessary for a rotavirus vaccine to have the greatest global impact on reducing morbidity and mortality.

VIRUS ON THE RUN: Rehydration therapy saved this infant from rotavirus-caused diarrhea. Within a decade new vaccines may conquer this deadly virus.

Leonard P. Ruiz, Jr.
President and CEO, BIOVIRx, Inc.
Shoreview, Minn.

QUANTUM NOTS

With regard to "Computing with Quantum Knots," by Graham P. Collins, is it possible to simulate a quantum computer on a conventional computer, at least in theory? If not, perhaps some of the difficulty in artificial intelligence is because biological systems have somehow discovered ways to use quantum effects.

John J. Boyer
GodTouches Digital Ministry
Madison, Wis.

COLLINS REPLIES: It is impossible to simulate a quantum computer efficiently *on a conventional classical computer. The simulation would require a vast supply of hardware to run algorithms in parallel, or it would take an extremely long time. Some researchers, most notably mathematician Roger Penrose, have speculated that biological brains might use quantum computation, but physicist Max Tegmark has argued that in the physical environment of a brain, the quantum coherence required would decay far too quickly to have any effect on the firing of neurons.*

ERRATA "An Antibiotic Resistance Fighter," by Gary Stix, noted incorrectly that mice did not develop resistance to the antibiotic ciprofloxacin. The article should have stated that the bacterium *Escherichia coli* did not become resistant. It also remarked that blocking the cutting of the protein LexA might undermine "drug effectiveness" in other microbes besides *E. coli.* Rather preventing the clipping of LexA might undermine the evolution of drug resistance in other microbes.

"Sharp Shooter," by Steven Ashley [Technicalities], stated that the APS sensor in the Sony R1 digital camera has a low signal-to-noise ratio. It has a high signal-to-noise ratio.

CLARIFICATIONS In "The Science behind Sudoku," by Jean-Paul Delahaye [June], the grid for puzzle *e* in the "Variations on a Theme" box should have included outlines for the domino pieces. Without the lines, the puzzle has two solutions. To find the correct version, go to www.sciam.com/ontheweb; visit that same site for extra puzzles and solutions to the grids in the article. (The address given in the article was incomplete.)

In Ask the Experts, Stephen M. Roth's answer to the question "Why does lactic acid build up in muscles?" stated that high lactate levels increase acidity in the muscle cells. Though associated with this condition, they are not the cause. For a more in-depth explanation, see www.sciam.com/ontheweb

GEOFF WARD

50, 100 & 150 Years Ago

FROM *SCIENTIFIC AMERICAN*

Little Neutral Particle ▪ Britain's Achilles' Heel ▪ Pickpockets Beware

AUGUST 1956

NEUTRINO FOUND—"A long and exciting adventure in physics has come to a triumphant end. The neutrino has been found. Frederick Reines and Clyde L. Cowan, Jr., of the Los Alamos Scientific Laboratory trapped the ghostly particle in an underground chamber near the Savannah River atomic pile. Phillip Morrison, in the January issue of *Scientific American,* compared the neutrino to the planet Neptune. The discovery of Neptune was a crowning achievement of classical physics: the motions of other planets showed it had to be there. The neutrino is a similar achievement of modern physics, and its discovery is a vindication of the law of the conservation of energy."

AUGUST 1906

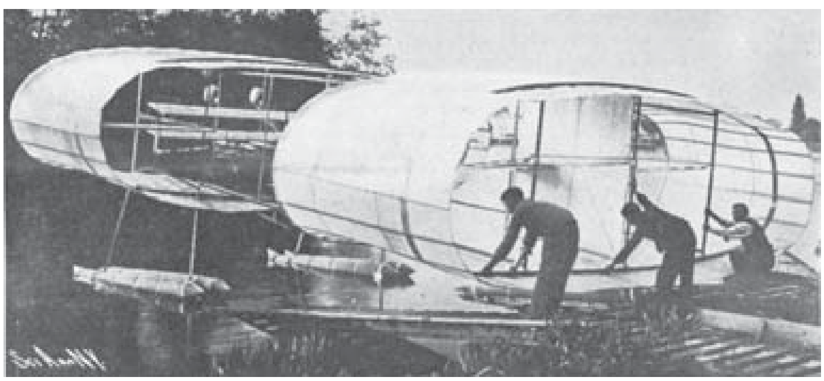
JAVA DEAD END—"Some years ago Eugène Dubois discovered in the island of Java some bones from a prehistoric animal, which might have formed the so-called missing link in the chain of descent of man from monkey. Julius Kollman is rather of the opinion that the direct antecedents of man should not be sought among the species of anthropoid apes of great height and with flat skulls, but much further back in the zoological scale, among the small monkeys with pointed skulls; from these he believes were developed the human pygmy races of prehistoric ages, with pointed skulls. Thus may be explained the persistency with which mythology and folk lore allude to pygmy people."

WAR GAME OF THE ATLANTIC—"So largely does Great Britain depend upon her over-sea commerce for food stuffs, that in time of war there would be no surer way of bringing that proud empire to its knees as a supplicant for peace than to capture, destroy, or drive from the high seas her merchant fleet. The recent naval maneuvers were planned with a view to determine just how great this peril might be. Although the swift cruisers of the 'raiding fleet' succeeded in doing considerable damage to the country's commerce, they were driven from the trade route which was selected for attack. The contention of the leading naval authorities is that Great Britain's commerce can never be so absolutely crippled as to decisively affect the issues of war."

CAUTIONARY TALE—"William S. Meade, who is said to have made a fortune of $250,000 in a process discovered by him for the preservation of meat, recently died in a New York lodging house, penniless. After he made his fortune, while on the Pacific coast, he befriended an old sea captain, who claimed to know the resting place of a sunken treasure boat, and upon the captain's death Meade was bequeathed a number of charts and directions in cipher for locating the craft. Meade's whole fortune was wasted in an effort to find this boat. At his own expense he sent out three expeditions. Two of them came to grief on the coast of South America, and the third was abandoned after cruising along the coast of Chile and Peru for several years in search of the treasure."

FLYING FERVOR—"What has greatly stimulated aeroplane work has been the founding of the Grand Prize of Aviation by the Aero Club of France. As regards the new aeroplane which Messrs. Bleriot and Voisin are constructing at their establishment in the suburbs of Paris, we present here a view of the machine [*see illustration*] during the first experimental flight upon the lake of Enghien. But a short flight was made, as it was found that some alterations were needed. Perhaps the most important of all recent French aeronautical craft is Santos-Dumont's aeroplane. The machine has been christened '14bis,' and has been constructed primarily with a view to competing for the $10,000 Deutsch-Archdeacon aeroplane prize."

SCIENTIFIC AMERICAN



BLERIOT AIRPLANE unveiled in France, 1906

AUGUST 1856

PICKPOCKET TRAP—"This contrivance externally resembles a watch in size and shape. Within the case is a bell and spring hammer, the latter connected with a fob chain. The supposition is, that the thief will pull the chain in order to obtain the prize. But instead of getting the watch, the *watch* gets him. The pull sounds the alarm bell, the owner of the watch grabs the rogue, and the policeman conducts him to limbo."

ENVIRONMENT

Soccer Goes Green

AT THE WORLD CUP, A NEW WAY TO OFFSET CARBON EMISSIONS **BY GUNJAN SINHA**

Soccer, beer and bratwurst were very likely the only things on fans' minds as they descended on Germany to celebrate the World Cup this June. But all that partying had a downside—pollution. One million soccer tourists consumed a lot of energy. Environmentalism is part of the German zeitgeist, so it is only fitting that the event had a "green goal," too. A consortium including FIFA, the international soccer federation, and the German football association DFB donated 1.2 million euros to make this year's play-off the first sporting event to offset its carbon dioxide emissions by investing in three renewable energy projects.


WORLD CUP 2006 brought in Brazilian, Croatian and other energy-hungry fans. Soccer officials hope to compensate for the emitted carbon by funding sustainable energy programs.

SEAN GALLUP *Getty Images*

Among the environmentally savvy, carbon-offset programs are the latest rage. At least a dozen companies offer the promise to mitigate greenhouse gas emissions from activities such as flying and driving and from events such as weddings and record releases. Voluntary offset programs, however, are not regulated, so consumers cannot be sure that they are investing in environmentally sound projects. What is more, reductions do not help much, because emissions from such activities are tiny: for instance, Germany's total carbon dioxide emissions are about 800 million tons per year—the World Cup emitted a mere 100,000 tons extra.

But governments are starting to pay heed to offsets. Europe established a cap-and-trade system last year that limits carbon dioxide emissions from about 50 percent of industry to reach its emissions goals as dictated by the Kyoto Protocol. Officials modeled the system on the sulfur dioxide trading market established in the U.S. in 1995, which has successfully cut levels of acid rain. As the trading market evolves, some environmentalists think that voluntary offset programs could join existing cap-and-trade market schemes to cut emissions even more substantially.

Right now, however, groups involved in voluntary projects are busy establishing credibility. "We wanted to serve as a model," says Christian Hochfeld of the Öko Institute in Berlin, an environmental think tank that developed the Green Goal project for the World Cup.

To that end, the institute chose projects that met criteria established by the World Wildlife Fund to better define high-quality development projects for industries affected by Kyoto. In Tamil Nadu, India, Women for Sustainable Development, a nonprofit organization, will oversee the installation of 700 to 1,000 biogas reactors—simple enclosed pits about the size of a well into which villagers dump cow dung. The fermenting mass generates gas, which fuels stoves and replaces kerosene. Two other sustainable energy projects will take place in South Africa. One will capture off-gas at a sewage treatment facility and burn it to supply electricity to Sebokeng, a township near Johannesburg. The other will replace a citrus farm's coal-fired heating system with one that burns sawdust—a by-product of wood processing usually discarded. The projects will offset all the soccer tournament's emissions.

World Cup organizers could have planted trees—as has been done by previous sporting events, such as the Super Bowl—or invested in other projects on home turf. But planting has been criticized because trees take years of growth to suck up an equivalent amount of released carbon. Also, not all renewable energy projects constitute an "offset." If an undertaking that would have happened anyway was jump-started through government subsidies—using wind power, for example—it cannot be considered a true offset, environmental groups say.

BIOGAS REACTOR, which consists of a well for fermenting dung, is being built in Tamil Nadu, India, as part of the World Cup's Green Goal program.

COURTESY OF GUNJAN SINHA

Currently, voluntary activities do not generate tradable emission credits. But imagine if they did. Suppose, for example, that anyone could earn credits for cutting consumption or increasing efficiency. Those villagers in India could earn credits for reducing their emissions that they could in turn sell, says Annie Petsonk, international counsel at Environmental Defense, a New York City–based nonprofit group. "How interesting would it be to have everyone participating? It would stimulate so much energy efficiency," she predicts. "We're talking about tapping economic power in favor of protecting the environment on a huge scale."

Gunjan Sinha is based in Berlin.

A WORK IN PROGRESS

The European Union instituted caps on greenhouse gas emissions for about 50 percent of Europe's industries to meet Kyoto Protocol goals. That system suffered a setback in May after governments realized that they had set emissions limits too high. Most companies came in far below their limits, rendering their pollution credits worthless and eliminating financial incentive for them to cut carbon dioxide output. The European Union is discussing how to tighten the scheme.

The U.S. is not part of the Kyoto treaty, but states are picking up the slack. Under the Regional Greenhouse Gas Initiative, Northeast and mid-Atlantic states plan to implement a market-based cap-and-trade program for all greenhouse gases emitted from power plants in the area. In California, a pending bill would cap greenhouse gas emissions from all industries in the state.

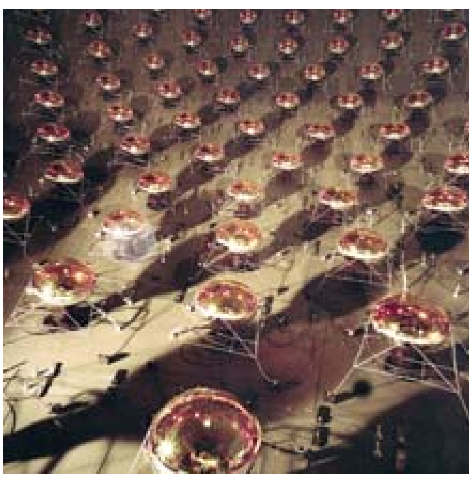
PHYSICS

The Neutrino Frontier

AT FERMILAB, PARTICLE SMASHING YIELDS TO FLAVOR CHANGING BY MARK ALPERT

Since 1983 researchers at the Fermi National Accelerator Laboratory in Batavia, Ill., have plumbed the subatomic realm by smashing high-energy protons and antiprotons together in the Tevatron, the world's most powerful particle collider. Next year, however, the high-energy frontier will move to Europe, where the even more powerful Large Hadron Collider will begin operations near Geneva. Fermilab intends to shut down the Tevatron by 2010. But rather than scrapping the device, lab officials have outlined an ambitious plan to use some of the collider's parts to enhance a promising research program: the study of the mysterious neutrino, whose strange properties may offer clues to new laws of physics.

Appropriately enough, the lab's namesake—physicist Enrico Fermi—coined the name for the particle, which means "little neutral one." Neutrinos come in three types, called flavors; the most common are electron neutrinos, which are produced in copious amounts by reactions in the sun. (The other flavors are muon and tau neutrinos.) Because neutrinos have no charge and interact with other particles only through the weak nuclear force and gravity, they pass through matter virtually unhindered and are devilishly difficult to detect. Until recently, most scientists thought neutrinos had no mass either, but in the late 1990s researchers found that the particles change flavor as they travel, and this transformation can happen only if they have mass.



MULTIPLIER TUBES inside the MiniBooNE tank can detect the light produced from neutrino interactions.

FERMILAB

Quantum theory predicts that these changes should be oscillatory, so physicists are now trying to measure how frequently neutrinos change flavor and the likelihood of each transition. Whereas previous experiments passively observed neutrinos created in the sun or in the earth's atmosphere, Fermilab investigators decided to generate intense beams of neutrinos using the same accelerators that supply protons to the Tevatron. (The high-energy protons crash into beryllium targets, producing pions that spit out neutrinos as they decay.) In 2002 researchers began firing a beam of muon neutrinos at a giant spherical tank half a kilometer away. Called the MiniBooNE detector, the 12-meter-wide tank contains some 800 tons of mineral oil and 1,500 photomultipliers that recognize the rare flashes of light caused by neutrino interactions with the oil. The MiniBooNE team is currently analyzing three years of data to determine how many of the particles changed to electron neutrinos while in flight.

In 2005 investigators started the MINOS

NEUTRINOS AND DARK ENERGY

One of the most intriguing hypotheses about neutrinos is that they may somehow be responsible for the accelerating expansion of the universe. Astrophysicists speculate that an unknown substance called dark energy has sped up the expansion over the past few billion years. Interestingly, the energy density needed to drive the current acceleration is roughly comparable to the estimated energy density of neutrinos. Rob Fardon, Ann E. Nelson and Neil Weiner of the University of Washington have proposed the existence of mass-varying neutrinos that would behave like a negative-pressure fluid, accelerating cosmic expansion as they move apart. Because the mass of these putative particles depends on their surroundings, researchers may be able to observe this phenomenon by looking for changes in flavor oscillations as neutrinos travel through different media.

experiment, which shoots a more powerful beam of muon neutrinos across 735 kilometers to a massive detector in the abandoned Soudan Iron Mine in northern Minnesota. (The long journey gives the particles more time to oscillate.) This past March, MINOS scientists announced that only about half of the expected muon neutrinos arrived at the Soudan detector, suggesting that the remainder changed flavor along the way. The results were consistent with those of the earlier K2K experiment, which studied muon neutrinos fired from an accelerator at Japan's KEK laboratory.

Fermilab researchers are now designing an experiment called MINERvA, which would improve the accuracy of the MINOS measurements by studying neutrino interactions with atomic nuclei. Scientists have also begun planning a study called NOvA, which would place another huge detector in northern Minnesota to hunt for electron neutrinos generated by the oscillations in the MINOS beam.

Because most neutrino detectors are designed to identify just one flavor, no single experiment can measure all the parameters of the oscillations. Fortunately, similar investigations in Japan may complement the U.S. studies. To increase their chances of success, Fermilab engineers are tweaking their accelerators to maximize the power of the MINOS beam, which in turn raises the number of neutrino interactions in the detector. Once the Tevatron is shut down, the lab intends to boost beam power even further by reconfiguring facilities now used to produce and store antiprotons.

Scientists are fascinated by neutrino oscillations because they may reveal phenomena that cannot be explained by the Standard Model, the highly successful but incomplete theory of particle physics. For example, the MiniBooNE results may confirm the existence of a fourth kind of neutrino—the so-called sterile neutrino—that is not subject to the weak force but may participate in novel interactions that have not yet been identified. "The neutrino is one of the least understood particles," says Richard Van de Water of the MiniBooNE team. "If there's extra physics, it's a good place for nature to hide it."

BIOLOGY

Sequencing Sea World

A GENETIC CENSUS OF THE OCEAN'S PRIMARY PREDATORS BY CHRISTINA REED

When biologists from the University of Connecticut wanted to take their laboratory's $85,000 DNA sequencer out to the Sargasso Sea in the Atlantic, the manufacturer, Applied Biosystems, balked. The warranty did not cover oceanographic expeditions. Even J. Craig Venter, famous for his role in decoding the human genome, had frozen his microbial samples from the Sargasso Sea for sequencing back on shore. His results showed that the surface water in the balmy sea around Bermuda teems with genetic material. The biologists from Connecticut wanted to go deeper into the ocean—and go beyond microbes—to test for diversity among the animals at the base of the food web. Working with a sequencer at sea would give them the best results.

The researchers are part of the Census of Marine Life, an international network of marine scientists that began a mission in 2000 to identify every living creature in the ocean by 2010. To this end, Peter Wiebe of the Woods Hole Oceanographic Institution designed a filter system of fine-mesh nets on a deepwater trawl. Onboard the National Oceanic and Atmospheric Administration's *Ronald H. Brown,* Wiebe, the chief scientist, and his team used the net device to scour the Sargasso Sea. Lashed down with a bungee cord in an air-conditioned room sat a brand-new, 140-kilogram DNA sequencer. Theirs was the first expedition to identify marine animals from 5,000 meters deep and sequence their gene markers, or "bar codes," while on the ship.

Ann Bucklin, director of the Marine Sciences and Technology Center at the University of Connecticut, presented the preliminary results in Amsterdam on May 15. She says this research will provide a baseline for measuring how pollution, overfishing, climate change and other human activity affect zooplankton and their environment.

During the 20-day expedition, 28 experts from 14 nations collected 1,000 individual specimens, including 120 species of fish, some new to science, and hundreds of species of zooplankton. Zooplankton are the drifting and swimming animals first in line to eat algae and other plant life, called phytoplankton. They also eat one another, and some can even take down small fish. Many organisms that spend their adult life stuck on the bottom of the seafloor start out as zooplankton in their larval stage. Whatever happens to zooplankton in the ocean has an immediate impact on the rest of the marine food web.

JELLYFISH RELATIVE siphonophore is actually a colony of animals. Some individuals attract and sting prey; others propel the colony. The genes of this and all marine creatures may be sequenced by 2010.

LARRY MADIN *Woods Hole Oceanographic Institution*

To examine the collected critters, marine scientists immediately rinsed the gooey and translucent mess of life-forms from the filters into buckets of cold water to keep as many animals as possible alive for visual identification. Change in pressure is not a problem for most of these grazers, which can descend hundreds of meters deep during the day. At night they rise back up near the surface to feed, treading carefully near the thermocline, the distinct boundary between the deeper, colder water and the warmer surface water. A change in temperature from near freezing to bathtub-water warm can kill the temperature-sensitive creatures.

NEED TO KNOW: GAS ATTACK

The increase in carbon dioxide in the atmosphere can alter ocean life. But just how much of an impact it will have is unknown. In the "deep open-ocean ecosystem, we have no idea how much is taking place," says Peter Wiebe of the Woods Hole Oceanographic Institution. A significant number of zooplankton and other sea life rely on very thin calcium carbonate shells, and changes in the pH caused by carbon dioxide, which is slightly acidic, can have severe consequences. "We need a good catalogue to see if something has changed," Wiebe adds. "We are charting the plankton in the sea like astronomers chart the stars in the sky."

When the animals die, they turn opaque and lose color as their proteins and DNA break down. Most zooplankton species are known from samples collected less than 1,000 meters deep. During this expedition, the scientists started at the surface and continued collecting throughout the water column at every 1,000 meters of depth.

The expedition was unique in having taxonomists working over the microscopes side by side with the molecular biologists preparing the species for sequencing. "These historically nonoverlapping skill sets mean that organisms collected for identification often aren't preserved in a way that permits DNA extraction," says Rob Jennings, a postdoctoral fellow at the University of Connecticut. This expedition changed that. "Training scientists to be adept at both taxonomic identification and DNA-based analysis is one of the top goals" of the project, he adds.

By the time they returned to port on April 30, the scientists had catalogued 500 animals and genetically bar-coded 220 of them. By 2010 the Census of Marine Life scientists expect to have bar-coded all 6,800 known species of zooplankton and potentially that many as yet undiscovered species as well.

Christina Reed, a freelance science journalist based in Seattle, writes frequently about ocean science.

ASTRONOMY

Venus de Seismo

NEW ORBITER BEGINS TO LISTEN FOR VENUSQUAKES **BY GEORGE MUSSER**

A**round 500 million** years ago, something awful seems to have happened on Venus. Maybe in spurts or maybe all at once, a fury of volcanism paved over nearly the entire surface. Some scientists think Earth's planetary sister could have supported life for billions of years, yet scarcely a trace now remains of that lost world. To fathom why a planet would have done such a thing to itself, researchers need to know its inner torment.

"Unraveling the mystery of why terrestrial planets evolve the way they do really requires that we understand the interior structure of Venus," says planetary scientist Ellen Stofan of Proxemy Research, an institute based in Laytonsville, Md. "This question links strongly back to the whole issue of why Earth is habitable and Venus apparently not."

At first glance, the European Space Agency's Venus Express orbiter (VEx), which arrived in April and began taking data in June, appears utterly unsuited to the task. It is little more than a copy of the Mars Express orbiter, with instruments originally designed for the Red Planet and sent to study Venus's atmosphere, all but neglecting the solid body. Last year, though, three planetary scientists argued that VEx could fill in an essential piece of the geologic puzzle, one that scientists had thought would take an expensive network of landers to detect: venusquakes.

Raphaël Garcia, Philippe Lognonné and Xavier Bonnin of the CNRS in Paris proposed to look for the low-frequency sound waves set off by tremors. On Earth, such waves make themselves felt in various ways, ranging from radio interference created as they ripple through the ionosphere to the infrasound that some scientists think animals perceive during quakes. It goes the other way, too: seismometers have registered deep bass notes from atmospheric turbulence and volcanic eruptions, as if the surface were acting as a giant microphone.

The dense Venusian atmosphere, usually a hindrance for observers, is a boon for aeroseismology. Not only does it pick up a larger fraction of the seismic energy than Earth's does—some 15 percent versus 0.04 percent or less—it extends farther from the planet, providing for more amplification as the sound propagates upward into thinner air and spreads its energy over fewer molecules. At a given altitude, the acoustic amplitude of a venusquake is 600 times that of a comparable earthquake. The wave ultimately deposits its energy in the rarefied upper atmosphere. Garcia says quakes might also generate the equivalent of tsunamis (which are buoyancy-driven rather than pressure-driven waves).

His team estimates that a magnitude 6 quake, occurring at a depth of 30 kilometers, should cause the air pressure and density to oscillate by as much as 10 percent. It would raise the temperature at an altitude of 170 kilometers by 10 degrees Celsius over an area 100 kilometers across. The heat pulse, lasting several minutes, is well within the means of the Venus Express spectrometer to detect—if it can be teased out from other atmospheric processes. "It is a bit too early to say if it will be possible or not," Garcia says.

Even the mere detection of a quake would be a breakthrough. "We are, of course, completely in the dark about how seismically active Venus might be," says Venus expert Sean Solomon of the Carnegie Institution of Washington. Quakes would reveal which geologic features are young, how fast the surface is deforming and how deeply faulted the planet is (strong quakes tend to be deeper)—helping to pin down the planet's recent history and the mantle convection that drives it.

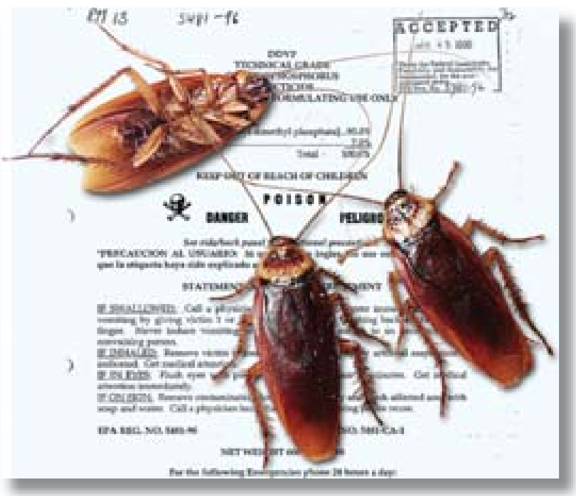
To be sure, probing internal structure with the resolution of Earth seismology would require multiple seismometers on the ground, each refrigerated to survive the 480 degree C temperature. The detection of quakes by VEx might shake loose support for a proposal by Stofan and her colleagues in the early 1990s to do just that. Only then might scientists come to learn why Venus is the goddess not of love but of tragedy.

POLICY

Slow-Acting

AFTER 25 YEARS, THE EPA STILL WON'T BAN A RISKY PESTICIDE BY PAUL RAEBURN

Dichlorvos, or DDVP, is a household pesticide related to World War II–era nerve agents. In May the Environmental Protection Agency proposed its continued sale, despite considerable evidence suggesting it is carcinogenic and harmful to the brain and nervous system, especially in children. On several occasions, the agency has come close to banning the pesticide—used in no-pest strips as well as in agriculture—but has always backed away. Environmentalists and labor unions charge that the latest decision was the product of backroom deals with industry and political interference—just as they did 25 years ago, when the EPA first considered a ban on DDVP and other similar pesticides.



IN HARM'S WAY: Attempts to ban the household use of the insecticide dichlorvos, or DDVP, have flopped, despite safety concerns since the 1970s.

"It's been two decades, and the riskiest uses—the home uses—are still on the market," says Aaron Colangelo, an attorney with the Natural Resources Defense Council (NRDC), which has sued the EPA over the pesticide. "It's bizarre, it's unlawful and it's a complete failure of the agency to protect public health."

DDVP is an organophosphate, one of a group of pesticides developed after World War II, when researchers discovered that insects' nervous systems were more sensitive to nerve agents than the human nervous system. The idea was that small amounts of these agents would be lethal to insects and harmless to people.

Questions about their safety arose in the 1970s. The EPA considered a ban on 13 pesticides in 1981, including DDVP, but took no action. The NRDC and labor unions sued the EPA. In the settlement, the agency said it would make a decision on the chemicals' safety by 1986.

The EPA missed that deadline. In 1988 it said DDVP deserved "special review," a designation that often leads quickly to a ban. The EPA again failed to act.

In 1996 a frustrated Congress passed the Food Quality Protection Act, giving the EPA until this year to make a firm decision on the safety of the estimated 800 pesticides in use in the U.S. The EPA has acted on many of them, but it is unlikely to meet its deadline for completing the pesticide review.

In its decision in May on DDVP—now the poster child for the agency's failure—the EPA added some restrictions to household use, saying the pest strips could be placed only in garages, cupboards, attics and crawl spaces. But that is unlikely to limit human exposure, insists David E. Camann of the Southwest Research Institute in San Antonio, Tex. "It's highly volatile, so it doesn't stay where it's put," he says. "It's going to migrate quickly through the air, and people are going to breathe it."

Colangelo says EPA officials told him these restrictions had been proposed by DDVP's maker, Amvac Chemical Corporation in Newport Beach, Calif., and that the EPA had quietly negotiated the details with the company—something it is not legally allowed to do. The EPA denies that, saying it pursues "a very open and transparent process." Amvac defends the product's safety and denies it colluded with the agency. "There were no 'backroom deals' with the EPA," Amvac said in a statement.

The agency's own employees charge, however, that industry is far too involved in these decisions. In a letter written May 24 by their union leaders, they complained that their colleagues in the pesticide program "feel besieged by political pressure exerted by agency officials perceived to be too closely aligned with the pesticide industry."

In June the NRDC once again took action, filing a petition charging that the latest decision was negotiated illegally and demanding an immediate ban on DDVP. The dispute is most likely headed to court.

Paul Raeburn writes about science, policy and the environment from New York City.

WAYS TO BUG OFF

Alternatives to organophosphate pesticides exist, notably pyrethrins, which are used in insect sprays, flea treatments and shampoos for head lice. These compounds, originally derived from chrysanthemum flowers, alter nerve cell function but in a different way than organophosphates do. They are not entirely safe, either. One better alternative is boric acid, especially when used in bait traps, which draw bugs and keep the chemical contained. Fortunately, the best ways to control insects in the home do not require any pesticides, just good housekeeping: washing dishes and emptying the garbage daily, sealing food containers, repairing screens, caulking cracks and gaps, and keeping bathrooms and kitchens as dry as possible.

EMILY HARRISON; SOURCES: ANTHONY BANNISTER *Gallo Images/Corbis*; TOM YOUNG *Corbis*

ANTHROPOLOGY

Paleolithic Juvenilia

WERE CAVE ARTISTS SEX- AND HUNTING-OBSESSED TEENAGE BOYS? BY JR MINKEL

Few images fire the imagination like Paleolithic cave paintings, part of the scant physical record left by humans who lived more than 10,000 years ago. To some scholars, this ancient art represents the handiwork of shamans; others detect traces of initiation rites or trancelike states. A new interpretation offers a more prosaic explanation for cave art: the expression of adolescent boys' preoccupation with hunting and sex.

During the late Paleolithic era, 10,000 to 50,000 years ago, humans roamed a vast steppe covering modern Europe, Asia and North America. These wandering hunters left behind myriad paintings on cave walls and artifacts depicting human figures and the large mammals of the day, including mammoth, elk, bison and horses. Early interpretations cast the images as religious icons or magical totems, perhaps part of hunting or fertility rituals performed by shamans. Humans definitely produced repetitive, stylized iconography over the past 10,000 years, says R. Dale Guthrie, a paleobiologist emeritus at the University of Alaska–Fairbanks. "Paleolithic art isn't like that," he contends. "It was done in a more naturalistic way, [showing] real animals eating, copulating, braying or bellowing, biting." To Guthrie, a hunter and amateur artist himself, cave painters seem more like natural historians than shamans.

DISEMBOWELED BISON gores a seemingly aroused man. This scene, from the Lascaux Caves in France, may reflect the typical focus of an adolescent male.

CAVES OF LASCAUX, DORDOGNE, FRANCE/THE BRIDGEMAN ART LIBRARY

Curious to discern more about the artists, Guthrie analyzed the handprints common to many image-bearing caves. He identified 201 handprints from Spanish and French caves that could be reasonably measured for width and length of fingers, palm and hand. He compared these data with measurements taken from 700 children, teenagers and adults at local Fairbanks schools. The groups are comparable, he reasoned, because both came from European stock and were well fed on high-protein diets.

Statistically, the cave handprints match up with modern children aged 10 to 16, Guthrie reports in his book, *The Nature of Paleolithic Art,* published earlier this year. "We've always known from the footprints in these caves that children are represented, but they're never given much to do by paleoanthropologists; they're regarded as invisible," says anthropologist Clive Gamble of Royal Holloway, University of London. Guthrie estimates the sex ratio of the handprints as largely male, by three or four to one. He argues that the subject matter of much Paleolithic art is consistent with its being created by adolescent boys, who would have been preoccupied with hunting and mating. Images of animals sometimes display lines through the beast's midsection, along with streaks of red pigment issuing from the mouth or loops below the belly—clear naturalistic hunting imagery, in his view. Voluptuous female imagery is even easier to understand, if modern adolescents are any guide.

Much cave art is spontaneous and playful, not shamanistic, in Guthrie's view. As evidence of its down-to-earth origins, he cites images that incorporate features of the cave wall or show a rudimentary skill level. Religious impulses and other motivations could still have played a role in some images, however. "I suspect there's not a one-size-fits-all answer" for Paleolithic art, says anthropologist Nicholas J. Conard of the University of Tübingen in Germany. "We're talking about 30,000 years of some fairly complex imagery."

JR Minkel's earliest surviving drawings depict killer shark-bots.

ART OF SURVIVAL

Teenage boys may have dominated cave paintings, according to R. Dale Guthrie of the University of Alaska–Fairbanks. But that is not to say that they had a monopoly on artistic expression. The constant environment of a cave tends to preserve its contents, Guthrie notes, so those who ventured inside are disproportionately represented. Jewelry, pottery, clothing and soft artifacts would have degraded more readily.

BY THE NUMBERS

The Crystal Crisis

METH ABUSE MOVES EASTWARD **BY RODGER DOYLE**

Methamphetamine has a fearsome reputation as a destroyer of families. Meth binges, which produce a sense of euphoria, may last a week; the severe paranoia and depression that come afterward persist for days or weeks, making it difficult to hold a job or take care of children. By comparison, cocaine binges rarely last more than three days.

The meth problem grew out of addiction to over-the-counter remedies available since the 1930s for the treatment of asthma, narcolepsy and other ailments. These products, which were based on amphetamine and its sister drug, the more potent methamphetamine, could also be used for staying awake and suppressing appetite, a boon to truck drivers, students and people wanting to keep the pounds off.

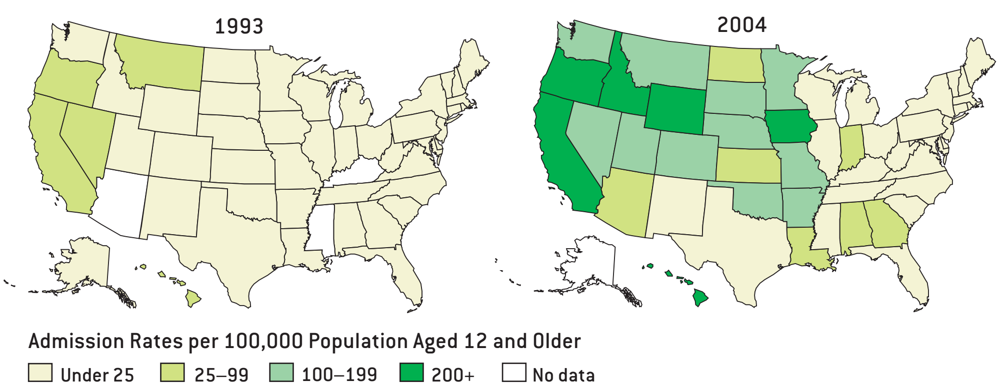
Methamphetamine was first produced as an illicit drug in 1962, and San Francisco motorcycle gangs were soon distributing it up and down the Pacific Coast. In their wake, many others followed: all that was required were commonly available ingredients such as rubbing alcohol, lye, and a supply of ephedrine or pseudoephedrine; the last are medications available without prescription. As a result, mom-and-pop laboratories have traditionally produced much of the illicit methamphetamine; Mexican gangs bring into the country most of the remainder. "Ice," or chunks of crystalline methamphetamine that can be smoked in a pipe, became available in Hawaii during the 1980s and soon spread to the mainland, where it has become the most prized form of the drug.

The maps, which show treatment facility admissions rates for methamphetamine abuse in 1993 versus 2004, illustrate that the drug is spreading east, expanding into the Northeast, the upper South, Texas, and East Central states. Epidemiologist Jane Maxwell of the University of Texas at Austin follows the epidemic closely. She believes that larger quantities of the drug are becoming available in the East as organized gangs become more involved in marketing and as the mom-and-pop labs find pseudoephedrine more difficult to obtain. Furthermore, she notes that these gangs are now selling a purer and hence more potent form of the drug.

Meth users are probably getting sicker. This conclusion comes from an analysis by the Substance Abuse and Mental Health Services Association (SAMHSA), which found that the number of meth users during the past month—a good indicator of hardcore usage—has remained at 600,000 from 2002 through 2004. SAMHSA found, however, that the proportion of these users abusing or becoming dependent on the drug rose from 11 percent in 2002 to 22 percent in 2004. That translates into 130,000 people who are at particular risk of violence and family breakdown.

Rodger Doyle can be reached at rodgerpdoyle@verizon.net

Admissions to Treatment for Meth Abuse



RODGER DOYLE; SOURCE: "TRENDS IN METHAMPHETAMINE/AMPHETAMINE ADMISSIONS TO TREATMENT: 1993–2000" IN *THE DASIS REPORT*, ISSUE 9, 2006, SUBSTANCE ABUSE AND MENTAL HEALTH SERVICES ADMINISTRATION (SAMHSA); DATA INCLUDE AMPHETAMINE

DOING THE METH

Methamphetamine—street names are meth, speed, ice, crystal, crank and glass—comes in three forms: as a white, odorless, crystalline powder; clear chunky crystals; and small, brightly colored tablets. It can be injected, snorted and eaten, but smoking is the most popular method. Smoking and injection provide the most stimulation.

In the U.S., 78 percent of users are white, 14 percent Hispanic and 4 percent black; 78 percent are male, and 82 percent are 26 years of age and over. In 2004, 1.4 million people used methamphetamine, and 600,000 used it in the past month.

FURTHER READING

History of the Methamphetamine Problem. M. Douglas Anglin, Cynthia Burke, Brian Perrochet, Ewa Stamper and Samia Dawud-Noursi in *Journal of Psychoactive Drugs*, Vol. 32, No. 2, pages 137–141; April–June 2000.

Methamphetamine Use, Abuse, and Dependence: 2002, 2003 and 2004. *The NSDUH Report* (National Survey on Drug Use and Health); September 16, 2005.

DATA POINTS: ANTIFREEZE

Core sediments retrieved by three icebreaker and drill ships revise what is known about the Arctic since about 55 million years ago. For a few million years, the North Pole felt downright Floridian thanks to the presence of greenhouse gases released by some unknown geologic process. The warmth as recorded by the core data is 10 degrees Celsius higher than climate models had predicted. They suggest that heat-trapping gases may exert a more potent climatic influence than previously thought.

North Pole's mean annual temperature: **−20 degrees C**

Temperature 55 million years ago: **23 degrees C**

Concentration of atmospheric carbon dioxide today, in parts per million: **380**

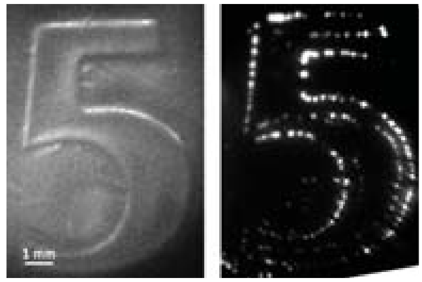
Concentration 55 million years ago: **2,000**

Rise in average global temperature as a result: **5 degrees C**

When "icehouse" conditions began: **45 million years ago**

NANOTECH

Alight Touch



IMPRINT of the number "5" glows (*right*) thanks to an electronic sensor with human-touch sensitivity.

Modern state-of-the-art fingertip-size sensors for touch are sensitive enough to detect features only about two millimeters wide. Chemical engineers at the University of Nebraska–Lincoln have now made sensors that reach the micron range, rivaling the delicacy of human touch. They fabricated a 100-nanometer-thick sheet of alternating self-assembled layers of gold and cadmium sulfide nanoparticles separated by insulating films. Once a voltage is applied through the film, pressure on it causes electrons to tunnel from the gold through the insulating films to the cadmium sulfide layers, which then glow. Images developed from this light capture features 40 microns wide and five high, detailed enough to show the wrinkles on Lincoln's clothing on a penny. Future touch sensors for robots based on this work will most likely rely not on light signals but rather on electrical impulses, researcher Ravi Saraf explains. The findings appear in the June 9 *Science*. —*Charles Q. Choi*

ECOLOGY

A Firm Stand

American chestnut trees once made up a quarter of eastern U.S. forests, with four billion of them stretching from Maine to Florida. Virtually all these chestnuts were killed within 50 years by a fungus introduced to the U.S. in 1904. But wildlife biologist Nathan Klaus of the Georgia Department of Natural Resources recently discovered half a dozen American chestnuts near President Franklin D. Roosevelt's Little White House in the southern end of the Appalachians, with the largest some 40 feet tall. "You occasionally find a single American chestnut in the wild, but such a large group of them is unprecedented," Klaus says. He notes that the chestnut-killing fungus may not have thrived on the dry, rocky mountaintop where this stand dwells. The American Chestnut Foundation announced May 19 it would breed the trees with blight-resistant Chinese chestnuts in the hopes of eventually developing a hardier, mostly American hybrid.—*Charles Q. Choi*

BEHAVIOR

Solitary Briny Confinement

In these days of pandemic anxiety, envy the Caribbean spiny lobster's ability to avoid sick cohorts. Marine biologists had noticed that in this typically social crustacean, sickly looking lobsters, infected with a lethal and contagious virus called PaV1, usually become isolated from the pack. To determine if the avoidance is purposeful, scientists set up two adjacent dens in a tank of seawater, tied either a sick or healthy lobster in one den and then introduced a second lobster. If the second creature was already infected, it bunked with its sick or healthy tank-mate equally often. But healthy lobsters were one fourth as likely to share dens with infected lobsters than with healthy ones—even before the sickies became contagious—probably because of chemical signals. The result marks the first observation of wild social animals avoiding diseased individuals, the researchers write in the May 25 *Nature*. —*JR Minkel*

SPINY LOBSTERS steer clear of sick brethren to avoid contagion.

VIVEK MAHESHWARI AND RAVI F. SARAF *University of Nebraska–Lincoln* (*top*); B. JONES AND M. SHIMLOCK NHPA (*bottom*); ILLUSTRATION BY MATT COLLINS

BIOLOGY

A Place for RNA

The only cellular structures known to possess genomes are the nucleus, mitochondria and chloroplasts. Now researchers find that centrosomes, which help to oversee cell division, apparently possess their own genetic machinery—and curiously, it's not DNA but its cousin, RNA. Scientists collaborating at the Marine Biological Laboratory in Woods Hole, Mass., purified five RNA sequences from surf clam eggs. Though abundant in the centrosomes, few to no copies of these RNAs were found elsewhere in the cell, and their sequences were not seen in any genome database. One of the RNAs appears to encode machinery involved in replicating DNA and RNA, which suggests that centrosomes can duplicate their genetic material. Relatively little is known about the inner workings of centrosomes even after a century of study, and the investigators suggest their discovery could explain centrosome evolution and function. The work is online June 5 via the *Proceedings of the National Academy of Sciences USA*. *—Charles Q. Choi*

ACOUSTICS

Lasers from Sound

Lasers are most familiar as pinpoint beams of coherent light, but the principle of lasing—amplification by stimulated emission—occurs just fine without light, as two reports of ultrasound lasers demonstrate. One device consists of layers of gallium arsenide and aluminum arsenide that emit and partially trap sound vibrations in the solid (phonons) oscillating in the terahertz range. A voltage creates a burst of phonons, which reverberate and multiply, and the amplified ultrasound exits one end. Physicists describe this so-called saser, for sound laser, in the June 2 *Physical Review Letters*. In the "uaser" (pronounced "wayzer"), piezoelectric oscillators vibrate an aluminum block, which feeds back into the oscillators and locks the vibrations into a single megahertz frequency. The system generates multidirectional ultrasound and might aid in studying so-called random lasers that likewise produce scattered, coherent light, co-designer Richard Weaver of the University of Illinois informed the Acoustical Society of America on June 8. *—JR Minkel*

FLUID DYNAMICS

Platonic Liquids

SPIN CYCLE: Water adopts trefoil, square and pentagonal shapes depending on the rotational speed.

KÅRE HARTVIG JENSEN

Isaac Newton spun a bucket of water like a top and observed a round vortex. Now Danish researchers have squared that circle by rotating just the bottom of a water-filled bucket. As they dialed up the speed to a few spins per second, the vortex adopted a trefoil shape, then became square, pentagonal and hexagonal. The sluggish outer layer of water may be amplifying small variations in the radius of the vortex, creating corners, says bucket spinner Tomas Bohr of the Technical University of Denmark (and grandson of Niels), whose group reports in the May 5 *Physical Review Letters*. Storms are similar in principle to a semistationary bucket, which may explain structured hurricane eyes, and the underlying principle might apply to bathtub drains, too, Bohr says. "I've never seen a triangular bathtub vortex," he adds, "but who knows?" *—JR Minkel*

BRIEF POINTS

- **Physicians typically prescribe a seven- to 10-day antibiotic regimen for pneumonia. A double-blind study, however, shows that a three-day treatment is just as effective.**

  BMJ, June 10

- **A 300-mile-wide crater under the Antarctic ice may be the evidence that an object 30 miles wide caused the earth's greatest mass extinction 250 million years ago. The one that polished off the dinosaurs was about one fifth that size.**

  American Geophysical Union Joint Assembly, Baltimore, May 23–26

- **And good for covering up, too: Fig trees may have marked the beginning of agriculture 11,400 years ago. Based on finds in the lower Jordan Valley, figs were domesticated 1,000 years before wheat, barley and legumes.**

  Science, June 2

- **Scientists discovered a few years ago that the mammalian heart has stem cells. Now, in a step toward using stem cells in cardiac repair, they have found where they lurk, based on mouse studies: in small niches near heart muscle cells.**

  Proceedings of the National Academy of Sciences USA, June 13

Folk Science

Why our intuitions about how the world works are often wrong By MICHAEL SHERMER

Thirteen years after the legendary confrontation over the theory of evolution between Bishop Samuel Wilberforce ("Soapy Sam") and Thomas Henry Huxley ("Darwin's bulldog"), Wilberforce died in 1873 in an equestrian fall. Huxley quipped to physicist John Tyndall, "For once, reality and his brain came into contact and the result was fatal."

When it comes to such basic forces as gravity and such fundamental phenomena as falling, our intuitive sense of how the physical world works—our folk physics—is reasonably sound. Thus, we appreciate Huxley's wry comment and note that even children get the humor of cartoon physics, where, for example, a character running off a cliff does not fall until he realizes that he has left terra firma.

Folk science often gets it wrong because we evolved in a radically different environment.

But much of physics is counterintuitive, as is the case in many other disciplines, and before the rise of modern science we had only our folk intuitions to guide us. Folk astronomy, for example, told us that the world is flat, celestial bodies revolve around the earth, and the planets are wandering gods who determine our future. Folk biology intuited an élan vital flowing through all living things, which in their functional design were believed to have been created ex nihilo by an intelligent designer. Folk psychology compelled us to search for the homunculus in the brain—a ghost in the machine—a mind somehow disconnected from the brain. Folk economics caused us to disdain excessive wealth, label usury a sin and mistrust the invisible hand of the market.

The reason folk science so often gets it wrong is that we evolved in an environment radically different from the one in which we now live. Our senses are geared for perceiving objects of middling size—between, say, ants and mountains—not bacteria, molecules and atoms on one end of the scale and stars and galaxies on the other end. We live a scant three score and 10 years, far too short a time to witness evolution, continental drift or long-term environmental changes.

Causal inference in folk science is equally untrustworthy. We correctly surmise designed objects, such as stone tools, to be the product of an intelligent designer and thus naturally assume that all functional objects, such as eyes, must have also been intelligently designed. Lacking a cogent theory of how neural activity gives rise to consciousness, we imagine mental spirits floating within our heads. We lived in small bands of roaming hunter-gatherers that accumulated little wealth and had no experience of free markets and economic growth.

Folk science leads us to trust anecdotes as data, such as illnesses being cured by assorted nostrums based solely on single-case examples. Equally powerful are anecdotes involving preternatural beings, compelling us to make causal inferences linking these nonmaterial entities to all manner of material events, illness being the most personal. Because people often recover from sickness naturally, whatever was done just before recovery receives the credit, prayer being the most common.

In this latter case, we have a recent scientific analysis of this ancient folk science supposition. The April issue of the *American Heart Journal* published a comprehensive study directed by Harvard Medical School cardiologist Herbert Benson on the effects of intercessory prayer on the health and recovery of patients undergoing coronary bypass surgery. The 1,802 patients were divided into three groups, two of which were prayed for by members of three religious congregations. Prayers began the night before the surgery and continued daily for two weeks after. Half the prayer recipients were told that they were being prayed for, whereas the other half were told that they might or might not receive prayers. Results showed that prayer itself had no statistically significant effect on recovery. Case closed.

Of course, people will continue praying for their ailing loved ones, and by chance some of them will recover, and our folk science brains will find meaning in these random patterns. But for us to discriminate true causal inferences from false, real science trumps folk science. SA

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

BRAD HINES

Virtuous Circles and Fragile States

Want to promote stable democracy in struggling nations? Send timely packages of food, seeds and medicine By JEFFREY D. SACHS

If U.S. leaders better understood the politics of impoverished and crisis-ridden countries, they would more effectively protect American national security by advancing the causes of economic development and democracy. Although the administration of George W. Bush has often stated its commitment to the spread of democracy, partly to combat the risks of terror, it relies excessively on military approaches and threats rather than strategic aid. Timely development assistance to places hovering between democracy and disarray can yield enormous benefits.

For nations in a deep crisis, the greatest danger is a self-fulfilling prophecy of disaster. Consider Liberia, just emerging from a prolonged civil war, and Haiti, which has suffered decades of intense political instability. Both nations have recently elected new democratic governments, but both face continuing possibilities of internal violence and disorder.

Rather than giving practical help, rich countries send consultants.

When the public thinks that a newly elected national government will succeed, local leaders throw their support behind it. Expectations of the government's longevity rise. Individuals and companies become much more likely to pay their taxes, because they assume that the government will have the police power to enforce the tax laws.

A virtuous circle is created. Rising tax revenues strengthen not only the budget but also political authority and enable key investments—in police, teachers, roads, electricity—that promote public order and economic development. They also bolster confidence in the currency. Money flows into the commercial banks, easing the specter of banking crises.

When the public believes that a government will fail, the same process runs in reverse. Pessimism splinters political forces. Tax payments and budget revenues wane. The police and other public officials go unpaid. The currency weakens. Banks face a withdrawal of deposits and the risk of banking panics. Disaster feeds more pessimism.

By attending to the most urgent needs of these fragile states, U.S. foreign policy can tilt the scales to favor the consolidation of democracy and economic improvement. To an informed and empathetic observer, the necessary actions will usually be clear. Both Liberia and Haiti lack electricity service, even in their capital cities. Both countries face massive crises of hunger and insufficient food production. Both suffer from pervasive infectious diseases that are controllable but largely uncontrolled.

But if each impoverished farm family is given a bag of fertilizer and a tin of high-yield seeds, a good harvest with ample food output can be promoted within a single growing season. A nationwide campaign to spread immunizations, antimalaria bed nets and medicines, vitamin supplements and deworming agents can improve the health of the population even without longer-term fixes of the public health system. Electric power can be restored quickly in key regions. And safe water outlets, including boreholes and protected natural springs, can be constructed by the thousands within a year.

All these initiatives require financial aid, but the costs are small. Far too often, however, the U.S. response is neglect. Rather than giving practical help, the rich countries and international agencies send an endless stream of consultants to design projects that arrive too late, if ever. They ignore emergency appeals for food aid. After a few months, the hungry, divided, disease-burdened public begins to murmur that "nothing has changed," and the downward spiral recommences. Pessimism breeds pessimism. Eventually the government falls, and the nascent democracy is often extinguished.

By thinking through the underlying ecological challenges facing a country—drought, poor crops, disease, physical isolation—and raising the lot of the average household through quick-disbursing and well-targeted assistance, U.S. foreign policy makers would provide an invaluable investment in democracy, development and U.S. national security. Liberia and Haiti are two important places to begin to make good on the Bush administration's pledge to spread democracy. SA

Jeffrey D. Sachs is director of the Earth Institute at Columbia University and of the U.N.'s Millennium Project.

BRUCE GILBERT/EARTH INSTITUTE

The Geometer of Particle Physics

Alain Connes's noncommutative geometry offers an alternative to string theory. In fact, being directly testable, it may be better than string theory By ALEXANDER HELLEMANS

If there is a mathematician eagerly waiting for the Large Hadron Collider near Geneva to start up next year, it is Alain Connes of the Collège de France in Paris. Like many physicists, Connes hopes that the Higgs particle will show up in detectors. The Higgs is the still missing crowning piece of the so-called Standard Model—the theoretical framework that describes subatomic particles and their interactions. For Connes, the discovery of the Higgs, which supposedly endows the other particles with mass, is crucial: its existence, and even its mass, emerges from the arcane equations of a new form of mathematics called noncommutative geometry, of which he is the chief inventor.

ALAIN CONNES: REDEFINING SPACE

- **Chief inventor of noncommutative geometry, a mathematical space wherein the order of events is more important than the location of objects.**
- **On its origin: "What is superb is that what has been obtained by a dialogue between experiment and theory turns out to be of an unsuspected mathematical beauty."**

Connes's idea was to extend the relation between geometric space and its commutative algebra of Cartesian coordinates, such as latitude and longitude, to a geometry based on noncommutative algebras. In commutative algebra, the product is independent of the order of the factors: $3 \times 5 = 5 \times 3$. But some operations are noncommutative. Take, for example, a stunt plane that can aggressively roll (rotate over the longitudinal axis) and pitch (rotate over an axis parallel to the wings). Assume a pilot receives radio instructions to roll over 90 degrees and then to pitch over 90 degrees toward the underside of the plane. Everything will be fine if the pilot follows the commands in that order. But if the order is inverted, the plane will take a nosedive. Operations with Cartesian coordinates in space are commutative, but rotations over three dimensions are not.

To gain a clearer vision of what goes on in nature, physicists sometimes resort to "phase space." Such a space is an alternative to Cartesian coordinates—a researcher can plot the position of an electron against its momentum, rather than simply its x and y locations. Because of the Heisenberg uncertainty principle, one cannot measure both quantities simultaneously. As a consequence, position times momentum does not equal momentum times position. Hence, the quantum phase space is noncommutative. Moreover, introducing such noncommutativity into an ordinary space—say, by making the x and the y coordinates noncommutative—produces a space that has noncommutative geometry.

Through such analyses, Connes discovered the peculiar properties of his new geometry, properties that corresponded to the principles of quantum theory. He has spent three decades refining his thinking, and even though he laid down the basics in a 1994 book, re-

NICOLAS GUERIN *Azimuts Productions*

searchers beat a path to listen to him. On a day plagued by typical March showers and wind, about 60 of the crème de la crème of French mathematicians fill Salle 5 at the Collège de France. Like a caged lion, the 59-year-old Connes walks quickly back and forth between two overhead projectors, talking rapidly, continually replacing transparencies filled with equations. Outside, police sirens scream amid student protestors trying to occupy the Sorbonne next door in response to the French government's proposed new employment law.

Connes seems oblivious to the commotion—even afterward, while crossing the rue Saint-Jacques past blue police vans and officers in riot gear, he keeps talking about how his research has led him to new insights into physics. As an example, Connes refers to the way particle physics has grown: The concept of spacetime was derived from electrodynamics, but electrodynamics is only a small part of the Standard Model. New particles were added when required, and confirmation came when these predicted particles emerged in accelerators.

But the spacetime used in general relativity, also based on electrodynamics, was left unchanged. Connes proposed something quite different: "Instead of having new particles, we have a geometry that is more subtle, and the refinements of this geometry generate these new particles." In fact, he succeeded in creating a noncommutative space that contains all the abstract algebras (known as symmetry groups) that describe the properties of elementary particles in the Standard Model.

The picture that emerges from the Standard Model, then, is that of spacetime as a noncommutative space that can be viewed as consisting of two layers of a continuum, like the two sides of a piece of paper. The space between the two sides of the paper is an extra discrete (noncontinuous), noncommutative space. The discrete part creates the Higgs, whereas the continuum parts generate the gauge bosons, such as the *W* and *Z* particles, which mediate the weak force.

Connes has become convinced that physics calculations not only reflect reality but hide mathematical jewels behind their apparent complexity. All that is needed is a tool to peer into the complexity, the way the electron microscope reveals molecular structure, remarks Connes, whose "electron microscope" is noncommutative geometry. "What I'm really interested in are the complicated computations performed by physicists and tested by experiment," he declares. "These calculations are tested at up to nine decimals, so one is certain to have come across a jewel, something to elucidate."

One jewel held infinities. Although the Standard Model proved phenomenally successful, it quickly hit an obstacle: infinite values appeared in many computations. Physicists, including Gerard 't Hooft and Martinus Veltman of the University of Utrecht in the Netherlands, solved this problem by introducing a mathematical technique called renormalization. By tweaking certain values in the models, physicists could avoid these infinities and calculate properties of particles that corresponded to reality.

Although some researchers viewed this technique as a bit like cheating, for Connes renormalization became another opportunity to explore the space in which physics lives. But it wasn't easy. "I spent 20 years trying to understand renormalization. Not that I didn't understand what the physicists were doing, but I didn't understand what the meaning of the mathematics was that was behind it," Connes says. He and physicist Dirk Kreimer of the Institut des Hautes Études Scientifiques near Paris soon realized that the recipe to eliminate infinities is in fact linked to one of the 23 great problems in mathematics formulated by David Hilbert in 1900—one that had been solved. The linkage gave renormalization a mathematically rigorous underpinning—no longer was it "cheating."

The relation between renormalization and noncommutative geometry serves as a starting point to unite relativity and quantum mechanics and thereby fully describe gravity. "We now have to make a next step—we have to try to understand how space with fractional dimensions," which occurs in noncommutative geometry, "couples with gravitation," Connes asserts. He has already shown, with physicist Carlo Rovelli of the University of Marseille, that time can emerge naturally from the noncommutativity of the observable quantities of gravity. Time can be compared with a property such as temperature, which needs atoms to exist, Rovelli explains.

To Connes, physics calculations not only reflect reality but hide mathematical jewels.

What about string theory? Doesn't that unify gravitation and the quantum world? Connes contends that his approach, looking for the mathematics behind the physical phenomena, is fundamentally different from that of string theorists. Whereas string theory cannot be tested directly—it deals with energies that cannot be created in the laboratory—Connes points out that noncommutative geometry makes testable predictions, such as the Higgs mass (160 billion electron volts), and he argues that even renormalization can be verified.

The Large Hadron Collider will not only test Connes's math but will also give him data to extend his work to smaller scales. "Noncommutative geometry now supplies us with a model of spacetime that reaches down to 10^{-16} centimeter, which still is a long way to go to reach the Planck scale, which is 10^{-33} centimeter," Connes says. That is not quite halfway. But to Connes, the glass undoubtedly appears half full. SA

Alexander Hellemans is based in Bergerac, France.

Nanotechnology's Future

Over the next two decades, this new field for controlling the properties of matter will rise to prominence through four evolutionary stages By MIHAIL C. ROCO

Today nanotechnology is still in a formative phase—not unlike the condition of computer science in the 1960s or biotechnology in the 1980s. Yet it is maturing rapidly. Between 1997 and 2005, investment in nanotech research and development by governments around the world soared from $432 million to about $4.1 billion, and corresponding industry investment exceeded that of governments by 2005. By 2015, products incorporating nanotech will contribute approximately $1 trillion to the global economy. About two million workers will be employed in nanotech industries, and three times that many will have supporting jobs.

Rearranging atoms leads to new properties.

Descriptions of nanotech typically characterize it purely in terms of the minute size of the physical features with which it is concerned—assemblies between the size of an atom and about 100 molecular diameters. That depiction makes it sound as though nanotech is merely looking to use infinitely smaller parts than conventional engineering. But at this scale, rearranging the atoms and molecules leads to new properties. One sees a transition between the fixed behavior of individual atoms and molecules and the adjustable behavior of collectives. Thus, nanotechnology might better be viewed as the application of quantum theory and other nano-specific phenomena to fundamentally control the properties and behavior of matter.

Over the next couple of decades, nanotech will evolve through four overlapping stages of industrial prototyping and early commercialization. The first one, which began after 2000, involves the development of passive nanostructures: materials with steady structures and functions, often used as parts of a product. These can be as modest as the particles of zinc oxide in sunscreens, but they can also be reinforcing fibers in new composites or carbon nanotube wires in ultraminiaturized electronics.

The second stage, which began in 2005, focuses on active nanostructures that change their size, shape, conductivity or other properties during use. New drug-delivery particles could release therapeutic molecules in the body only after they reached their targeted diseased tissues. Electronic components such as transistors and amplifiers with adaptive functions could be reduced to single, complex molecules.

Starting around 2010, workers will cultivate expertise with systems of nanostructures, directing large numbers of intricate components to specified ends. One application could involve the guided self-assembly of nanoelectronic components into three-dimensional circuits and whole devices. Medicine could employ such systems to improve the tissue compatibility of implants, or to create scaffolds for tissue regeneration, or perhaps even to build artificial organs.

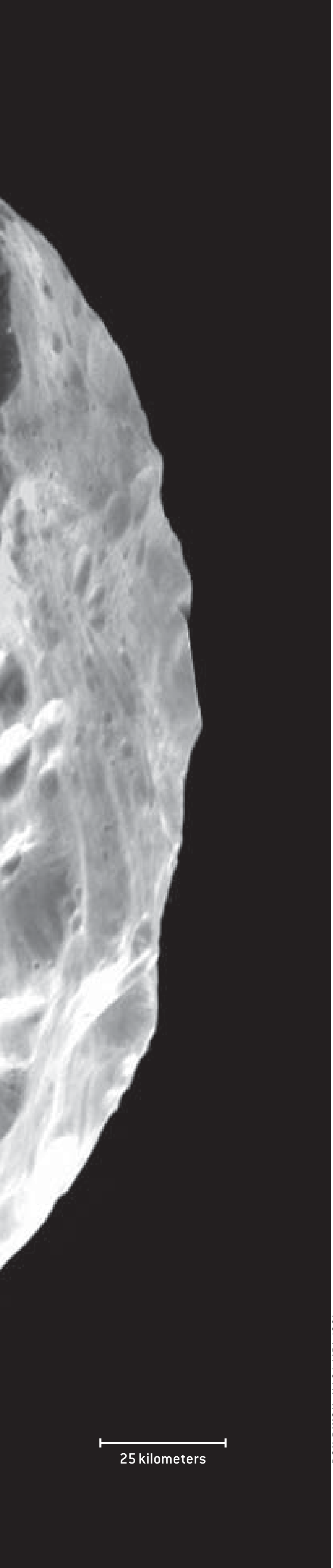
After 2015–2020, the field will expand to include molecular nanosystems—heterogeneous networks in which molecules and supramolecular structures serve as distinct devices. The proteins inside cells work together this way, but whereas biological systems are water-based and markedly temperature-sensitive, these molecular nanosystems will be able to operate in a far wider range of environments and should be much faster. Computers and robots could be reduced to extraordinarily small sizes. Medical applications might be as ambitious as new types of genetic therapies and antiaging treatments. New interfaces linking people directly to electronics could change telecommunications.

Over time, therefore, nanotechnology should benefit every industrial sector and health care field. It should also help the environment through more efficient use of resources and better methods of pollution control. Nanotech does, however, pose new challenges to risk governance as well. Internationally, more needs to be done to collect the scientific information needed to resolve the ambiguities and to install the proper regulatory oversight. Helping the public to perceive nanotech soberly in a big picture that retains human values and quality of life will also be essential for this powerful new discipline to live up to its astonishing potential. SA

Mihail C. Roco is senior adviser for nanotechnology to the National Science Foundation and a key architect of the National Nanotechnology Initiative.

COURTESY OF MIHAIL C. ROCO

GUNKED-UP, ICY BODY, Saturn's largest irregular moon, Phoebe, looks like a comet that was plucked from solar orbit. The craters are named after the Argonauts of Greek mythology: the largest, at top, is Jason; immediately to its left is Erginus; and the one in shadow along the bottom rim is Oileus.



DON DIXON; NASA/JPL/SSI

The STRANGEST Satellites in the Solar System

Found in stretched, slanted, loop-d-loop orbits, an odd breed of planetary satellites opens a window into the formation of the planets

By David Jewitt, Scott S. Sheppard and Jan Kleyna

FIVE YEARS AGO TWO OF US WHILED AWAY A CLOUDY NIGHT ON THE SUMMIT OF MAUNA KEA IN HAWAII by placing bets on how many moons remained to be discovered in the solar system. Jewitt wagered $100 that a dedicated telescopic search could find, at most, 10 new ones. After all, he reasoned, in the entire 20th century, astronomers had come across only a few. Sheppard more optimistically predicted twice as many, given the increased sensitivity of modern astronomical facilities.

Sheppard is now a richer man. Since that night, our team has discovered 62 moons around the giant planets, with more in the pipeline. Other groups have found an additional 24. (In strict astronomical parlance, they are "satellites," not "moons." There is only one moon and it is Earth's satellite. But even astronomers generally adopt the popular usage.) No one predicted that the family of the sun had so many members lurking in the shadows. They are classified as "irregular," meaning that their orbits are large, highly elliptical and tilted with respect to the equators of their host planets. So-called regular moons, such as Earth's or the large Galilean satellites of Jupiter, have comparatively tight, circular and nearly equatorial orbits.

Odder still, most of the irregulars have retrograde orbits, which means they each trundle around their host planet in a direction opposite to the sense of the rotation of the planet. In contrast, regular moons have prograde orbits. For example, as seen from a position above Earth's North Pole, our moon travels counterclockwise—the same direction in which Earth rotates on its axis and revolves around the sun. The other planets also move counterclockwise, a pattern that presumably reflects the swirling of the disk of gas and dust out of which they emerged four and a half billion years ago. Regular moons share this motion because, astronomers think, they coalesced from disks around their respective planets. So the contrary behavior of the irregular moons is a sign of a different origin.

These bodies are not well explained by standard models, and a wave of fresh theoretical work is under way. It seems that they are products of a long-gone epoch when the gravitational tug of the newly formed planets scattered—or snatched—small bodies from their original orbits. Studying them promises to illuminate the early stages in the development of the solar system.

Black Sheep

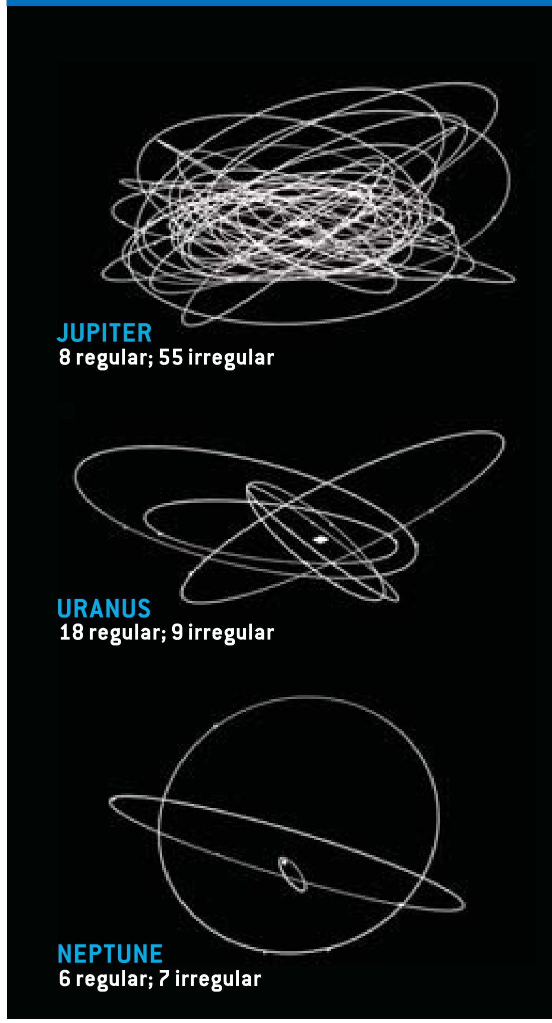
ALTHOUGH THE FIRST known irregular moon, Neptune's Triton, was discovered in 1846, most escaped detection until recently because they tend to be smaller and thus fainter than their regular counterparts. Adding to the challenge, they are distributed over a much larger region of space. For instance, Jupiter's outermost regular satellite, Callisto, orbits 1.9 million kilometers from the planet, whereas its known irregular moons range as far away as 30 million kilometers. That distance is comparable to the size of Jupiter's gravitational realm, or Hill sphere, beyond which the sun would pry loose any moon. If visible to the eye, Jupiter's Hill sphere would appear 10 degrees across—20 times the angular diameter of the full moon. It is huge compared with the fields of view of most telescopes.

Scanning such a vast area for moons demands the newest, largest digital detectors and the analysis of up to 100 gigabytes of data a night [*see box on page 46*]. Our own Hawaii Moon Survey focused initially on Jupiter, whose proximity allows us to probe small moons that would be too faint to detect around the other, more distant giant planets. Teams led by Brett Gladman of the University of British Columbia, Matthew Holman of the Harvard-Smithsonian Center for Astrophysics (CfA) and J. J. Kavelaars of the National Research Council of Canada's Herzberg Institute of Astrophysics have mounted parallel efforts to survey Saturn, Uranus and Neptune.

All four giant planets, irrespective of mass, turn out to have similar irregular moon systems. Extrapolating from the findings so far, we estimate that each has about 100 irregular moons larger than one kilometer in diameter. The bodies occupy a wide range of sizes, with smaller ones being much more abundant. In Jupiter's case, the size range extends from the largest irregular, J6 Himalia, at about 180 kilometers in diameter, down to the smallest objects at only one or two kilometers across.

The orbits of these moons are some of the most complicated in the solar system. Because they roam so far from their host planet, they are tugged by both planetary and solar gravity, and their orbits precess rapidly—that is, the long axis of the ellipse representing the orbit rotates.

A SWARM OF MOONS

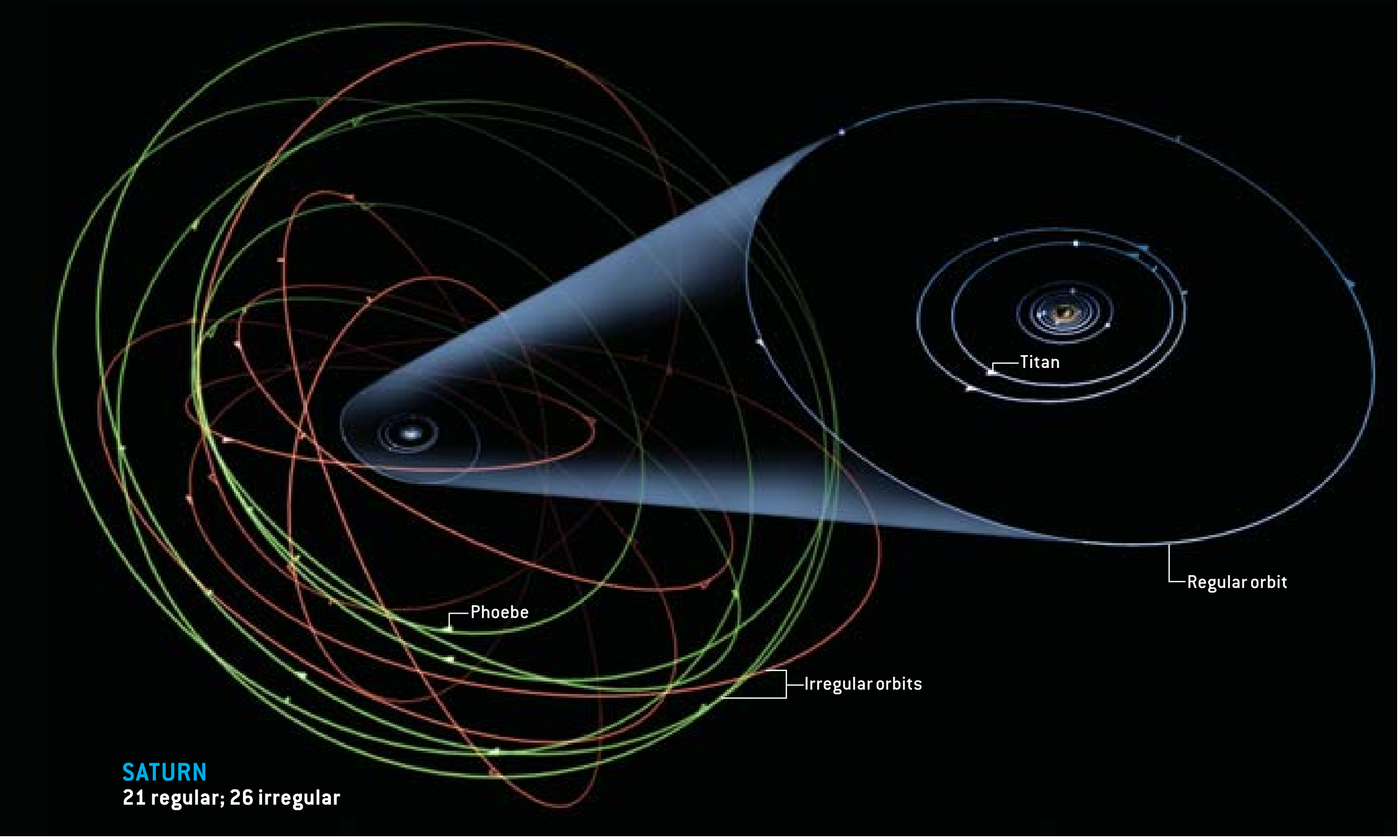


DON DIXON

Overview/*Irregular Moons*

- Astronomers used to think that most planetary moons formed from disks around their respective planets—reproducing, in miniature, the formation of the solar system itself. These moons orbit in the same plane as the planet's equator and in the same direction as the planet's spin. The few bodies not fitting this pattern were deemed "irregular."
- A recent flood of discoveries using advanced digital detectors shows that irregular moons are actually the majority. Their long, looping, slanted orbits indicate that they did not form in situ but instead in paths encircling the sun. In essence, they are asteroids or comets that the planets somehow captured.
- Neither the source region nor the mechanism of capture is well understood. The moons might have come from the Kuiper belt beyond Neptune or from regions closer in. Their capture may have involved collisions or other interactions in a younger, more densely populated solar system.

The full extent of the system of moons around Saturn was barely known until recent years. The satellites fall into two broad categories: regular moons (*blue*), such as Titan and Iapetus, have tight, nearly coplanar orbits; irregular moons, such as Phoebe, have wider, variously oriented orbits. Some revolve in the same direction as Saturn rotates (*red*); others go the opposite way (*green*). Similar systems surround the other giant planets (*far left*). These diagrams show a sampling of the total number of moons.



The rotation is so rapid that it is not even accurate to represent the paths as closed loops. Instead the moons trace out strange, looping trajectories akin to figures from the children's toy Spirograph.

Cosmic Polyrhythm

WHEN THE VARIOUS influences on the moons act in synchrony, the situation gets especially complex. For instance, if the rate of precession matches the rate at which the host planet orbits the sun, the moon is said to be in an "evection" resonance. The otherwise modest effects of solar gravity accumulate over time, destabilizing the orbit; the ellipse elongates to such an extent that the moon either collides with the planet (or one of its larger moons) or breaks out of the Hill sphere and falls into the gravitational clutches of the sun. Prograde orbits are more vulnerable than retrograde ones. If irregular moons were originally equally likely to be either prograde or retrograde, this resonance could explain why most moons are now retrograde.

Another resonance, known as the Kozai resonance, couples the tilt and shape of the orbit. Moons that are hauled into inclined orbits wind up on highly stretched ellipses, again leading potentially to their ejection or destruction. That may be why observers have found no moons with inclinations between 50 and 130 degrees. In short, the irregular moons we see today appear to be the survivors of gravitational interactions that cleared out many of their brethren.

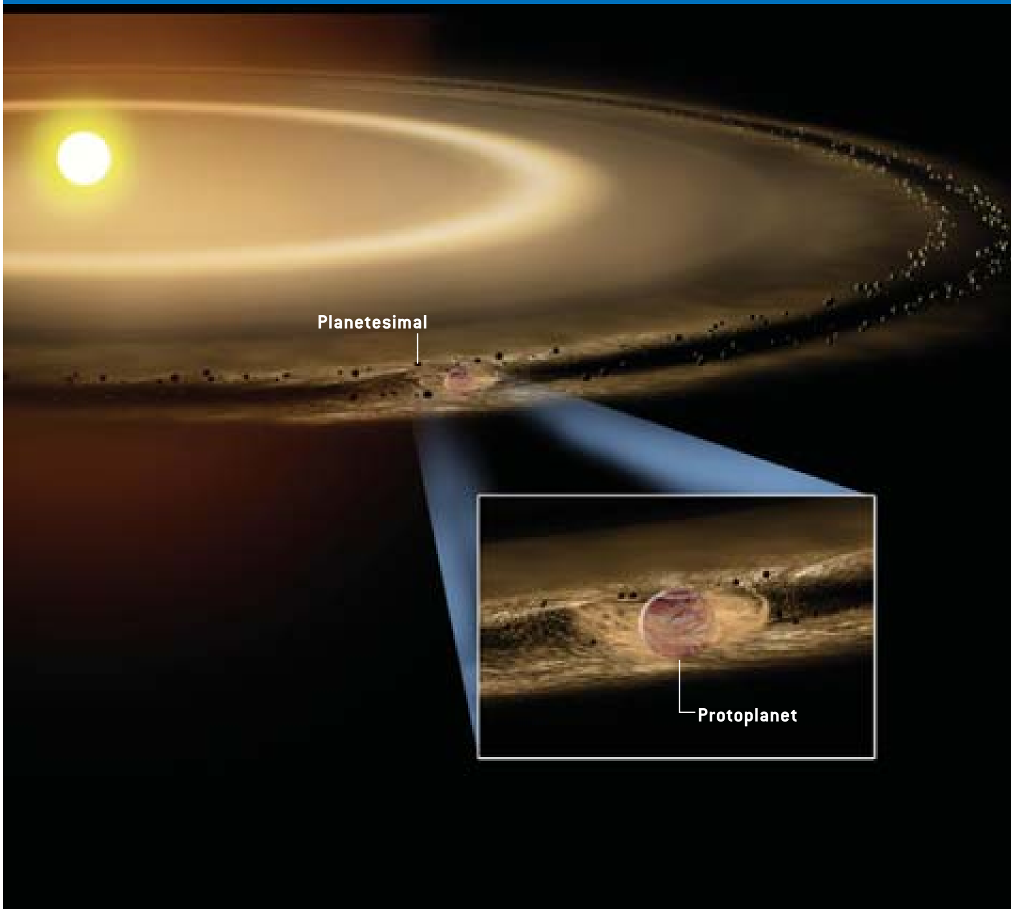
Still other features of the orbits require processes beyond those of gravity. The moons belong to distinct groups, or families, having similar orbits. Jupiter's groups, for instance, have up to 17 members each. The most straightforward interpretation is that the members of a group are pieces of a larger moon that was shattered by an impact and continue to follow in that body's orbit. If so, many of the irregular moons we see today are the second generation—one step removed from the original population.

David Nesvorny of the Southwest Research Institute in Boulder, Colo., and his collaborators have modeled the collisional disruption of moons in detail. They find that in the present day it is rare for a moon to collide with another moon or with an interplanetary body such as a comet. Therefore, the existence of groups hints at an earlier time when the populations of irregular moons or comets (or both) were much larger than now and collisions were much more frequent.

Beyond learning something about the orbits of irregular moons, astronomers have made some progress in discerning other properties. Most of the moons are so faint that they have been able to uncover very little about their composition. Tommy Grav of CfA and Terry Rettig of the University of Notre Dame have found, however, that moons within a group tend to have similar colors. Color is a proxy for composition, so this discovery implies a likeness in makeup—further supporting the idea that group members are fragments of a larger, bygone parent body.

One of the few irregular moons that astronomers know in any detail is Saturn's Phoebe, which NASA's Cassini spacecraft visited in June 2004. Cassini obtained very high resolution images, which showed a high density of craters on Phoebe's surface, and also recorded the spectra of reflected sunlight, which revealed water and carbon dioxide ices. The two irregular moons of Neptune seen by the Voyager 2 space probe, Nereid and Triton, also have icy surfaces. The ices hint that these objects formed relatively far from the sun, like comets. The irregular moons of Jupiter are pitch-black and appear to be devoid of ice, probably because they are closer to the sun and too warm for ice to be stable. In this sense, Jupiter's irregular moons closely resemble comets that have lost their volatile compounds.

HOW TO SNAG A MOON



The strange orbital properties of the irregular moons indicate that they started off in orbit around the sun and later were captured by their current host planets. Astronomers have proposed three capture mechanisms.

For all three, the initial stage is the formation of asteroid-size bodies called planetesimals. Many agglomerate to form the rocky cores of the giant planets. The leftovers are vulnerable to being captured.

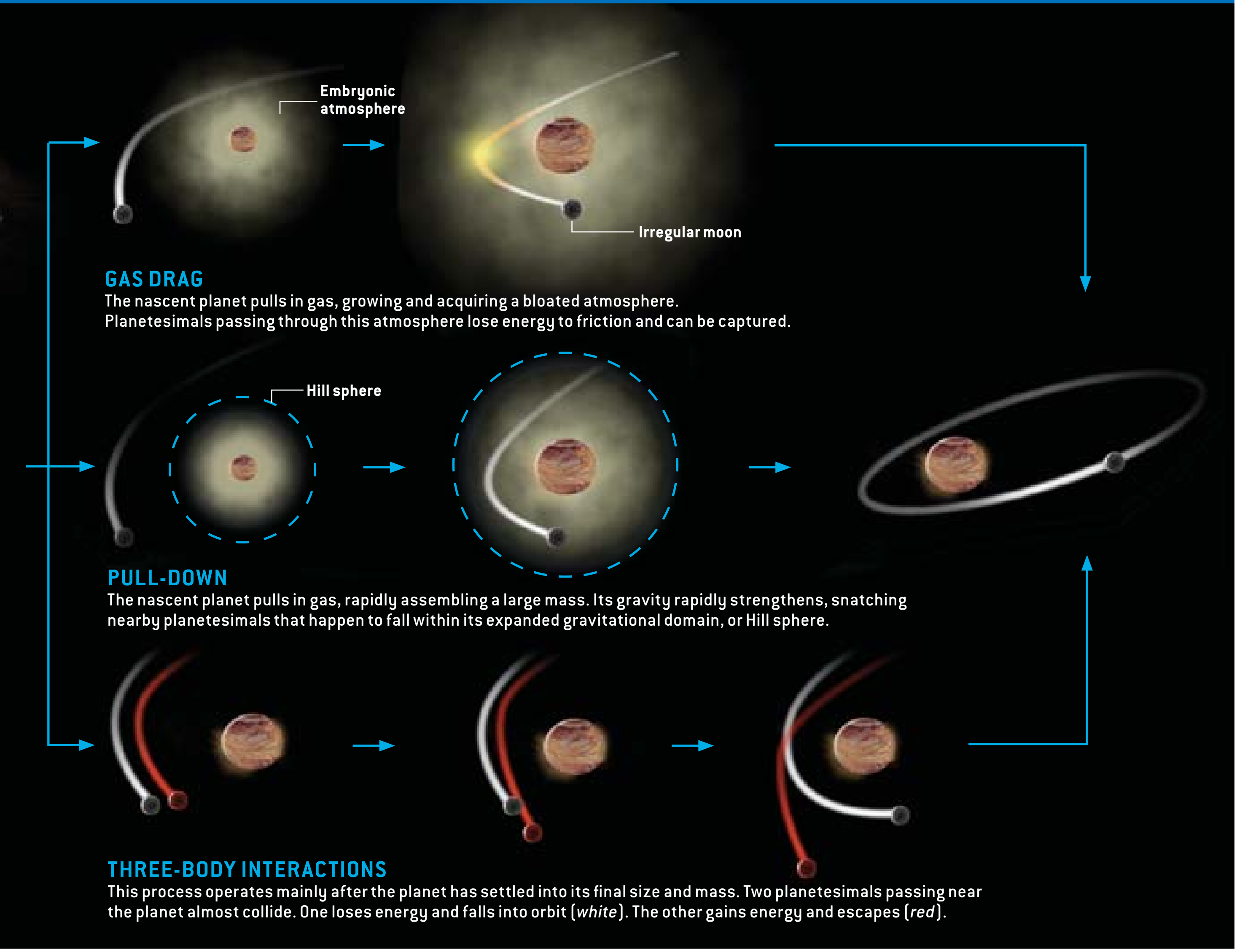
What a Drag

THE PROPERTIES of the irregular moons—especially their retrograde orbits—suggest that they did not form in situ. Instead they must be leftover planetary building blocks, like asteroids or comets, that originally orbited the sun and were somehow captured by the planets. Understanding how that happened is not easy. In the complex interplay of solar and planetary gravity, asteroids and comets are routinely pulled into short-lived orbits around the giant planets. Temporary capture is analogous to the trapping of leaves in a vortex on a windy autumn day. The leaves enter the vortex, swirl around for perhaps a few dozen times and then are blown out in an unpredictable way.

Examples of this type of capture include the well-known comet D/Shoemaker-Levy 9 (the "D" stands for "defunct"), which entered a temporary orbit around Jupiter sometime in the 20th century and rammed into the planet in 1994. Had it not met an untimely death, the comet would have been ejected back into heliocentric orbit within a few hundred years. Astronomers know of several objects that survived temporary capture by Jupiter and returned to orbiting the sun.

THE AUTHORS

DAVID JEWITT, SCOTT S. SHEPPARD and *JAN KLEYNA* are the world's most prolific discoverers of planetary moons. Jewitt traces his interest in astronomy to age seven, when he was astonished by a spectacular meteor shower visible against the sodium-lit night skies of industrial north London. He is now a professor at the University of Hawaii and a member of the National Academy of Sciences. Sheppard, his former graduate student, recently became a Hubble postdoctoral fellow in the department of terrestrial magnetism at the Carnegie Institution of Washington. Kleyna grew up on a farm in Maine, enjoys incomprehensible art-house cinema and is now a Parrent postdoctoral fellow at the University of Hawaii, where he mainly studies dark matter in dwarf galaxies.



But for a body to be permanently captured from heliocentric orbit into a bound, stable orbit around a planet, it must lose some of its initial energy. Essentially the body has to be slowed down to prevent it from escaping again. No efficient process of energy dissipation operates in the solar system today. Moon capture, then, must have occurred long ago, at a time when the solar system had different properties. In the 1970s theorists proposed three possible mechanisms, all functioning during or soon after the epoch of planet formation.

The first, advanced by James B. Pollack and Joseph A. Burns, then at the NASA Ames Research Center, and Michael E. Tauber of Cornell University, argues that the moons lost energy to friction generated as they passed through the vastly extended atmospheres of the embryonic gas giant planets. Jupiter and Saturn, quite unlike Earth and other terrestrial planets, are composed primarily of hydrogen and helium. Most probably, they formed when a core of rock and ice, of roughly 10 Earth-masses, pulled in vast quantities of gas from the primordial disk surrounding the young sun. Before settling into their modern, relatively compact forms, the planets may have passed through a transient, distended phase, during which their atmospheres extended hundreds of times farther than they do now.

In true Goldilocks style, a passing asteroid or comet would have met one of three distinct fates, depending on its size. If it was too small, it burned up in the bloated atmosphere, like a meteor. If it was too large, it plowed through unimpeded and continued in orbit about the sun. If it was just right, it slowed down and was captured. This process is a natural version of the aerobraking procedure that many planetary probes have used to enter orbit.

One problem with the gas-drag model is that it does not explain the presence of irregular satellites around Uranus and Neptune. Those planets are not gas giants but rather ice giants—dominated by rock and ice, with rela-

DON DIXON

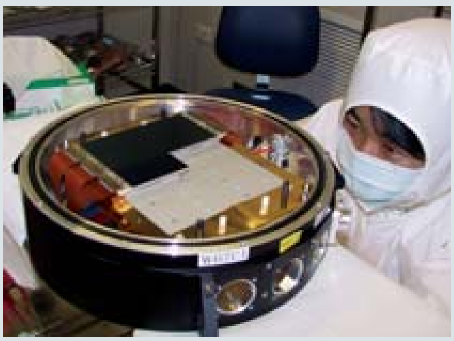
Watchers of the Skies

Far-flung, tiny, dimly lit: irregular moons are among the most challenging observational targets in the solar system. Finding them requires the world's most powerful survey telescopes—that is, instruments that scan broad swaths of the sky rather than concentrating on single, limited areas. Our team made most of our discoveries using the Canada-France-Hawaii Telescope and Subaru Telescope on Mauna Kea in Hawaii. They are equipped with digital detectors of more than 100 million pixels each.

The central problem is to distinguish objects in the solar system from more distant stars and galaxies. Observers use two methods. The first involves a distance measurement. We compare three images of the same area, spaced some time apart. During that time, Earth moves partway around the sun, changing our vantage point and causing bodies to appear to shift position; the closer the body, the more it appears to move.

The second method involves a velocity measurement. We take tens of images of one field, offset them depending on the expected orbital speed of the irregular moons we are looking for, and add them together. In the summed image, background stars appear as streaks and the irregular moons as bright dots. Because this method uses more images of a given area of sky, it is more sensitive to faint objects than the first approach but takes longer to perform a full survey. To make sure the bodies are moons rather than asteroids or comets, we monitor them for several months and work with Brian Marsden of the CfA to check whether they orbit their respective planets.

—D.J., S.S.S. and J.K.



FOR SURVEYING large areas of the sky, one of the best detectors is the Subaru Prime Focus Camera, a mosaic of 10 eight-megapixel CCD chips.

MOVING PINPRICK OF LIGHT: Jupiter's satellite S/2003 J14 was discovered on February 26, 2003, in these two images taken 39 minutes apart. The other objects here are background stars. Thought to be about two kilometers across, the moon has an orbit that stretches 31 million kilometers away from the giant planet.

tively modest veneers of hydrogen and helium. Because of their greater distance from the sun and the consequently lower density of material in the outer regions of the circumsolar disk, their cores took a longer time to reach the critical mass needed to precipitate gaseous collapse. Before that happened, the solar nebula had largely dissipated, and so Uranus and Neptune never had extended atmospheres, like those of Jupiter and Saturn. How can gas drag operate when there is not much gas?

Three's a Crowd

THE SECOND METHOD also places capture during the planetary growth phase. The accretion of gas onto the cores of the gas giants would have caused their mass to shoot upward in a self-reinforcing process, leading to sudden growth in the size of the Hill sphere around each planet. Asteroids and other objects that were unlucky enough to be nearby at the moment of this runaway growth would have found themselves trapped by the abruptly extended reach of the planets' gravity. This mechanism of capture was first expounded by Thomas A. Heppenheimer and Carolyn Porco, both then at the California Institute of Technology. They called it, somewhat confusingly, "pull-down" capture.

Like gas drag, however, this mechanism has trouble accounting for the moons around Uranus and Neptune, neither of which underwent a runaway growth in mass. Most models indicate that these planets grew slowly by accumulating asteroid- and comet-size bodies, perhaps taking tens or hundreds of millions of years to reach their present-day masses. Even Jupiter and Saturn would have had to grow within a matter of millennia to make pull-down capture work, and many modelers are uncomfortable with such a short growth timescale. An alternative model for forming Uranus and Neptune, proposed by Alan Boss of the Carnegie Institution of Washington, is that they started out as massive as Jupiter and Saturn and were whittled down by ionizing radiation from nearby massive stars. The irregular moons are even harder to understand in this model, because a shrinking planet would tend to lose moons rather than grabbing them.

In both the gas-drag and pull-down models, the irregular moons were acquired early in solar system history, probably before Earth had reached a recognizable state. A third and very different scenario was proposed in 1971 by Bepi Colombo and Fred Franklin, both then at CfA. They suggested that collisions between two bodies in the Hill sphere of a planet could dissipate enough energy to allow one of them to be captured. This idea, called three-body capture, received relatively little attention in the 35 intervening years, perhaps because such collisions are exceedingly rare now.

Yet newer work shows that no collision is needed. The three bodies need only interact gravitationally. If they exchange energy, one can gain energy at the expense of the others. The process is a scaled-up version of the gravitational slingshot effect that space mission planners use to boost deep space probes. This past May, Craig Agnor of the University of California, Santa Cruz, and Doug Hamilton of the University of Maryland suggested another form of three-body capture in which a binary object is sheared apart by the gravity of a planet, leading one component to be ejected and the other pulled into orbit.

Planetary Movements

THREE-BODY CAPTURE is appealing in light of the new finding that all four giant planets have retinues of irregular moons. The process works for both gas giants and ice giants. It does not require a massive gaseous envelope or runaway planetary growth; all it needs is a sufficient number of collisions or near misses occurring close to the planets. These types of interactions would have been most probable near the end of the planet formation epoch, after the Hill spheres had grown to their present proportions but before the leftover debris of planet formation had been cleared out. Three-body capture might be able to account for why each planet has roughly the same number of irregular moons: although Uranus and Neptune are less massive than Jupiter and Saturn, they are farther from the sun, so their Hill spheres are comparable in size.

Even if three-body interactions explain how the irregular moons were captured, where did they come from to begin with? Researchers have suggested two distinct possibilities. The moons could be asteroids and comets that had agglomerated in the same general region of the solar system as the planet that eventually snatched them. Most of their cohorts were incorporated into the bodies of the planets or catapulted out of the solar system. The irregular moons were the lucky ones, neither eaten nor consigned to wander in the rarefied space between the stars.

Another possibility emerges from a recent model in which the solar system remained choked with debris until some 700 million years after the planets formed. Strong gravitational interactions between Jupiter and Saturn then set up oscillations that shook the entire system. Billions of asteroids and comets were scattered as the major planets lurched into their present, more stable orbits. A tiny fraction of the scattered bodies could have been captured. In this scenario, proposed last year by K. Tsiganis and his colleagues of the Observatory of Côte d'Azur, most of the bodies shaken loose originally formed beyond Neptune in the Kuiper belt [see "The Kuiper Belt," by Jane X. Luu and David C. Jewitt; SCIENTIFIC AMERICAN, May 1996].

Spectral measurements should one day be able to test these two hypotheses. If the irregular moons of different planets have different compositions, it would support the first hypothesis, in which moons formed near their eventual host planets. If they have similar compositions, that would argue for the second hypothesis, in which the moons all formed together and then dispersed. Thus, the moons could reveal whether the solar system went through a turbulent rearrangement.

LARGEST IRREGULAR MOON, Neptune's Triton, has baffled scientists since its discovery in 1846. Recent work suggests that it and a partner orbited the sun in mutual embrace, until Neptune sundered them and claimed Triton as its own.

NASA/JPL

Exploration of the irregular moon systems is ongoing. Two things are already evident: First, the capture of these moons must have occurred early in the solar system's history, either in association with planet formation or with its immediate aftermath. The modern solar system simply offers no suitable mechanism through which moons could be captured. Second, the similarities among the irregular moon populations of all four outer planets suggest that they arose by three-body interactions, the only known mechanism that is about as effective for Neptune as it is for Jupiter.

Like skid marks on a road after a car crash, the irregular moons swooping around the giant planets provide us with tantalizing clues about past events that we could never have witnessed directly. SA

MORE TO EXPLORE

The Discovery of Faint Irregular Satellites of Uranus. J. J. Kavelaars et al. in *Icarus,* Vol. 169, No. 2, pages 474–481; June 2004.

Discovery of Five Irregular Moons of Neptune. Matthew J. Holman et al. in *Nature,* Vol. 430, pages 865–867; August 19, 2004.

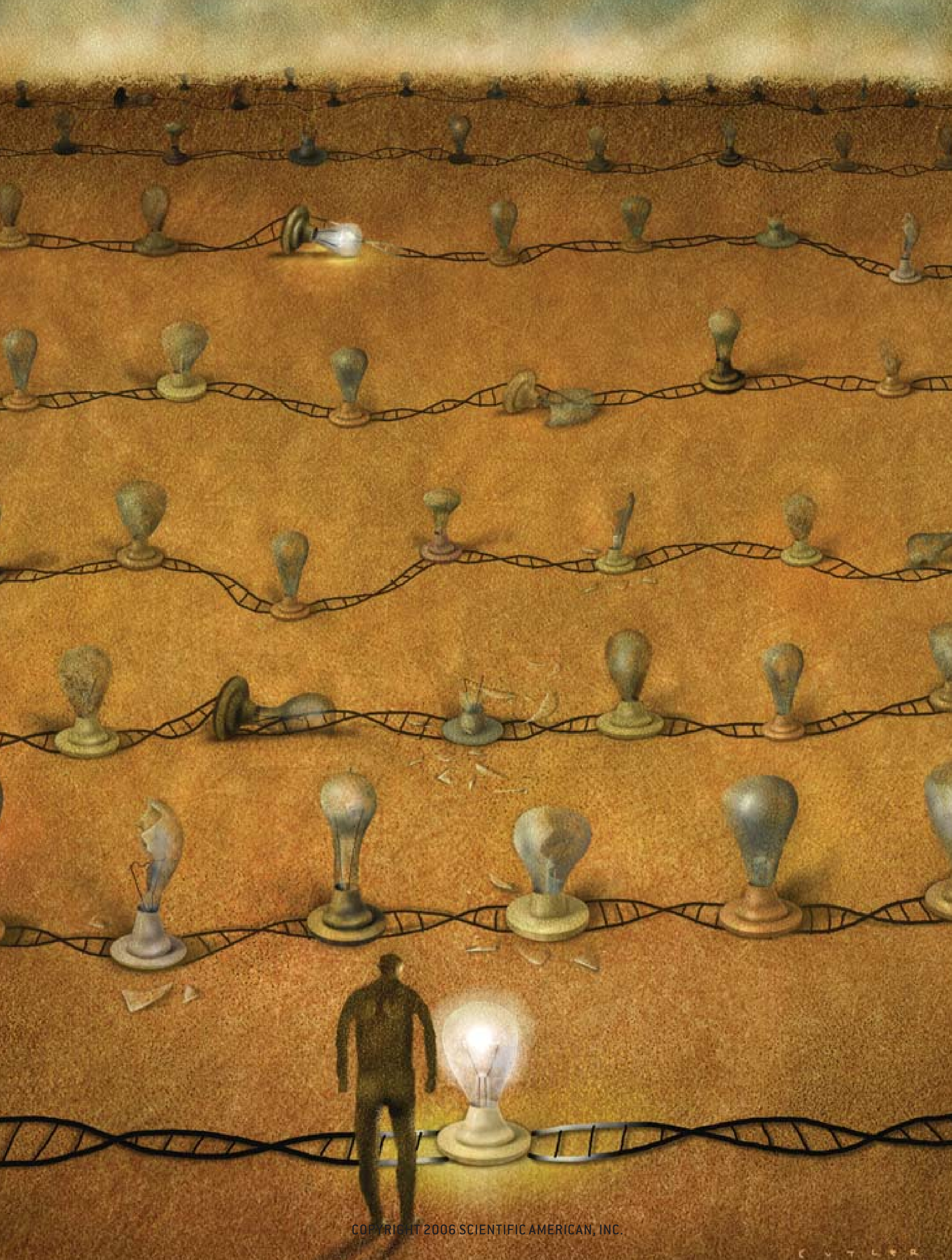
Photometry of Irregular Satellites of Uranus and Neptune. Tommy Grav, Matthew J. Holman and Wesley C. Fraser in *Astrophysical Journal,* Vol. 613, No. 1, pages L77–L80; September 2004. Available online at **arxiv.org/abs/astro-ph/0405605**

Irregular Satellites in the Context of Giant Planet Formation. David Jewitt and Scott Sheppard in *Space Science Reviews,* Vol. 116, Nos. 1–2, pages 441–456; January 2005.

Cassini Imaging Science: Initial Results on Phoebe and Iapetus. C. C. Porco et al. in *Science,* Vol. 307, pages 1237–1242; February 25, 2005.

Neptune's Capture of Its Moon Triton in a Binary-Planet Gravitational Encounter. Craig B. Agnor and Douglas P. Hamilton in *Nature,* Vol. 441, pages 192–194; May 11, 2006.

Hawaii Irregular Satellite Survey Web site: **www.ifa.hawaii.edu/~jewitt/irregulars.html**



Disabled genes, molecular relics scattered across the human genomic landscape, have a story of their own to tell. And it is still unfolding

The Real Life of Pseudogenes

By Mark Gerstein and Deyou Zheng

Our genetic closet holds skeletons. The bones of long-dead genes—known as pseudogenes—litter our chromosomes. But like other fossils, they illuminate the evolutionary history of today's more familiar forms, and emerging evidence indicates that a few of these DNA dinosaurs may not be quite so dead after all. Signs of activity among pseudogenes are another reminder that although the project to sequence the human genome (the complete set of genetic information in the nuclei of our cells) was officially finished, scientists are still just beginning to unravel its complexities.

It is already clear that a whole genome is less like a static library of information than an active computer operating system for a living thing. Pseudogenes may analogously be vestiges of old code associated with defunct routines, but they also constitute a fascinating record contained within the overall program of how it has grown and diversified over time. As products of the processes by which genomes remodel and update themselves, pseudogenes are providing new insights into those dynamics, as well as hints about their own, possibly ongoing, role in our genome.

DAVE CUTLER STUDIO

Copied, Not Fake

"FALSE" GENES, which look like real genes but have no apparent function, were first recognized and dubbed pseudogenes during the late 1970s, when early gene hunters began trying to pinpoint the chromosomal locations associated with production of important molecules. For example, while seeking the gene responsible for making betaglobin, a key component of the hemoglobin protein that transports oxygen through the bloodstream, scientists identified a DNA sequence that looked like a globin gene but could not possibly give rise to a protein. Essential functional parts of the gene's anatomy were disabled by mutations, making it impossible for cellular machinery to translate the gene into a useful molecule.

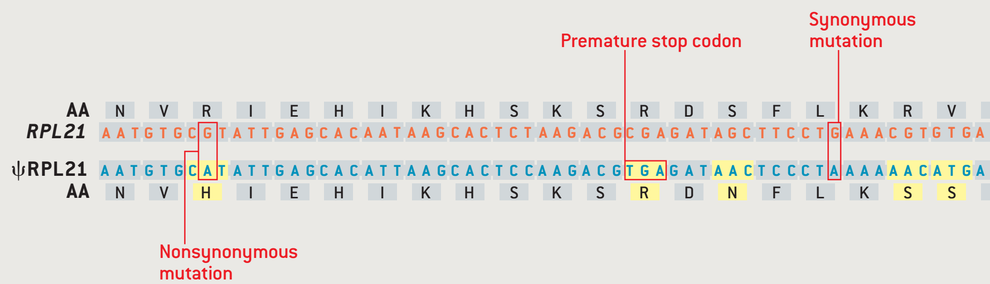
Only the far more recent completion of sequencing projects covering the full genomes of humans and other organisms allowed geneticists to get an aerial view of the genomic landscape and to appreciate how riddled with such oddities it is. The human genome is made up of more than three billion pairs of nucleotides, the building blocks of DNA molecules. Yet less than 2 percent of our genomic DNA directly encodes proteins. Perhaps a third is noncoding sequences within genes, called introns. The remaining tracts between genes constitute the great majority of our DNA, and much of it is effectively genomic dark matter whose function is still largely a mystery. It is in these seemingly barren expanses that most pseudogenes are randomly scattered like rusted car parts on the landscape—and in surprising numbers.

PSEUDOGENE BIRTH AND GENE DEATH

Two distinct processes can duplicate genes, and together they allow genomes to grow and diversify over evolutionary time. If errors in a copy destroy its ability to function as a gene, however, it becomes a pseudogene instead (*right*). The mutations that can kill a gene (*below*) range from gross deletions (such as the loss of the promoter region that signals the start of a gene sequence) to minute changes in the DNA sequence that skew the meaning of the gene's protein-encoding segments, called exons.

GENE DEATH

Genes die and become pseudogenes when mutations generated during the gene-copying process or accumulated over time render them incapable of giving rise to a protein. Cellular machinery reads the DNA alphabet of nucleotide bases (abbreviated A, C, G, T) in three-base increments called codons, which name an amino acid building block in a protein sequence or encode "stop" signals indicating the end of a gene. Even single-base mutations in codons can alter their amino acid meaning, and base deletions or insertions can affect neighboring codons by shifting the cellular machinery's reading frame. The alignment shown here of a partial sequence for a human gene (*RPL21*) against one of its pseudogene copies (ΨRPL21), along with each codon's corresponding amino acid (AA), illustrates some of the disabling mutations typically found in pseudogenes.



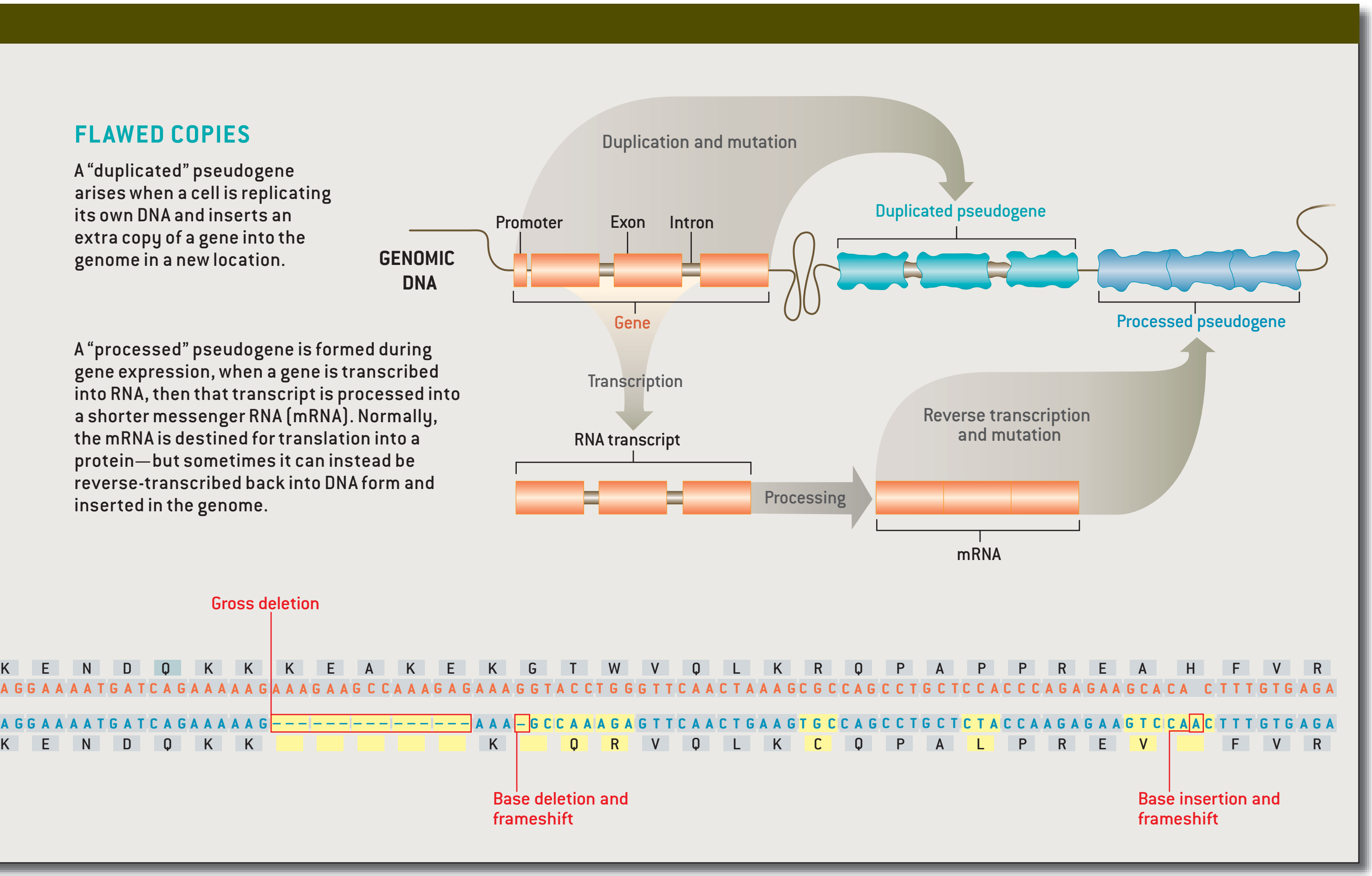
With ongoing annotation of the human genome sequence, our research group, along with others in Europe and Japan, have identified more than 19,000 pseudogenes, and more are likely to be discovered. Humans have only an estimated 21,000 protein-coding genes, so pseudogenes could one day be found to outnumber their functional counterparts. Their sheer prevalence has raised many questions, including how they came into existence, why there are so many of them and why, if they are really useless, they have been retained in our genome for so long.

The answer to the first question is already fairly well understood. A small fraction of pseudogenes are believed to have once been functional genes that simply "died" from disabling changes to their nucleotide sequences. But most pseudogenes are disabled duplicates of working genes. They may have been dead on arrival, having suffered lethal damage during the copying process, or they may have accumulated debilitating mutations over time that collectively rendered them incapable of functioning.

Critical to a working gene is an intact anatomy that includes uninterrupted nucleotide spans called exons, which correspond to amino acid sequences in the encoded protein. Introns typically separate the exons, and at the beginning of a gene is a segment known as a promoter that serves as the starting point for cellular machinery to recognize the

Overview/*The Pseudogenome*

- Pseudogenes are the molecular remains of broken genes, which are unable to function because of lethal injury to their structures.
- The great majority of pseudogenes are damaged copies of working genes and serve as genetic fossils that offer insight into gene evolution and genome dynamics.
- Identifying pseudogenes involves intensive data mining to locate genelike sequences and analysis to establish whether they function.
- Recent evidence of activity among pseudogenes, and their potential resurrection, suggests some are not entirely dead after all.



FLAWED COPIES

A "duplicated" pseudogene arises when a cell is replicating its own DNA and inserts an extra copy of a gene into the genome in a new location.

A "processed" pseudogene is formed during gene expression, when a gene is transcribed into RNA, then that transcript is processed into a shorter messenger RNA (mRNA). Normally, the mRNA is destined for translation into a protein—but sometimes it can instead be reverse-transcribed back into DNA form and inserted in the genome.

LUCY READING-IKKANDA

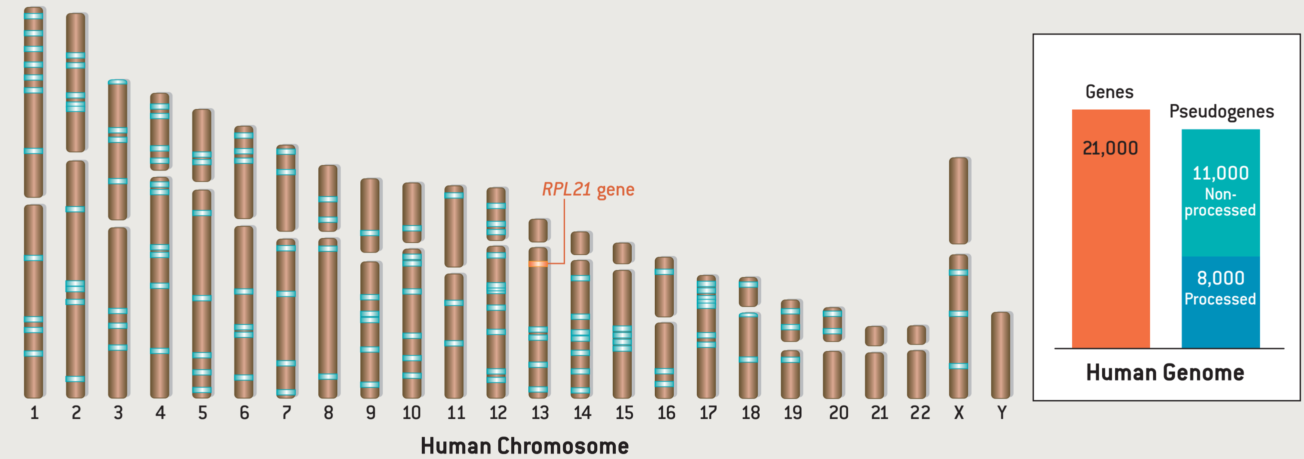
gene on a chromosome. When a cell expresses a gene, it first recruits essential molecular players to the promoter site, which travel down the gene's length, transcribing it into a preliminary RNA copy. A splicing process next cuts introns out of the raw transcript and joins exonic sequences to produce an edited messenger RNA (mRNA) version of the gene. The mRNA is then read by a ribosome, a cellular machine that translates its sequence into the string of amino acids that forms a protein, the molecule that ultimately carries out the gene's function.

Pseudogenes can be born in two ways, each of which yields a distinctive facsimile of the original parent gene. Just before dividing, a cell duplicates its entire genome, and during that process, an extra copy of a gene can be inserted into the chromosomes in a new location. Alternatively, a new version of a gene can also be created through reverse transcription: during gene expression, the mRNA is copied back into a sequence of DNA that is inserted into the genome. Known as retrotransposition, this phenomenon can occur because of the activity of another type of transposable genetic actor, known as a long interspersed nuclear element, or LINE, that behaves like a genomic virus. LINEs carry their own machinery for making DNA copies of themselves to insert into the genome, and mRNA transcripts that are in the vicinity when LINEs are active can be swept up and retrotransposed as well.

These two processes, duplication and retrotransposition, are major forces that remodel genomes over the course of evolutionary time, generating new variation in organisms. They are the means by which genomes grow and diversify, because many replicated genes remain active. But if the gene copy contains disabling typos or is missing pieces of the original, such as the promoter, it will become a pseudogene. Those arising from duplication of an entire gene are recognizable because they contain both introns and exons. Pseudogenes made from mRNA lack introns and are described as processed pseudogenes.

Although the overall distribution of most pseudogenes across human chromosomes seems completely random, certain kinds of genes are more likely to give rise to pseudogenes. Geneticists organize functional genes into families based on their similarity to one another in both sequence and purpose. Only about a quarter of these family groups are associated with a pseudogene, and some families have spun off an inordinate number of copies. For example, the family of 80 human genes that produce ribosomal proteins has given rise to about 2,000 processed pseudogenes—roughly a tenth of the genome's identified total. In one extreme case, a single ribosomal protein gene known as *RPL21* has spawned more than 140 pseudogene copies.

This disparity probably derives from



PSEUDOGENE DESCENDANTS (*blue*) of the ribosomal protein gene *RPL21* (*orange*) are scattered across the human chromosomal landscape. Overall distribution of pseudogenes in the human genome appears to be completely random, although some local genome regions tend to contain more pseudogenes. Those DNA regions may be analogous to certain geochemical environments that better preserve mineral fossils. Identification of genes and pseudogenes is an ongoing process, but to date more than 19,000 pseudogenes have been identified in the human genome—only slightly less than the current tally of around 21,000 human genes (*inset*). About 8,000 of our pseudogenes are processed; the rest include duplicated pseudogenes and other nonprocessed subcategories.

the activity levels of different genes. Those responsible for basic cellular housekeeping functions, such as the genes in the ribosomal protein family, are abundantly expressed, providing more opportunities to create processed pseudogenes.

Because pseudogenes have been accreting this way in our genomes for so long, some are relics of genes eliminated during the course of evolution, and no functional version exists today. Others are copies of a gene that has so evolved over time, the pseudogene's sequence may reflect an older, earlier version of its parent. Consequently, intergenic regions can be seen as vast molecular fossil beds offering a silent record of events in our evolutionary past.

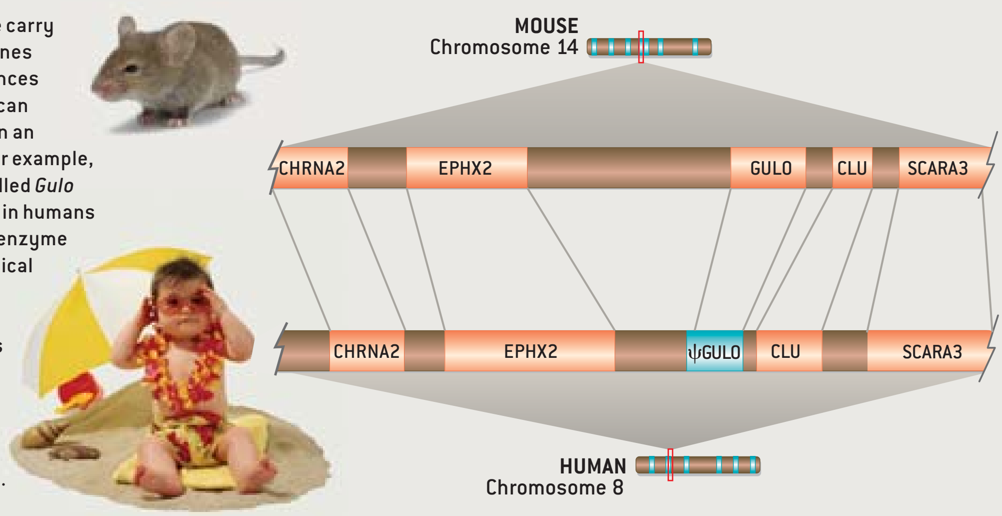
Family Histories

THE PRINCIPLES of natural selection appear to extend to individual genes, strongly constraining mutations in the sequences of functional genes. Beneficial mutations in a gene that improve the organism's fitness therefore tend to be preserved, whereas a sequence change that impairs a gene's function leads it to be discarded.

Once consigned to the genomic junk pile, however, pseudogenes are released from this selection pressure and are free to accumulate all kinds of mutations, including changes that would be deleterious to normal genes. Scientists can use this tendency to derive a kind of molecular clock from the nucleotide changes in pseudogenes and use it to study the overall dynamics and evolution of the genome. Tracking the evolutionary path of genes and pseudogenes helps molecular biologists to uncover instances of gene birth and death just as the study of mineral fossils tells paleontologists about the creation and extinction of species.

Our group has surveyed pseudogenes in the genomes of many forms of life,

CHROMOSOMES of humans and mice carry a very similar array of functional genes (*orange*) but reveal distinct differences in their pseudogenes (*blue*), which can highlight important turning points in an organism's evolutionary history. For example, the counterpart of a mouse gene called *Gulo* has become a pseudogene (ΨGulo) in humans and other primates. *Gulo* makes an enzyme that is the last element in a biochemical pathway for synthesizing vitamin C. Most mammals possess the active gene, but the primate lineage seems to have lost it more than 40 million years ago. When the *Gulo* gene became a pseudogene, primates became dependent on food sources of vitamin C to avoid scurvy.



LUCY READING-IKKANDA (*illustrations*); DIGITAL VISION/GETTY IMAGES (*mouse*); PAUL BARTON *zefa/Corbis* (*baby*)

Differences in pseudogenes offer hints about diverse life histories.

ranging from bacteria to more complex organisms, such as yeast, worms, flies and mice, and their prevalence across a wide range of creatures is striking. The number of pseudogenes in different genomes varies greatly, more so than genes, and it is not readily predictable, because it is neither strictly proportional to the size of a genome nor to the total number of genes. Comparisons of pseudogenes in related genomes can nonetheless reveal important information about the history of specific genes and the general workings of molecular evolution.

One of the largest known gene families in mammals, for example, consists of more than 1,000 different genes encoding olfactory receptors, the cell-surface proteins that confer our sense of smell. Detailed analyses of olfactory receptor (OR) genes and pseudogenes by Doron Lancet and Yoav Gilad of the Weizmann Institute of Science in Rehovot, Israel, show that humans have lost a large number of functional olfactory receptor genes during evolution, and we now have fewer than 500 of them in our genome. For comparison, versions of about 300 human olfactory receptor pseudogenes are still functional genes in the genomes of rats and mice.

This difference is not surprising given that most animals depend more for their survival on the sense of smell than humans do. In fact, humans have considerably more olfactory receptor pseudogenes than chimpanzees do, indicating that we lost many of those functional genes after our split from the ape lineage. Apes, however, have a higher proportion of olfactory receptor pseudogenes (30 to 40 percent of the OR family) than rodents or dogs do, suggesting that some influence has allowed the entire ape lineage to get by with a somewhat reduced sense of smell.

Lancet and his colleagues found in studies of apes, monkeys and other distant primate cousins that the greatest loss of olfactory receptor genes—that is, the greatest increase in OR pseudogenes—occurred in ape and monkey lineages that evolved the ability to see color in three wavelengths of visible light. The link may suggest that a sensory trade-off took place over time in the primate lineage when better eyesight made an acute sense of smell less critical.

Often, genes involved in an organism's response to its environment are subject to extensive duplication and diversification over time, leading to large gene families, such as the olfactory receptor repertoire. Many dead-on-arrival pseudogene copies are an immediate byproduct of this process. But the subsequent death of additional duplicates, which gives rise to new pseudogenes, is also frequently connected to changes in an organism's environment or its circumstances. Consequently, differences in the pseudogenes of animals offer hints about their diverse life histories that are not as easily detected in comparisons of working genes, which are strongly constrained by their function.

Analysis of the mouse genome, for example, has shown that 99 percent of human genes have a corresponding version in the mouse. Although the human and mouse lineages diverged some 75 million years ago, nearly all of the human genome can be lined up against equivalent regions, known as syntenic blocks, in the mouse genome. Yet despite this similarity in functional genes and overall genome structure, just a small fraction of the known human pseudogenes have an obvious counterpart in the mouse.

What is more, some of the specific gene families giving rise to pseudogenes differ significantly between mouse and human. Using the rate of sequence decay relative to the parent genes to determine their age, it is also clear that many pseudogenes in the human and mouse genomes have arisen at different times. These observations indicate that very disparate events have led to independent bursts of retrotransposition that created pseudogenes in each of the lineages.

THE AUTHORS

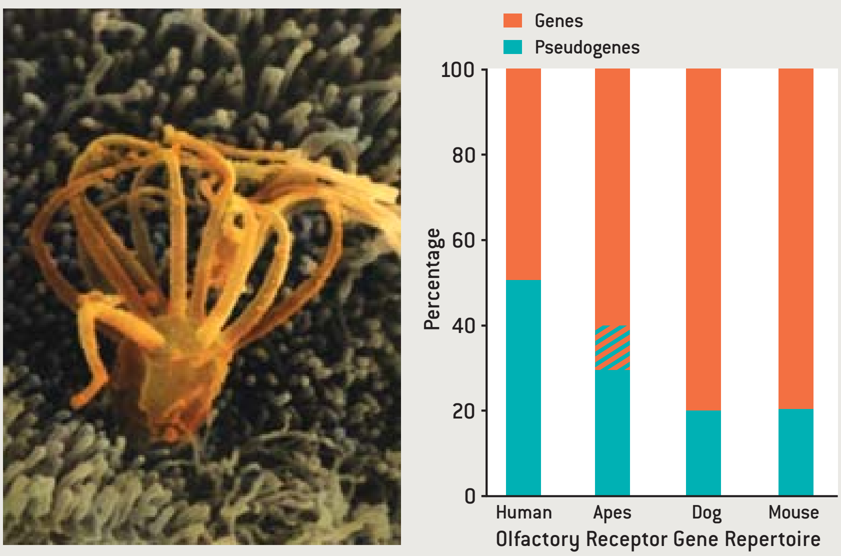
MARK GERSTEIN and *DEYOU ZHENG* are bioinformaticians at Yale University, where Gerstein is A. L. Williams Professor of Biomedical Informatics and co-director of the Yale Program in Computational Biology and Bioinformatics. Zheng, after completing his Ph.D. at Rutgers University, joined Gerstein's group in 2003 to begin investigating pseudogene activity and evolution. Both authors were initially interested in studying molecular structure and simulation, as described in Gerstein's previous article for *Scientific American* with Michael Levitt, "Simulating Water and the Molecules of Life" (November 1998). But Gerstein and Zheng were intrigued by the enormous data analysis challenges posed by the sequencing of the human genome and chose to start scanning and sifting the regions of DNA between genes.

Scanning and Sifting

STUDIES OF PSEUDOGENES in their own right are really just beginning, because these fossil genes were long viewed as little more than a nuisance. Early efforts to catalogue pseudogenes were largely driven by the need to distinguish them from true genes when annotating genome sequences. Identifying pseudogenes is not as straightforward as recognizing genes, however. Based on characteristic elements, pattern-seeking computer algorithms can scan DNA sequences and identify genes with moderate success. Recognition of pseudogenes, in contrast, relies primarily on their similarity to genes and their lack of function. Computers can detect similarity by exhaustively aligning chunks of intergenic DNA against all possible parent genes. Establishing a suspected pseudogene's inability to function is more challenging.

Just as a living organism can die of many different causes, a variety of deleterious mutations affecting any step in the process of making a protein can disable a copied gene, turning it into a pseudogene. But the sequence itself can offer clues to whether a mutation is debilitating. We can look for premature "stop" signs, as well as insertions or deletions of nucleotides that shift the reading frame of cellular machinery that decodes the gene's information for making a protein. These disablements cannot be tolerated by true genes and are thus typical manifestations of pseudogenes.

More subtly, the theory of neutral evolution introduced by mathematical biologist Motoo Kimura in the 1960s holds that nonfunctional DNA sequences can change freely, without the constraint of natural selection. Thus, individual nucleotide mutations can be divided into two types: those that would preserve the amino acid sequence of the protein encoded by a gene, known as synonymous changes, and nonsynonymous changes that would alter the meaning of the sequence. Because changing a protein's amino acid sequence can abolish its function, a gene under selective pressure will be more likely to contain synonymous mutations, whereas a nonfunctional DNA sequence will not be subject to that constraint.



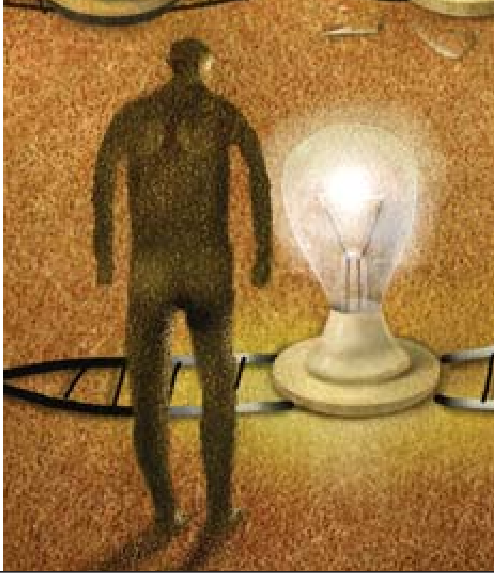
CILIA projecting from human olfactory epithelium (*left*) are studded with invisible odorant molecule receptors that detect smells. A family of more than 1,000 genes encoding those olfactory receptors in mammals has been identified, although individual species vary widely in their total number of olfactory receptor genes and the percentage of that repertoire that has died and become pseudogenes. Large-scale pseudogenization is most often seen among genes that, like the olfactory receptor family, are responsible for responses to the environment. An organism's pseudogenes may therefore reflect species-specific changes in circumstances during its evolutionary history.

Comparison of pseudogenes among genomes has revealed a puzzling phenomenon, however: a few pseudogenes appear to be better preserved than one would expect if their sequences were drifting neutrally. Such pseudogenes may therefore be under evolutionary constraint, which implies that they might have some function after all. One way to try to ascertain whether pseudogenes are functioning is to see whether they are transcribed into RNA. Recent experiments by Thomas Gingeras of Affymetrix and by Michael Snyder of Yale University have found that a significant fraction of the intergenic regions in the human genome are actively transcribed. In their studies, in fact, more than half the heavily transcribed sequences map to regions outside of known genes. What is more, a number of those transcriptionally active intergenic areas overlap with pseudogenes, suggesting that some pseudogenes may have life left in them.

Our research group is part of a consortium of laboratories working to understand what is going on in the dark matter of the genome. We are now in the pilot phase of a project to create an "encyclopedia of DNA elements" (referred to as ENCODE) whose ultimate goal is to identify all of the genome's parts and their function. Previous studies as well as preliminary ENCODE data indicate that at least one tenth of the pseudogenes in the human genome are transcriptionally active. Knowing that so many pseudogenes are transcribed does not tell us their function, but together with evidence that certain pseudogenes are better preserved than background intergenic sequences, it certainly challenges the classical view of pseudogenes as dead.

One possibility is that pseudogenes play some ongoing part in regulating the activity of functional genes. Molecular biologists have come to understand in recent years that many genes in higher organisms do not code for a final protein product, but instead their RNA transcripts act to control other genes. These regulatory RNA molecules can variously activate or repress another gene or can interfere with the translation of that

RICHARD COSTANZO *Virginia Commonwealth University (micrograph)*; COLORIZATION BY EMILY HARRISON; LUCY READING-IKKANDA *(illustration)*



Nature may have figured out a way to reuse the broken parts of genes.

gene's mRNA transcript into a functional protein. And at least two examples of pseudogenes behaving in a similar manner have been documented so far.

The first was reported in 1999 by Michael O'Shea's research group at the University of Sussex in England. The investigators found that in the neurons of a common pond snail, both the gene for nitric oxide synthase (NOS) and its related pseudogene are transcribed into RNA but that the RNA transcript of the NOS pseudogene inhibits protein production from the transcript of the normal NOS gene.

Then, in 2003, Shinji Hirotsune of the Saitama Medical School in Japan traced deformities in a group of experimental baby mice to the alteration of a pseudogene. The inactivity of an important regulatory gene called *Makorin1* had derailed the development of the mice, but Hirotsune had not done anything to *Makorin1*. He had accidentally disrupted the Makorin1 pseudogene, which affected the function of its counterpart, the *Makorin1* gene.

Perhaps two dozen examples of specific pseudogenes that appear to be active in some way—often only in certain cells of an organism—have been identified, although the findings are still preliminary. Because many pseudogenes have sequences highly similar to those of their parent genes, it is very tempting to speculate that the NOS and Makorin1 pseudogenes are not just isolated cases. Yet it is hard to imagine that these two pseudogenes had the specific roles they now perform when they first arose. Instead their activity may be the result of selection preserving happy accidents or of nature having figured out an efficient way to reuse the broken parts of genes by converting them into regulatory elements.

Protogenes

AN EXCITING ERA of molecular paleontology is just beginning. We have barely scratched the surface of the pseudogene strata, and once we drill deeper, the number of identified pseudogenes will most likely grow and we may find more surprises. Large-scale pseudogene identification is a very dynamic data-mining process. Current techniques rely heavily on sequence comparison to well-characterized genes, and although they can readily identify recently generated pseudogenes, very ancient and decayed sequences are probably escaping detection. As the sequence and annotation of the human genome itself are refined and updated, characterization of pseudogenes will improve as well.

Recent hints that not all pseudogenes are entirely dead have been intriguing, and some evidence also exists for the possibility of pseudogene resurrection—a dead gene turning back into a living one that makes a functional protein product. Careful sequence comparisons have shown that one cow gene for a ribonuclease enzyme was a pseudogene for much of its history but appears to have been reactivated during recent evolutionary time. Slight differences in the pseudogene complements of individual people have also been found—for example, a few olfactory receptor pseudogenes straddle the fence: in most people they are pseudogenes, but in some they are intact, working genes. These anomalies could arise if random mutation reversed the disablement that originally produced the pseudogene. Might they account for individuals' differing sensitivities to smell? Perhaps, although it is too early to guess at the scope or significance of this unexpected source of genetic variation among humans.

Our studies have suggested, however, that in yeast, certain cell-surface protein pseudogenes are reactivated when the organism is challenged by a stressful new environment. Thus, pseudogenes may be considered not only as dead genes (which nonetheless provide fascinating new insights into our past) but also as potentially unborn genes: a resource tucked away in our genetic closet to be drawn on in changing circumstances, one whose possible roles in our present and future genomes are just beginning to unfold. SA

MORE TO EXPLORE

Human Specific Loss of Olfactory Receptor Genes. Yoav Gilad et al. in *Proceedings of the National Academy of Sciences USA,* Vol. 100, No. 6, pages 3324–3327; March 18, 2003.

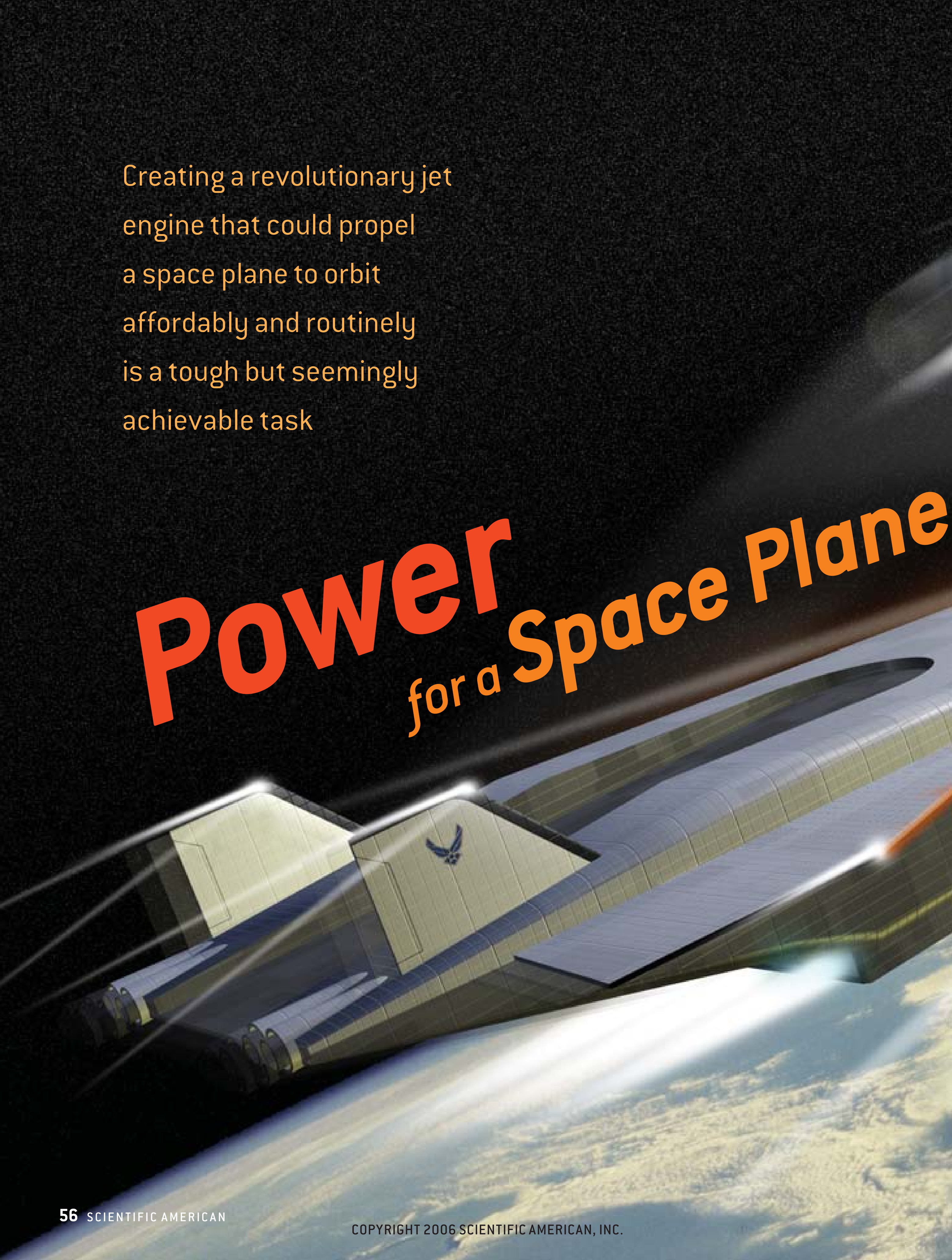
Pseudogenes: Are They "Junk" or Functional DNA? Evgeniy S. Balakirev and Francisco J. Ayala in *Annual Review of Genetics,* Vol. 37, pages 123–151; December 2003.

Large-Scale Analysis of Pseudogenes in the Human Genome. Zhaolei Zhang and Mark Gerstein in *Current Opinion in Genetics & Development,* Vol. 14, No. 4, pages 328–335; August 2004.

www.pseudogene.org/

Creating a revolutionary jet engine that could propel a space plane to orbit affordably and routinely is a tough but seemingly achievable task

Power for a Space Plane



By Thomas A. Jackson

Engineers have long dreamed of building an aircraft that could soar from a runway to outer space and then back again—the way Luke Skywalker's X-wing fighter does in the *Star Wars* movies. But one thing has stood in their way: jet engines need oxygen to burn fuel, and the upper atmosphere does not have enough of it to sustain combustion. So flying to space requires rocket propulsion, for which both fuel and oxidizer are carried inside the vehicle. In even the most advanced launch system in use today, the space shuttle, about half the launch weight is the liquid oxygen and solid oxidizer it must lug aloft to keep its rocket fuel burning all the way to orbit.

One answer could come from a supersonic combustion ramjet, known as a scramjet, which would scoop oxygen from the atmosphere as it ascends. The weight savings from capturing air in flight instead of carrying it means the scramjet can deliver about four times the thrust of a rocket for every pound of propellant consumed. At last, after decades of intermittent development, practical scramjets appear poised to take wing.

TWO-STAGE ORBITAL BOOSTERS, such as this conceptual Quicksat military space plane, may one day be able to send people and objects into space at lower cost than today's rockets. They would be made possible by successful supersonic combustion ramjet—scramjet—engines.

Researchers plan to conduct critical, full-scale ground tests of engines in 2007 and 2008, with a series of barrier-busting flight tests of the technology scheduled for 2009.

Unlike a rocket that punches straight up to orbit, a scramjet-powered craft would climb like an airplane on aerodynamic lift generated by its wings and fuselage, making it more maneuverable and safer (if a flight were aborted, the vehicle could glide back to earth). It would take off and accelerate to supersonic speeds using conventional jet engines. (Supersonic velocities begin at Mach 1, or 760 miles per hour at sea level.) The scramjets would then take over and propel it into the hypersonic regime—Mach 5 to Mach 15 (the theoretical limit of scramjet performance). Finally, small rocket engines would accelerate the payload the rest of the way to orbit. Mach 5 is five times the speed of sound, or about a mile every second. By comparison, the fastest air-breathing manned airplane, the now retired U.S. Air Force SR-71 Blackbird, topped out at velocities around Mach 3.2.

Such capabilities would lead to a revolution in flight. A reusable space plane's ability to fly like a conventional aircraft could help make such trips relatively routine, leading to dramatic reductions in the cost of sending people or objects into orbit. The new engine's tremendous performance could enable a military aircraft or missile to deliver weapons on target anywhere on the globe substantially sooner than is currently possible. Scramjets might someday even power long-range hypersonic passenger airliners, permitting, say, a two-hour flight from New York to Sydney.

Many research groups around the world are working to overcome the huge technical challenges involved in achieving hypersonic flight with scramjets. My discussion will focus on the U.S. Air Force and Pratt & Whitney Hypersonic Technology (HyTech) scramjet engine program, the one with which I am most familiar. Other vigorous development efforts are ongoing by the U.S. Navy, NASA, the Defense Advanced Research Projects Agency (DARPA) and engineering teams in Australia, the U.K., Japan and elsewhere [*see table on page 60*].

The Road to Flight

THE SCRAMJET is not a new propulsion concept. Initial patents were filed in the 1950s, and by the mid-1960s several scramjet engines had been ground-tested at speeds up to Mach 7.3. General Electric, United Technologies, Marquardt, Johns Hopkins University Applied Physics Laboratory and the NASA Langley Research Center built engines that typically burned hydrogen (the same fuel used by the space shuttle and many other liquid-rocket boosters). In the mid-1980s the U.S. government initiated the scramjet-powered National Aerospace Plane program. After an investment of almost $2 billion, however, the project was canceled in 1994 as part of budget cuts at the end of the cold war. In 2004 NASA completed its Hyper-X program, in which it successfully flew two hydrogen-fueled scramjet engines for a few seconds, each at a single speed and altitude. Late that year the X-43A scramjet research vehicle reached a record-setting Mach 9.6 [*see box on page 61*]. The ongoing U.S. Air Force effort is to use next-generation scramjet technology to accelerate a vehicle through a range of velocities and altitudes, fuel the engine with a liquid hydrocarbon and cool the engine structure with that same fuel.

Scramjets are members of a family of so-called air-breathing jet engines that rely on variations of a basic thrust-generating principle to operate at different ranges of speeds and altitudes. In general, jets work by compressing atmospheric air, combining it with fuel, burning the mixture and then expanding the combustion products out the back to provide thrust. Most familiar to commercial airline passengers are gas turbine engines, which contain five basic components: an air inlet; a compressor (a wheel of airfoils on a rotating shaft) that sucks in air and raises its pressure; a combustion chamber (or combustor) into which fuel is injected and burned; a turbine wheel that spins when hot combustion gases rush past its airfoils to drive the compressor wheel shaft; and a nozzle through which the high-temperature exhaust expands to generate thrust. Current turbojet engines can power aircraft to velocities up to a bit more than Mach 3 [*see box on opposite page*]. At higher speeds, the rotating components suffer damage from overheating.

At speeds above approximately Mach 2.5, a jet engine needs no compressor or turbine if it is designed so the oncoming air is "ram-compressed." Thus, a ramjet engine has just an air

Scramjets may someday power A TWO-HOUR PASSENGER AIRLINER FLIGHT from New York to Sydney.

Overview/*Hypersonic Engines*

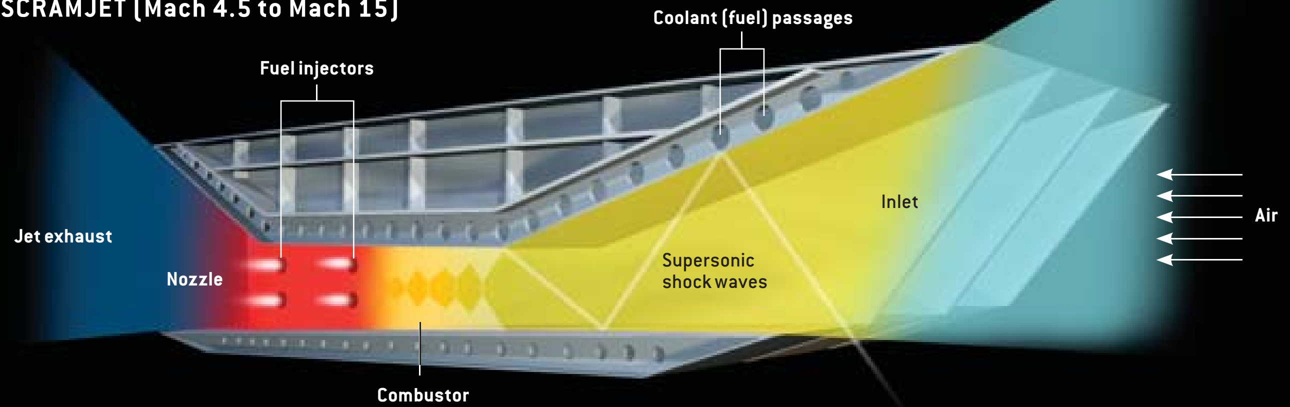
- Supersonic combustion ramjets, or scramjets, could power missiles and other weapons, space planes and even long-range passenger airliners at hypersonic speeds—Mach 5 to Mach 15. (Mach 1 is the speed of sound, or 760 miles per hour at sea level.)
- Scramjets ingest supersonic air, mix it with fuel, and burn it to create tremendous propulsive thrust. Unlike rockets, they do not need to carry oxygen and oxidizer, saving weight and providing as much as four times the thrust per given weight unit of propellant.
- Although a scramjet is simple in concept—it requires no spinning turbine wheels—the technical challenges to building one that can operate in different flight regimes for extended periods present many hurdles.

DON FOLEY; COURTESY OF EARTH SCIENCES AND IMAGE ANALYSIS LABORATORY, NASA JOHNSON SPACE CENTER (*Earth*) (*preceding pages*)

A FAMILY OF FLIGHT ENGINES

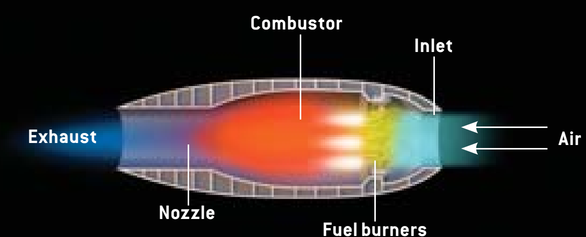
Scramjets are part of a family of jet engines that operate on similar principles. In general, each generates propulsive thrust by compressing incoming air, blending it with fuel, burning the mixture and expanding the combustion products out the rear end.

SCRAMJET (Mach 4.5 to Mach 15)



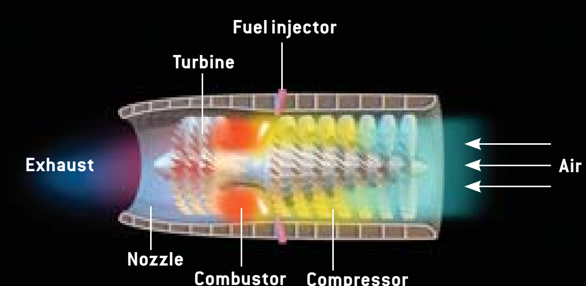
Supersonic-speed air enters the inlet where the constricted flow path "rams" the air—decelerating and compressing it, thus converting some of its kinetic energy to heat. Injectors pump fuel into the air in the combustor, where the still supersonic mixture burns rapidly, transforming the fuel's chemical energy into thermal energy. The squeezed interior pathway confines the swelling, high-temperature mixture, further raising its pressure. When the exhaust gases reach the nozzle, where the pathway widens, the mass expands and accelerates outward, changing its thermal energy into kinetic energy for thrust.

RAMJET (Mach 2.5 to Mach 5 or 6)



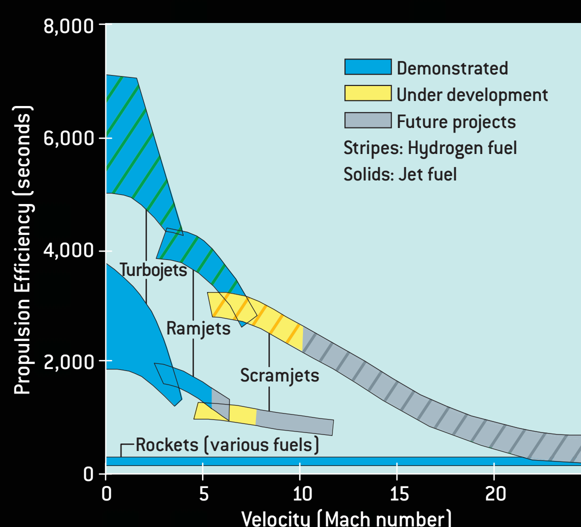
Ramjet operation resembles that of the scramjet except that its internal airflow remains at subsonic velocities.

TURBOJET (Mach 0 to Mach 3)



Because it travels more slowly, the turbojet needs rotating turbine wheels to compress incoming air and produce thrust.

PERFORMANCE WINDOWS



Each flight engine design is best suited to a certain set of vehicle speed and altitude conditions. Hydrogen fuel offers better engine performance but presents problems with packaging in a small space and with the existing fuel distribution infrastructure. Hydrocarbon jet fuels are easier to handle but provide less energy per weight unit. Propulsion efficiency, a measure of relative engine effectiveness, is the jet thrust per weight unit flow rate of propellant.

DON FOLEY (*illustrations*); LUCY READING-IKKANDA (*graph*)

Selected Scramjet R&D Programs

In addition to the U.S. Air Force's HyTech and X-51A Scramjet Engine Demonstrator programs [*see main article*], other national and international research efforts are developing scramjet engine technology.

PROGRAM	DATES	INSTITUTION	ACHIEVEMENTS OR GOALS
Hyper-X	1996–2004	NASA	The Hyper-X (X-43A) effort flew small test vehicles to demonstrate hydrogen-fueled scramjet engines. In one flight test the X-43A reached nearly Mach 10
HyShot	2001–present	University of Queensland, Australia (supported by partners in Australia, the U.K., the U.S., Germany, South Korea and Japan)	In July 2002 the HyShot team conducted the first-ever successful test flight of a scramjet. It flew downward at about Mach 7.6 for six seconds
Hypersonic Flight Demonstration (HyFly)	2002–present	DARPA and the Office of Naval Research (ONR)	The HyFly program is building a Mach 6-plus ramjet/scramjet-powered cruise missile. The Johns Hopkins University Applied Physics Laboratory developed the engine for the rocket-boosted craft
Freeflight Atmospheric Scramjet Test Technique (FASTT, under HyFly)	2003–present	Alliant Techsystems (with DARPA and ONR support)	The rocket-boosted, kerosene-fueled scramjet vehicle reached Mach 5.5 during a 15-second flight on December 10, 2005
Falcon	2003–present	DARPA	The Falcon program is to construct an unmanned hypersonic military aircraft capable of reaching any point on the globe in about two hours. The technology may eventually have nonmilitary applications, potentially aiding in the development of a single-stage orbital space plane

inlet, a combustor and a nozzle [*see box on preceding page*]. The ramjet's specially designed inlet simultaneously pressurizes the air and slows it to subsonic velocities. Injectors add fuel to the airflow, whereupon the air-fuel mixture ignites and burns. The hot exhaust gases again accelerate to around the speed of sound as they pass through a narrow throat, or mechanical choke, and then expand out the conelike nozzle to supersonic velocities. As the aircraft's Mach number rises above 5, deceleration of the inlet air raises the temperature within the engine to a point at which it is difficult to add more heat efficiently via combustion. Thus, Mach 5 to Mach 6 is the practical limit of the ramjet.

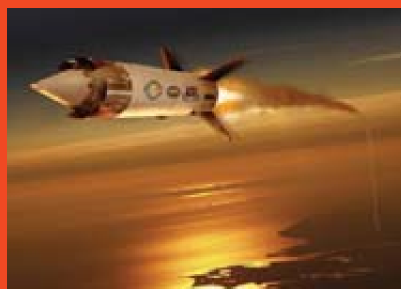
Anatomy of a Scramjet

TO GENERATE more propulsive thrust and operate at a higher flight speed than the ramjet, the scramjet reduces the initial compression of the airflow so that it does not slow nearly as much—ideally, it remains supersonic throughout the combustion process. Like a ramjet, the scramjet contains no moving parts in the airflow path: it essentially consists of a constricted tube shaped like two funnels attached at their narrow ends [*see box on preceding page*]. In operation, supersonic air entering the inlet (the first funnel) becomes pressurized and hot. In the area of restricted flow in the central passage (the combustor), fuel is injected into the passing air and ignited, which heats the gas further. The resulting exhaust gases rocket out of the nozzle (the second funnel) at a higher speed than that of the inlet air.

Somewhat like certain sharks that swim forward ceaselessly to maintain their oxygen supply, a ramjet or scramjet engine must travel forward rapidly to force air into its intake before it can start up and create thrust. The need to get a running start means that a scramjet-powered orbital launch vehicle would have to integrate another propulsion system, such as a rocket or a gas turbine engine, to get it moving. Once it reaches the necessary speed, a space plane pilot would engage the scramjet for the flight to the upper atmosphere, where a rocket would take over for final insertion into orbit. Designing a propulsion system that integrates the different engine cycles is an optimization problem that is influenced by factors such as payload size, intended orbit, required range and speed of atmospheric transport, and weapons-carrying capability.

The chief difficulty in scramjet operation is the short residence time of air in the engine—a few thousandths of a second—so the task of burning the fuel resembles lighting a match inside a tornado and somehow keeping it alight. The trick to getting a scramjet to work lies in the extremely sophisticated shaping of the tube's inner geometry and in where along that tube heat is released from combustion. A practical scramjet generates stable thrust by precisely controlling the speed and pressure of air flowing through the engine and by metering the fuel into the combustor so it burns fully and releases its energy exactly as needed. Careful control of the relation between the flow area and the heat release negates the need for the mechanical choke of the ramjet and enables the scramjet to maintain supersonic flow through the combustor.

Scramjet researchers know that it is crucial to manage closely the thermal energy in the engine. Heat flows to the structure from friction and from the combustion process; locally, this heat flux can be greatly amplified by internal shock waves impinging on the engine wall. The kinetic energy of the ingested hypersonic airflow, if fully converted to thermal en-



SCRAMJETS come in various shapes, but all require rockets to boost them up to speed so they can start up. NASA's aircraftlike X-43A research vehicle (*above*), which set the speed record in November 2004 for an air-breathing-engine jet—Mach 9.6, or nearly 7,000 miles per hour—was launched from an Orbital Sciences Pegasus rocket. HyShot scramjets blasted off onboard Terrier-Orion rockets (*far right*), as did the similarly configured FASTT scramjet, shown in a future military missile application (*center*).

NASA/DRYDEN FLIGHT RESEARCH CENTER (*X-43A*); ATK (*FASTT*); CHRIS STACEY *University of Queensland* (*HyShot*)

ergy, is more than enough to melt the metal structure of the engine. Yet without sufficient deceleration, the air travels through the engine too fast and at too low a temperature and pressure to support fuel combustion.

To prevent the engine structure from melting because of air friction from the supersonic flow, engineers use "active cooling" schemes. Pumps force a constant flow of heat-absorbing fuel through passages built into the engine and airframe components to siphon off potentially destructive heat. The process has the ancillary benefit of preparing the fuel for rapid combustion in the engine. This cooling technique has been applied successfully to conventional rockets for decades, typically employing liquid hydrogen as the coolant. Using a hydrocarbon fuel in such an environment is more challenging because a thermally stressed hydrocarbon can readily decompose to solid coke, which can plug coolant passages. Other drawbacks are that active-cooling systems involve additional weight and complexity and that they must stay active, because loss of the fuel coolant will lead to catastrophic structural failure.

Successful scramjet operation is thus a delicate balancing act, further complicated by the fact that a given airflow geometry is optimized for only a single set of flight conditions (speed, altitude, and so forth). Ideally, the physical dimensions and shape of a scramjet's flow path would adapt constantly as the vehicle accelerates and changes height, but mobile, heat-resistant inner surfaces and mechanical joints with such capabilities might be beyond today's materials and structures. The need to continuously move very hot interior surfaces of the engine and to seal passages against leakage of high-temperature engine gases remains a barrier to realizing the full potential of the scramjet cycle.

A Case Study

DESPITE THE TECHNICAL HURDLES inherent in scramjets, researchers have attained recent successes that hold promise for the future. One of these achievements is the work of the U.S. Air Force's HyTech program, which began in 1995. HyTech's collaboration of government, industry, and university scientists and engineers concentrated on what the team thought would be a manageable piece of the scramjet engineering challenge. First, the members focused on small, expendable scramjet engines, such as those for a missile. This engine would be small enough to fit in existing ground-test facilities, easing technical evaluation. It would need to operate only once, deferring the complication of developing reusable flight structures to later research. By limiting the operational range from Mach 4 to Mach 8 and by specifying a fixed-geometry flow path, the program minimized design complexity.

Last, they chose to operate HyTech with JP-7 jet fuel, a liquid hydrocarbon originally developed for the SR-71 Black-

THE AUTHOR

THOMAS A. JACKSON is deputy for science at the Aerospace Propulsion Division of the U.S. Air Force Research Laboratory's Propulsion Directorate in Ohio, where he sets the direction for scientific investigations into advanced air-breathing propulsion engine technology. Jackson received a Ph.D. in mechanical engineering from the University of California, Irvine, in 1985 and holds an M.A. in the management of technology from the Massachusetts Institute of Technology's Sloan School of Management. His research has focused mainly on combustion and fuel injection technologies in propulsion engines. A father of six, Jackson enjoys playing tennis, renovating old houses and repairing broken toys.

bird program. As mentioned earlier, in a fuel-cooled scramjet the fuel serves as the heat sink—the means by which excess heat is managed. In a thermally balanced system the amount of fuel required to absorb excess structural heat should not exceed the quantity of fuel required for combustion. HyTech planners want this balance to occur at Mach 8, and JP-7 has proved well suited to the task.

For an air-breathing power plant to compete favorably in effectiveness with a rocket-powered launch vehicle, performance studies indicate that it must be able to operate well at about half its maximum speed. Thus, engineers aimed for a Mach 4 scramjet start-up speed, which is difficult to achieve because the temperature of air entering the combustor at that speed is far below that at which the fuel autoignites. Hence, the engine would require an ignition assist, such as a chemical additive to lower the fuel's autoignition temperature or a device that would set fire to the fuel by generating a very hot gas and injecting it into the air-fuel mixture. At higher flight Mach numbers, igniting and stabilizing the flame is much easier, until very high flight speeds at which the short residence time inside the engine makes maintaining combustion a challenge.

By 2003 the HyTech team developed engine components and integrated engine subsystems that met or exceeded most of the original program goals. But even after extensive ground-based engine testing, significant development uncertainties existed. These remaining question marks, which are all associated with maintaining performance in transient conditions—as velocity, altitude and throttle settings change—are extremely difficult to investigate in wind tunnels and so would be best addressed in test flights.

That is why the U.S. Air Force Scramjet Engine Demonstrator (SED), now designated the X-51A, will take some of these HyTech systems into the air [*see illustration below*] in 2009. The program is a follow-on flight evaluation for the technology still being refined in the HyTech program. Through ground testing and extensive computational analysis, HyTech engineers have developed a flight-weight, actively cooled scramjet that is suitable for testing in the SED program.

Faced with the current inability to alter interior engine shape in flight to adjust performance for fast-changing speeds and altitudes, the SED engineering team chose to build a fixed-geometry flow path that is a compromise between adequate acceleration through the lower end of its velocity range (Mach 4.5 to Mach 7) and efficient cruise performance at the highest speed of Mach 7. Managing the distribution of fuel within the engine is the primary means of controlling the engine—its thrust, rate of acceleration and maintenance of stable operation.

The engine is constructed primarily of steel that is actively cooled with internal fuel flow. In addition, heat-resistant ceramic components substitute for steel in certain leading edges—the areas in the front that take the direct brunt of the hot airflow—that are too sharp to contain coolant passages. Reliably joining these cooled and uncooled parts is challenging but critical. Clearly, overly rapid structural breakdown (before the missile reaches its target) would lead to catastrophic failure. Another problem is that mismatching thermal expansion of the ceramic and metal components would distort airflow geometries and confound attempts to control scramjet performance. Engineers developed a heat-resistant carbon-carbon composite material with tongue-and-groove joints that overcame the problem.

The use of JP-7 to operate and cool the scramjet is essential to the X-51A's success. Up to now, hydrogen has been the fuel of choice for most scramjet programs. In contrast to hydrogen, most hydrocarbon fuels are less reactive, contain less energy per unit weight and have lower heat capacity to cool hot structures. Hydrocarbon fuels are, however, commonly used in all other Air Force flight applications, so a global infrastructure for distribution and handling already exists. In addition, hydrocarbons package better, exhibiting a greater energy content per unit volume, so they require less room onboard than

DON FOLEY (*illustration*); PAUL BAGBY NASA (*photograph*)

X-51A SCRAMJET ENGINE DEMONSTRATOR (*above*) will flight-test the U.S. Air Force's HyTech scramjet and airframe technology in 2009. The U.S. Air Force's latest scramjet prototype (*right*) recently completed a series of ground tests at the NASA Langley Research Center's Eight-Foot High Temperature Tunnel. Engineers tested the flight-weight engine upside down to route plumbing and instrumentation lines out through the pedestal below.

a quantity of hydrogen with the equivalent energy content.

To offset the hydrocarbon fuel's lower reactivity and heat capacity disadvantages, HyTech exploits JP-7's endothermic potential—that is, its ability to soak up heat chemically. When such fuels take in heat from their surroundings in the absence of oxygen and in the presence of an appropriate catalyst, their long, complex polymeric chains decompose into short, simple ones. In the process, the fuel absorbs up to five times its latent heat capacity—the energy absorbed in simple heating of the liquid. What is more, after endothermic heating the fuel becomes a hot gas that contains as much as 10 percent more energy than the chemical energy of the unheated liquid fuel. Finally, the resulting low-molecular-weight hydrocarbons are more reactive than their parent fuel molecules, making them easier to burn in the brief time the fuel is inside the scramjet.

Engineers have already produced a fixed-geometry engine of sufficient size to power a missilelike vehicle that can start burning hydrocarbon fuel (such as JP-7) at Mach 4.5 and then accelerate to Mach 7. Also in hand are active cooling and temperature-resistant structural technologies that enable the engine to maintain thermal balance as long as fuel remains onboard. In 2009 the X-51A free-flight vehicle will be boosted to scramjet light-off speed by a rocket and then released in hopes of confirming these critical technologies in flight.

Future Challenges

ASSUMING THAT the SED flight tests are successful, much will remain to be done before applications such as rapid-response weapons delivery, sustained hypersonic cruise and more affordable access to space can be realized.

Scramjets must be able to function reliably over a wide range of Mach numbers. As I mentioned, current gas turbines are effective from Mach 0 to Mach 3 or 4, whereas the rocket is essential for segments of the flight regime above approximately Mach 15. At those higher velocities, thermal heating becomes intolerable for scramjets at the altitudes required to ingest sufficient air for fuel combustion. Researchers therefore need to develop scramjets that can fill as much of the intervening niche from Mach 4 to Mach 15 as possible. In some applications the scramjet will have to be fully integrated with a lower-speed cycle such as the gas turbine. That means the engines' operating regimes will need to overlap to permit a smooth transition. Engineers will also have to endeavor to keep the extra mass of the various propulsive systems from overloading any multiengine craft while finely controlling the handoffs between them.

The X-51A, with its fixed-geometry design, cannot substantially lower its operating limit. Variable internal engine geometry will be required to allow a scramjet to run at Mach numbers lower than 4. Though not part of the SED effort, U.S. Air Force and NASA engineers have demonstrated in a HyTech engine a prototype variable-geometry inlet in which movable flaps change its aerodynamic configuration.

Fuel technology may also constrain the scramjet's utility at both ends of the current operating envelope. The X-51A is designed to run only after sufficient structural heating occurs to convert the JP-7 fuel into a gaseous state. For lower Mach speeds, next-generation scramjet combustors may have to operate briefly on liquid fuel or a propellant that contains both liquid and gas phases, before changing over to a fully gaseous fuel later in the flight. Liquids are 1,000 times as dense as gases, which makes fueling a scramjet and maintaining stable combustion and thrust during a transition from liquid propellant to gaseous fuel extremely tricky. This capability has, however, been demonstrated in component tests conducted within the HyTech research effort. At the high-speed end of the envelope, the heat capacity of aviation jet fuels, even those that can decompose endothermically like JP-7, will be found lacking as speeds approach Mach 8. Flying faster will require dramatically different fuels and advanced heat-resistant materials or perhaps the use of hydrogen, despite its attendant logistics and vehicle-packaging challenges.

Burning fuel in a scramjet resembles LIGHTING A MATCH AND KEEPING IT ALIGHT inside a tornado.

The HyTech program's initial focus has been on aircraft-launched, missile-size vehicles. For applications such as sustained hypersonic cruise and space access, much larger vehicles will be required. DARPA's Falcon program and the Air Force's Robust Scramjet program, both having begun in 2003, are wrestling with the issues of greater engine size and airflow capacities 100 times that of current HyTech devices.

Recent scramjet technology development efforts have made great strides in overcoming critical barriers to achieving sustained, high-speed flight. With continued progress, we hope to continue to inch, if not stride, toward something resembling a *Star Wars* X-wing engine in the not too distant future. SA

MORE TO EXPLORE

Ramjets. Edited by Gordon L. Dugger. American Institute of Aeronautics and Astronautics Selected Reprint Series, 1969.

A Procedure for Optimizing the Design of Scramjet Engines. P. J. Waltrup, F. S. Billig and R. D. Stockbridge in *Journal of Spacecraft and Rockets,* Vol. 16, No. 3, pages 163–171; May–June 1979.

Research on Supersonic Combustion. F. S. Billig in *Journal of Propulsion and Power,* Vol. 9, No. 4, pages 499–514; July–August 1993.

Hypersonic Airbreathing Propulsion. William H. Heiser, David T. Pratt, Daniel H. Daley and Unmeel B. Mehta. American Institute of Aeronautics and Astronautics Education Series, 1994.

Investigation of Scramjet Injection Strategies for High Mach Number Flows. D. W. Riggins, C. R. McClinton, R. C. Rogers and R. D. Bittner in *Journal of Propulsion and Power,* Vol. 11, No. 3, pages 409–418; May–June 1995.

EFFORTFUL STUDY is the key to achieving success in chess, classical music, soccer and many other fields. New research has indicated that motivation is a more important factor than innate ability.

THE EXPERT MIND

Studies of the mental processes of chess grandmasters have revealed clues to how people become experts in other fields as well

By Philip E. Ross

A man walks along the inside of a circle of chess tables, glancing at each for two or three seconds before making his move. On the outer rim, dozens of amateurs sit pondering their replies until he completes the circuit. The year is 1909, the man is José Raúl Capablanca of Cuba, and the result is a whitewash: 28 wins in as many games. The exhibition was part of a tour in which Capablanca won 168 games in a row.

How did he play so well, so quickly? And how far ahead could he calculate under such constraints? "I see only one move ahead," Capablanca is said to have answered, "but it is always the correct one."

He thus put in a nutshell what a century of psychological research has subsequently established: much of the chess master's advantage over the novice derives from the first few seconds of thought. This rapid, knowledge-guided perception, sometimes called apperception, can be seen in experts in other fields as well. Just as a master can recall all the moves in a game he has played, so can an accomplished musician often reconstruct the score to a sonata heard just once. And just as the chess master often finds the best move in a flash, an expert physician can sometimes make an accurate diagnosis within moments of laying eyes on a patient.

But how do the experts in these various subjects acquire their extraordinary skills? How much can be credited to innate talent and how much to intensive training? Psychologists have sought answers in studies of chess masters. The collected results of a century of such research have led to new theories explaining how the mind organizes and retrieves information. What is more, this research may have important implications for educators. Perhaps the same techniques used by chess players to hone their skills could be applied in the classroom to teach reading, writing and arithmetic.

The *Drosophila* of Cognitive Science

THE HISTORY of human expertise begins with hunting, a skill that was crucial to the survival of our early ancestors. The mature hunter knows not only where the lion has been; he can also infer where it will go. Tracking skill increases, as repeated studies show, from childhood onward, rising in "a linear relationship, all the way out to the mid-30s, when it tops out," says

ETHAN HILL

John Bock, an anthropologist at California State University, Fullerton. It takes less time to train a brain surgeon.

Without a demonstrably immense superiority in skill over the novice, there can be no true experts, only laypeople with imposing credentials. Such, alas, are all too common. Rigorous studies in the past two decades have shown that professional stock pickers invest no more successfully than amateurs, that noted connoisseurs distinguish wines hardly better than yokels, and that highly credentialed psychiatric therapists help patients no more than colleagues with less advanced degrees. And even when expertise undoubtedly exists—as in, say, teaching or business management—it is often hard to measure, let alone explain.

Skill at chess, however, can be measured, broken into components, subjected to laboratory experiments and readily observed in its natural environment, the tournament hall. It is for those reasons that chess has served as the greatest single test bed for theories of thinking—the "*Drosophila* of cognitive science," as it has been called.

The measurement of chess skill has been taken further than similar attempts with any other game, sport or competitive activity. Statistical formulas weigh a player's recent results over older ones and discount successes according to the strength of one's opponents. The results are ratings that predict the outcomes of games with remarkable reliability. If player A outrates player B by 200 points, then A will on average beat B 75 percent of the time. This prediction holds true whether the players are top-ranked or merely ordinary. Garry Kasparov, the Russian grandmaster who has a rating of 2812, will win 75 percent of his games against the 100th-ranked grandmaster, Jan Timman of the Netherlands, who has a rating of 2616. Similarly, a U.S. tournament player rated 1200 (about the median) will win 75 percent of the time against someone rated 1000 (about the 40th percentile). Ratings allow psychologists to assess expertise by performance rather than reputation and to track changes in a given player's skill over the course of his or her career.

Another reason why cognitive scientists chose chess as their model—and not billiards, say, or bridge—is the game's reputation as, in German poet Johann Wolfgang von Goethe's words, "the touchstone of the intellect." The feats of chess masters have long been ascribed to nearly magical mental powers. This magic shines brightest in the so-called blindfold games in which the players are not allowed to see the board. In 1894 French psychologist Alfred Binet, the co-inventor of the first intelligence test, asked chess masters to describe how they played such games. He began with the hypothesis that they achieved an almost photographic image of the board, but he soon concluded that the visualization was much more abstract. Rather than seeing the knight's mane or the grain of the wood from which it is made, the master calls up only a general knowledge of where the piece stands in relation to other elements of the position. It is the same kind of implicit knowledge that the commuter has of the stops on a subway line.

The blindfolded master supplements such knowledge with details of the game at hand as well as with recollections of salient aspects of past games. Let us say he has somehow forgotten the precise position of a pawn. He can find it, as it were, by considering the stereotyped strategy of the opening—a well-studied phase of the game with a relatively limited number of options. Or he can remember the logic behind one of his earlier moves—say, by reasoning: "I could not capture his bishop two moves ago; therefore, that pawn must have been standing in the way. . . ." He does not have to remember every detail at all times, because he can reconstruct any particular detail whenever he wishes by tapping a well-organized system of connections.

Of course, if the possession of such intricately structured knowledge explains not only success at blindfold play but also other abilities of chess masters, such as calculation and planning, then expertise in the game would depend not so much on innate abilities as on specialized training. Dutch psychologist Adriaan de Groot, himself a chess master, confirmed this notion in 1938, when he took advantage of the staging of a great international tournament in Holland to compare average and strong players with the world's leading grandmasters. One way he did so was to ask the players to describe their

Much of the chess master's advantage over the novice derives from the FIRST FEW SECONDS OF THOUGHT.

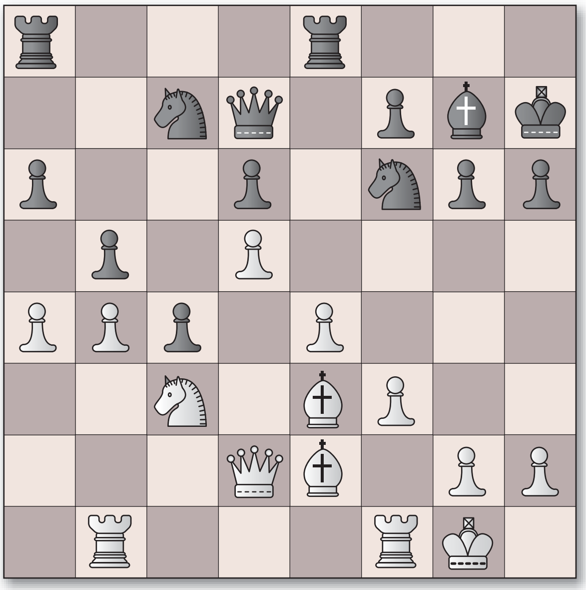
Overview/*Lessons from Chess*

- Because skill at chess can be easily measured and subjected to laboratory experiments, the game has become an important test bed for theories in cognitive science.
- Researchers have found evidence that chess grandmasters rely on a vast store of knowledge of game positions. Some scientists have theorized that grandmasters organize the information in chunks, which can be quickly retrieved from long-term memory and manipulated in working memory.
- To accumulate this body of structured knowledge, grandmasters typically engage in years of effortful study, continually tackling challenges that lie just beyond their competence. The top performers in music, mathematics and sports appear to gain their expertise in the same way, motivated by competition and the joy of victory.

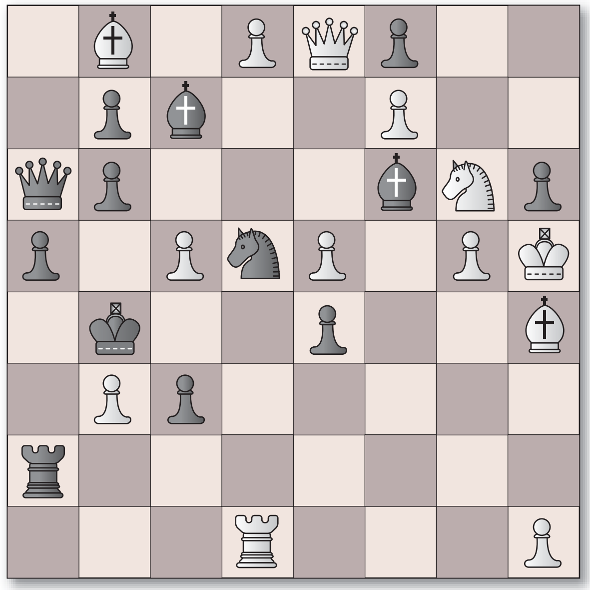
A GRANDMASTER'S MEMORY

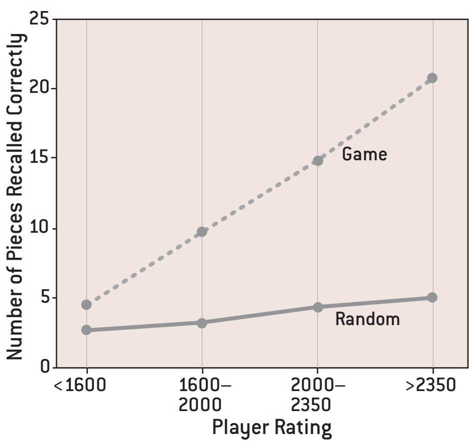
Experiments indicate that the memory of chess masters is tuned to typical game positions. In 13 studies conducted between 1973 and 1996 (the results were compiled in a review article published in 1996), players at various skill levels were shown positions from actual games (*a*) and positions obtained by randomly shuffling the pieces (*b*). After observing the positions for 10 seconds or less, the players were asked to reconstruct them from memory. The results (*graph at bottom*) showed that chess masters (with ratings of 2200 or higher) and grandmasters (generally 2500 or higher) were significantly better than weaker players at recalling the game positions but only marginally better at remembering the random positions. This finely tuned long-term memory appears to be crucial to chess expertise.

a

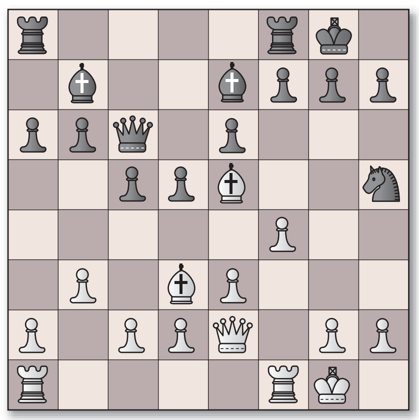


b





A structured knowledge of chess positions enables a grandmaster to spot the correct move quickly. The position at the right comes from a famous 1889 game between Emanuel Lasker (*white*) and Johann Bauer (*black*). Although a novice player would have to analyze the position extensively to see the winning move for white, any grandmaster would immediately recognize it. The correct move is shown on page 71.

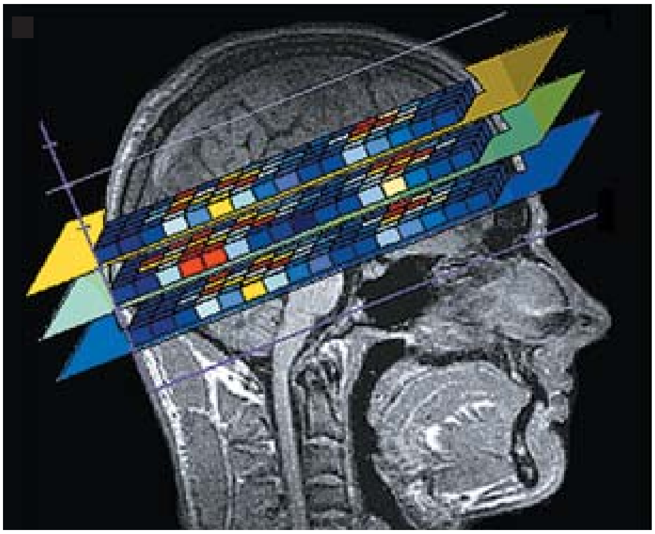


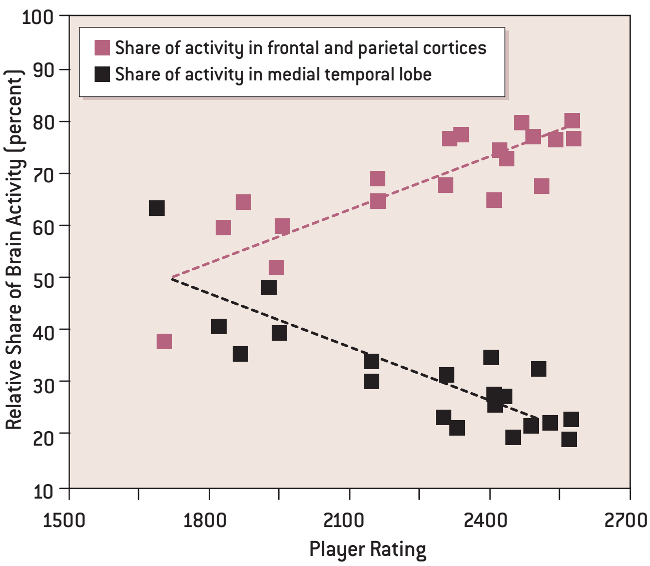
LUCY READING-IKKANDA; SOURCE: "CHUNKING MECHANISMS IN HUMAN LEARNING," BY FERNAND GOBET, PETER C. R. LANE, STEVE CROKER, PETER C.-H. CHENG, GARY JONES, IAIN OLIVER AND JULIAN M. PINE, IN *TRENDS IN COGNITIVE SCIENCES*, VOL. 5, NO. 6; JUNE 2001

thoughts as they examined a position taken from a tournament game. He found that although experts—the class just below master—did analyze considerably more possibilities than the very weak players, there was little further increase in analysis as playing strength rose to the master and grandmaster levels. The better players did not examine more possibilities, only better ones—just as Capablanca had claimed.

Recent research has shown that de Groot's findings reflected in part the nature of his chosen test positions. A position in which extensive, accurate calculation is critical will allow the grandmasters to show their stuff, as it were, and they will then search more deeply along the branching tree of possible moves than the amateur can hope to do. So, too, experienced physicists may on occasion examine more possibilities than physics students do. Yet in both cases, the expert relies not so much on an intrinsically stronger power of analysis as on a store of structured knowledge. When confronted with a difficult position, a weaker player may calculate for half an hour, often looking many moves ahead, yet miss the right continuation, whereas a grandmaster sees the move immediately, without consciously analyzing anything at all.

De Groot also had his subjects examine a position for a





BRAIN ACTIVITY in chess masters is different from the pattern observed in novices. In a 2001 report researchers used magnetoencephalography—the measurement of magnetic fields produced by electric currents in the brain—on subjects playing chess against a computer. In weaker players (*image at top*) more activity occurred in the brain's medial temporal lobe (*left side of colored slices*) than in the frontal and parietal cortices (*right side*), which suggests that the amateurs were analyzing unusual new moves. In grandmasters, however, the frontal and parietal cortices were more active, indicating that they were retrieving information from long-term memory (*data points at right in graph*).

limited period and then try to reconstruct it from memory. Performance at this task tracked game-playing strength all the way from novice to grandmaster. Beginners could not recall more than a very few details of the position, even after having examined it for 30 seconds, whereas grandmasters could usually get it perfectly, even if they had perused it for only a few seconds. This difference tracks a particular form of memory, specific to the kind of chess positions that commonly occur in play. The specific memory must be the result of training, because grandmasters do no better than others in general tests of memory.

Similar results have been demonstrated in bridge players (who can remember cards played in many games), computer programmers (who can reconstruct masses of computer code) and musicians (who can recall long snatches of music). Indeed, such a memory for the subject matter of a particular field is a standard test for the existence of expertise.

The conclusion that experts rely more on structured knowledge than on analysis is supported by a rare case study of an initially weak chess player, identified only by the initials D.H., who over the course of nine years rose to become one of Canada's leading masters by 1987. Neil Charness, professor of psychology at Florida State University, showed that despite the increase in the player's strength, he analyzed chess positions no more extensively than he had earlier, relying instead on a vastly improved knowledge of chess positions and associated strategies.

Chunking Theory

IN THE 1960S Herbert A. Simon and William Chase, both at Carnegie Mellon University, tried to get a better understanding of expert memory by studying its limitations. Picking up where de Groot left off, they asked players of various strengths to reconstruct chess positions that had been artificially devised—that is, with the pieces placed randomly on the board—rather than reached as the result of master play [*see box on preceding page*]. The correlation between game-playing strength and the accuracy of the players' recall was much weaker with the random positions than with the authentic ones.

Chess memory was thus shown to be even more specific than it had seemed, being tuned not merely to the game itself but to typical chess positions. These experiments corroborated earlier studies that had demonstrated convincingly that ability in one area tends not to transfer to another. American psychologist Edward Thorndike first noted this lack of transference over a century ago, when he showed that the study of Latin, for instance, did not improve command of English and that geometric proofs do not teach the use of logic in daily life.

Simon explained the masters' relative weakness in reconstructing artificial chess positions with a model based on meaningful patterns called chunks. He invoked the concept to explain how chess masters can manipulate vast amounts of stored information, a task that would seem to strain the working memory. Psychologist George Miller of Princeton University famously estimated the limits of working memory—the scratch pad of the mind—in a 1956 paper entitled "The Magical Number Seven, Plus or Minus Two." Miller showed that people can contemplate only five to nine items at a time. By packing hierarchies of information into chunks, Simon argued, chess masters could get around this limitation, because by using this method, they could access five to nine chunks rather than the same number of smaller details.

Take the sentence "Mary had a little lamb." The number of information chunks in this sentence depends on one's knowledge of the poem and the English language. For most native speakers of English, the sentence is part of a much larg-

OGNJEN AMIDZIC ET AL. IN *NATURE*, VOL. 412; AUGUST 9, 2001 (*top*); LUCY READING-IKKANDA (*bottom*) er chunk, the familiar poem. For someone who knows English but not the poem, the sentence is a single, self-contained chunk. For someone who has memorized the words but not their meaning, the sentence is five chunks, and it is 18 chunks for someone who knows the letters but not the words.

In the context of chess, the same differences can be seen between novices and grandmasters. To a beginner, a position with 20 chessmen on the board may contain far more than 20 chunks of information, because the pieces can be placed in so many configurations. A grandmaster, however, may see one part of the position as "fianchettoed bishop in the castled kingside," together with a "blockaded king's-Indian-style pawn chain," and thereby cram the entire position into perhaps five or six chunks. By measuring the time it takes to commit a new chunk to memory and the number of hours a player must study chess before reaching grandmaster strength, Simon estimated that a typical grandmaster has access to roughly 50,000 to 100,000 chunks of chess information. A grandmaster can retrieve any of these chunks from memory simply by looking at a chess position, in the same way that most native English speakers can recite the poem "Mary had a little lamb" after hearing just the first few words.

Even so, there were difficulties with chunking theory. It could not fully explain some aspects of memory, such as the ability of experts to perform their feats while being distracted (a favorite tactic in the study of memory). K. Anders Ericsson of Florida State University and Charness argued that there must be some other mechanism that enables experts to employ long-term memory as if it, too, were a scratch pad. Says Ericsson: "The mere demonstration that highly skilled players can play at almost their normal strength under blindfold conditions is almost impossible for chunking theory to explain because you have to know the position, then you have to explore it in your memory." Such manipulation involves changing the stored chunks, at least in some ways, a task that may be likened to reciting "Mary had a little lamb" backward. It can be done, but not easily, and certainly not without many false starts and errors. Yet grandmaster games played quickly and under blindfold conditions tend to be of surprisingly high quality.

Ericsson also cites studies of physicians who clearly put information into long-term memory and take it out again in ways that enable them to make diagnoses. Perhaps Ericsson's most homely example, though, comes from reading. In a 1995 study he and Walter Kintsch of the University of Colorado found that interrupting highly proficient readers hardly slowed their reentry to a text; in the end, they lost only a few seconds. The researchers explained these findings by recourse to a structure they called long-term working memory, an almost oxymoronic coinage because it assigns to long-term memory the one thing that had always been defined as incompatible with it: thinking. But brain-imaging studies done in 2001 at the University of Konstanz in Germany provide support for the theory by showing that expert chess players activate long-term memory much more than novices do [*see illustration on opposite page*].

Fernand Gobet of Brunel University in London champions a rival theory, devised with Simon in the late 1990s. It extends the idea of chunks by invoking highly characteristic and very large patterns consisting of perhaps a dozen chess pieces. Such a template, as they call it, would have a number of slots into which the master could plug such variables as a pawn or a bishop. A template might exist, say, for the concept of "the isolated queen's-pawn position from the Nimzo-Indian Defense," and a master might change a slot by reclassifying it as the same position "minus the dark-squared bishops." To resort again to the poetic analogy, it would be a bit like memorizing a riff on "Mary had a little lamb" by substituting rhyming equivalents at certain slots, such as "Larry" for "Mary," "pool" for "school" and so on. Anyone who knows the original template should be able to fix the altered one in memory in a trice.

The 10-year rule states that it takes approximately a decade of heavy labor to MASTER ANY FIELD.

A Proliferation of Prodigies

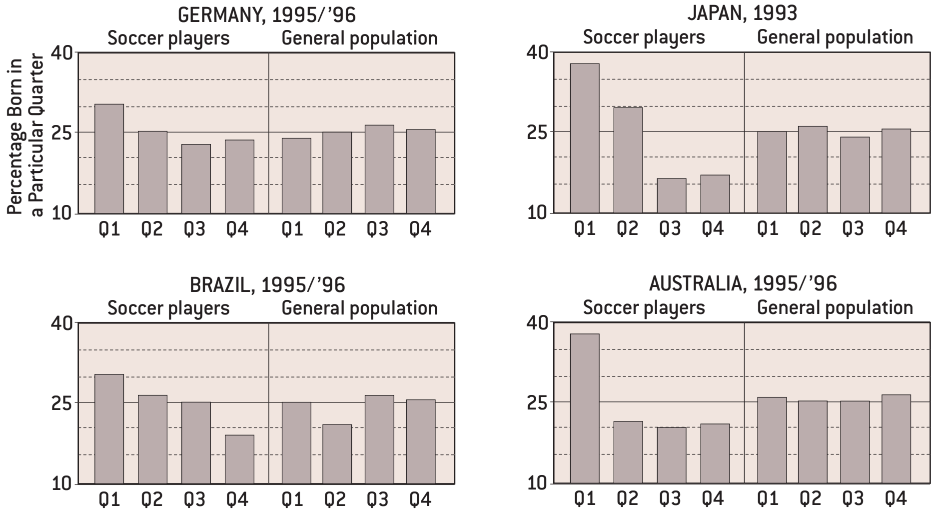
THE ONE THING that all expertise theorists agree on is that it takes enormous effort to build these structures in the mind. Simon coined a psychological law of his own, the 10-year rule, which states that it takes approximately a decade of heavy labor to master any field. Even child prodigies, such as Gauss in mathematics, Mozart in music and Bobby Fischer in chess, must have made an equivalent effort, perhaps by starting earlier and working harder than others.

According to this view, the proliferation of chess prodigies in recent years merely reflects the advent of computer-based training methods that let children study far more master games and to play far more frequently against master-strength programs than their forerunners could typically manage. Fischer made a sensation when he achieved the grandmaster title at age 15, in 1958; today's record-holder, Sergey Karjakin of Ukraine, earned it at 12 years, seven months.

Ericsson argues that what matters is not experience per se but "effortful study," which entails continually tackling challenges that lie just beyond one's competence. That is why it is possible for enthusiasts to spend tens of thousands of hours playing chess or golf or a musical instrument without ever advancing beyond the amateur level and why a properly trained student can overtake them in a relatively short time. It is interesting to note that time spent playing chess, even in

TRAINING TRUMPS TALENT

A 1999 study of professional soccer players suggests that they owe their success more to training than to talent. In Germany, Brazil, Japan and Australia, the players were much more likely than average to have been born in the first quarter (Q1) after the cutoff date for youth soccer leagues (*graphs at right*). Because these players were older than their teammates when they joined the leagues, they would have enjoyed advantages in size and strength, allowing them to handle the ball and score more often. Their success in early years would have motivated them to keep improving, thus explaining their disproportionate representation in the professional leagues. Intense motivation and training can also explain the feats of famous child prodigies such as Austrian composer Wolfgang Amadeus Mozart (*left*) and American golfer Tiger Woods (*right*).



NOTE: The cutoff dates were August 1 for Germany, Brazil and Australia, and April 1 for Japan.

LUCY READING-IKKANDA; SOURCE: "THE RELATIVE AGE EFFECT IN SOCCER," BY JOCHEN MUSCH AND ROY HAY, IN *SOCIOLOGY OF SPORT JOURNAL*, VOL. 16; 1999 (*graphs*); MOZART MUSEUM, SALZBURG, AUSTRIA/BRIDGEMAN ART LIBRARY (*Mozart*); ALAN LEVENSON *Corbis* (*Woods*)

tournaments, appears to contribute less than such study to a player's progress; the main training value of such games is to point up weaknesses for future study.

Even the novice engages in effortful study at first, which is why beginners so often improve rapidly in playing golf, say, or in driving a car. But having reached an acceptable performance—for instance, keeping up with one's golf buddies or passing a driver's exam—most people relax. Their performance then becomes automatic and therefore impervious to further improvement. In contrast, experts-in-training keep the lid of their mind's box open all the time, so that they can inspect, criticize and augment its contents and thereby approach the standard set by leaders in their fields.

Meanwhile the standards denoting expertise grow ever more challenging. High school runners manage the four-minute mile; conservatory students play pieces once attempted only by virtuosi. Yet it is chess, again, that offers the most convincing comparison over time. John Nunn, a British mathematician who is also a grandmaster, recently used a computer to help him compare the errors committed in all the games in two international tournaments, one held in 1911, the other in 1993. The modern players played far more accurately. Nunn then examined all the games of one player in 1911 who scored in the middle of the pack and concluded that his rating today would be no better than 2100, hundreds of points below the grandmaster level—"and that was on a good day and with a following wind." The very best old-time masters were considerably stronger but still well below the level of today's leaders.

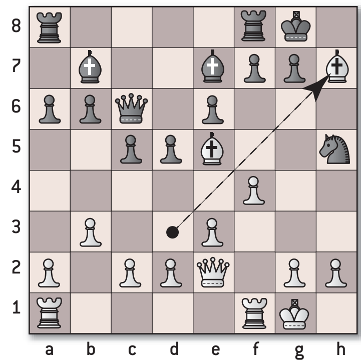
Then again, Capablanca and his contemporaries had neither computers nor game databases. They had to work things out for themselves, as did Bach, Mozart and Beethoven, and if they fall below today's masters in technique, they tower above them in creative power. The same comparison can be made between Newton and the typical newly minted Ph.D. in physics.

At this point, many skeptics will finally lose patience. Surely, they will say, it takes more to get to Carnegie Hall than practice, practice, practice. Yet this belief in the importance of innate talent, strongest perhaps among the experts themselves and their trainers, is strangely lacking in hard evidence to substantiate it. In 2002 Gobet conducted a study of British chess players ranging from amateurs to grandmasters and found no connection at all between their playing strengths and their visual-spatial abilities, as measured by shape-memory tests. Other researchers have found that the abilities of professional handicappers to predict the results of horse races did not correlate at all with their mathematical abilities.

Although nobody has yet been able to predict who will become a great expert in any field, a notable experiment has shown the possibility of deliberately creating one. László Polgár, an educator in Hungary, homeschooled his three daughters in chess, assigning as much as six hours of work a day, producing one international master and two grandmasters—the strongest chess-playing siblings in history. The youngest Polgár, 30-year-old Judit, is now ranked 14th in the world.

The Polgár experiment proved two things: that grandmasters can be reared and that women can be grandmasters. It is no coincidence that the incidence of chess prodigies multiplied after László Polgár published a book on chess education. The number of musical prodigies underwent a similar increase after Mozart's father did the equivalent two centuries earlier.

Thus, motivation appears to be a more important factor than innate ability in the development of expertise. It is no accident that in music, chess and sports—all domains in which expertise is defined by competitive performance rather than academic credentialing—professionalism has been emerging at ever younger ages, under the ministrations of increasingly dedicated parents and even extended families.



WHITE'S WINNING MOVE is bishop takes pawn on the h7 square. Black's king then captures the bishop, and the white queen captures the black knight at h5, with check, forcing the black king back to g8. White's other bishop then captures the pawn on g7, where it is taken by the black king. The double-bishop sacrifice paves the way for a queen-and-rook attack, forcing black to give up his queen to stave off mate. Emanuel Lasker, the game's winner, went on to become the world chess champion in 1894, a title he retained for 27 years before losing to José Raúl Capablanca.

The preponderance of psychological evidence indicates that experts are made, not born. What is more, the demonstrated ability to turn a child quickly into an expert—in chess, music and a host of other subjects—sets a clear challenge before the schools. Can educators find ways to encourage students to engage in the kind of effortful study that will improve their reading and math skills? Roland G. Fryer, Jr., an economist at Harvard University, has experimented with offering monetary rewards to motivate students in underperforming schools in New York City and Dallas. In one ongoing program in New York, for example, teachers test the students every three weeks and award small amounts—on the order of $10 or $20—to those who score well. The early results have been promising. Instead of perpetually pondering the question, "Why can't Johnny read?" perhaps educators should ask, "Why should there be anything in the world he can't learn to do?" SA

The preponderance of psychological evidence indicates that EXPERTS ARE MADE, NOT BORN.

Furthermore, success builds on success, because each accomplishment can strengthen a child's motivation. A 1999 study of professional soccer players from several countries showed that they were much more likely than the general population to have been born at a time of year that would have dictated their enrollment in youth soccer leagues at ages older than the average [*see box on opposite page*]. In their early years, these children would have enjoyed a substantial advantage in size and strength when playing soccer with their teammates. Because the larger, more agile children would get more opportunities to handle the ball, they would score more often, and their success at the game would motivate them to become even better.

Teachers in sports, music and other fields tend to believe that talent matters and that they know it when they see it. In fact, they appear to be confusing ability with precocity. There is usually no way to tell, from a recital alone, whether a young violinist's extraordinary performance stems from innate ability or from years of Suzuki-style training. Capablanca, regarded to this day as the greatest "natural" chess player, boasted that he never studied the game. In fact, he flunked out of Columbia University in part because he spent so much time playing chess. His famously quick apprehension was a product of all his training, not a substitute for it.

Philip E. Ross, a contributing editor at Scientific American, *is a chess player himself and father of Laura Ross, a master who outranks him by 199 points.*

MORE TO EXPLORE

The Rating of Chessplayers, Past and Present. Arpad E. Elo. Arco Publishing, 1978.

Thought and Choice in Chess. Adriaan de Groot. Mouton de Gruyter, 1978.

Expert Performance in Sports: Advances in Research on Sport Expertise. Edited by Janet L. Starkes and K. Anders Ericsson. Human Kinetics, 2003.

Moves in Mind: The Psychology of Board Games. Fernand Gobet, Alex de Voogt and Jean Retschitzki. Psychology Press, 2004.

The Cambridge Handbook of Expertise and Expert Performance. Edited by K. Anders Ericsson, Paul J. Feltovich, Robert R. Hoffman and Neil Charness. Cambridge University Press, 2006.

LUCY READING-IKKANDA

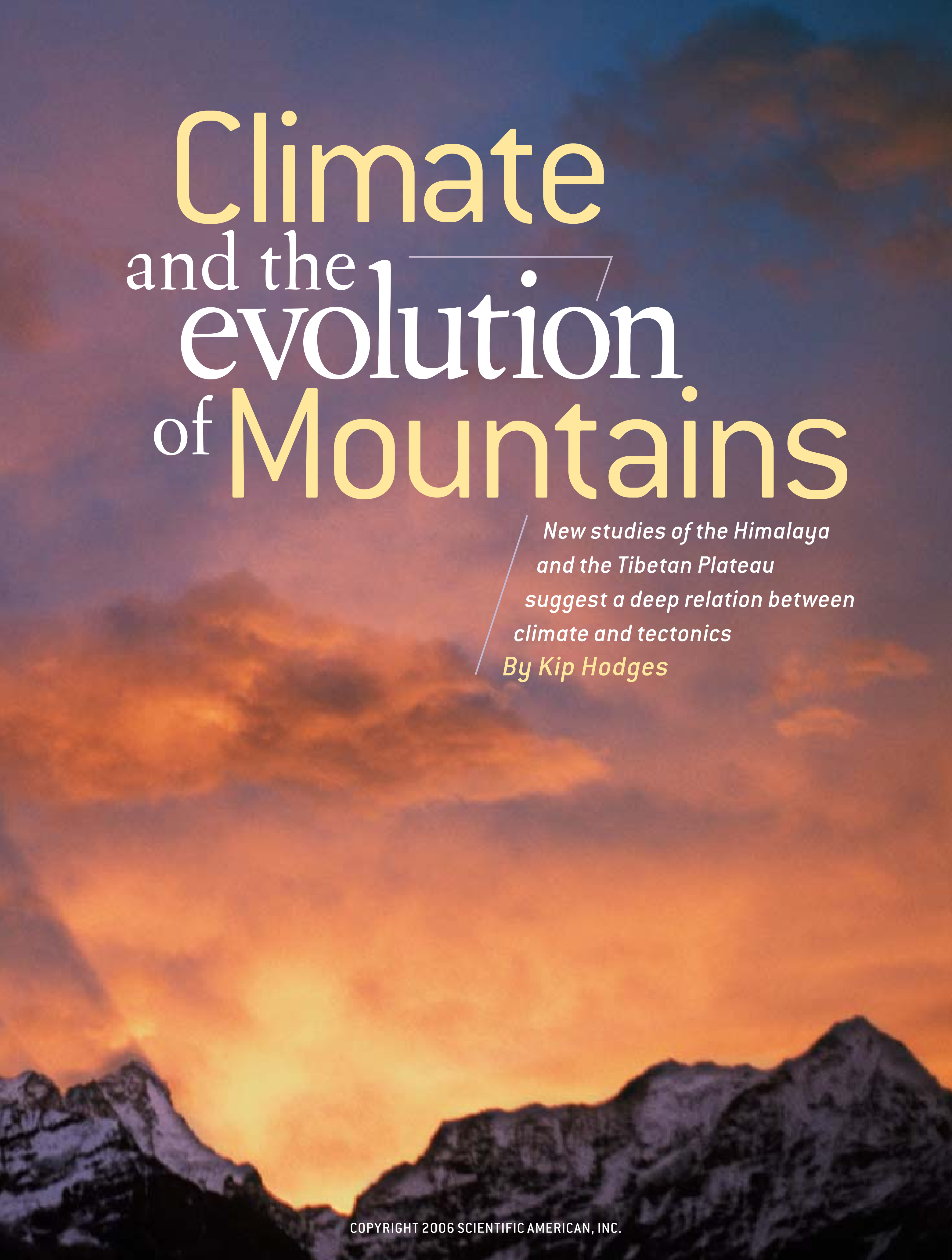
Six years ago I learned an important lesson about mountains and weather—from a farmer's horse. I was studying the geology of the ancient Kingdom of Mustang, now part of Nepal. Mustang lies high on the edge of the Tibetan Plateau, at the headwaters of the remarkable Kali Gandaki River, which carves a deep valley between the 8,000-meter-high peaks of Annapurna I and Dhaulagiri as it descends south to the Himalayan foothills. The farmer told me that the horse was perfectly good anytime for riding around on the plateau. But he mysteriously warned that if I were riding into the valley the horse would be good only in the morning.

I didn't understand his meaning until just after noon, when in a narrow part of the valley the horse and I found ourselves facing into a gale-force wind that seemed to come out of nowhere. As the winds intensified, the horse went slower and slower until it finally stopped, shook its head and turned around. No amount of prodding would get it to head into the wind with me on its back. As I dismounted and dragged it to the valley by its bridle, the thought occurred to me that this horse might know more about the weather patterns of the Himalaya than I did.

The winds of the Kali Gandaki Valley—some of the strongest upvalley winds on the planet, regularly reaching speeds of a steady 20 meters per second (just under 45 miles per hour)—are almost a daily occurrence. Air in the valley heats up after sunrise and rolls up from the valley as the day progresses. The predictable Kali Gandaki winds are just a local example of how climate can be affected by mountainous topography.

In turn, climate significantly influences the shape of the earth's surface. Wind, cold, heat, rain and ice all act as sculptors, and geologists have long studied the

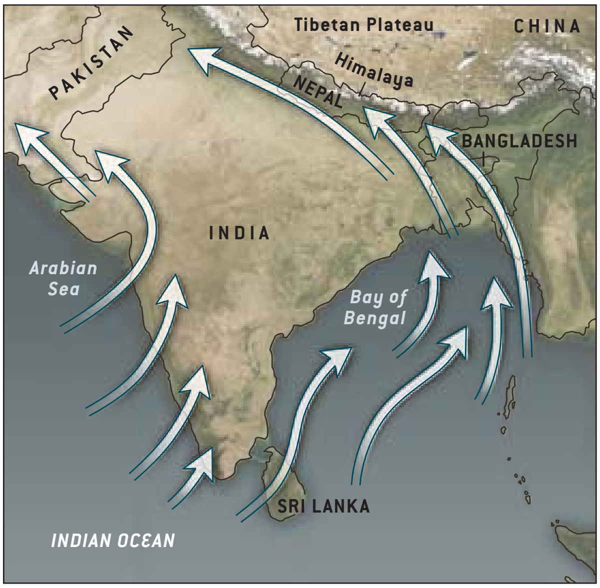
THE HIMALAYA at sunrise.



Climate and the evolution of Mountains

New studies of the Himalaya and the Tibetan Plateau suggest a deep relation between climate and tectonics

By Kip Hodges



MOIST AIR CURRENTS (*white arrows*) from the Indian Ocean drive the annual monsoons that occur during summer in India. When air masses tracking farther to the east flow over the Bay of Bengal, they pick up additional moisture, causing tropical depressions to form. Unable to climb high enough to pass over the Himalaya, the storms pound the land in their path; in places, as much rain falls as in a full year in the Amazon rain forest.

relation between climate and surface topography. But new research points to an unexpected, literally deep relation between climate and the evolution of mountain systems. Nowhere is this link better illustrated than in the intimate connection between the Indian monsoon and the continuing evolution of the Himalaya.

Monsoon Machinations

ONE OF THE MOST dramatic weather phenomena on earth, the annual Indian monsoon arises because springtime air temperatures rise rapidly over the Indian subcontinent but more slowly over the Indian Ocean. As a consequence, southwesterly winds bring a landward flow of moisture. Near Sri Lanka, some of the moisture-laden air begins to track northward across southern and western India while some heads eastward into the Bay of Bengal [*see illustration above*].

The eastward-tracking air mass picks up additional moisture from the bay, promoting the formation of a series of tropical depressions off the northeast coast of the subcontinent. These storms—attracted by a low-pressure trough that develops in northern India, along the southern flank of the mountains—race northward across Bangladesh and India. Finally, they strike the Himalaya.

The storms are then deflected westward, bringing on a monsoon season that begins in early June in northeastern India and Bangladesh, mid-June in Nepal, and late June in western India. Convection cells of moisture-laden air continually rise, like the hotter soup at the bottom of a pot bubbling to the surface, in a vain attempt to breach the wall of the Himalaya. As the air rises in these cells, the moisture condenses, resulting in torrential rain.

Meteorological stations in the Himalaya typically measure meters of rainfall every monsoon season (which can exceed the annual totals in much of the Amazon rain forest), concentrated in the Himalayan foothills at elevations between about 1,000 and 3,500 meters. Little precipitation makes it over the crests of the Himalayan ranges and onto the Tibetan Plateau, and so the climatic contrast across the Himalaya is extreme. In some regions, a traveler can pass from tropical forests, through rugged alpine terrain, and into high-altitude desert in less than 200 kilometers.

Obviously, the Himalaya, standing guard on the Tibetan Plateau, influence regional climate. But it turns out that climate can affect what goes on deep underneath the surface. How is such an effect possible? The answer has emerged from taking the understanding of how mountain ranges are built and adding new insights into how they interact with their surroundings. This synthesis builds on extensive research into several different geologic forces that control the formation and continual reshaping of mountains.

Mountain Building 101

THE TIBETAN PLATEAU and the Himalayan ranges are collectively known to earth scientists as the Himalayan-Tibetan orogenic system (after *oros*, Greek for "mountain," and *genic*, meaning "producing"). This region includes the 100 tallest mountains on the planet. The plateau, which is roughly equivalent in surface area to the Iberian Peninsula, has an average elevation higher than all but eight of the mountains found in the U.S. and Canada.

The theory of plate tectonics gives a

Overview/*Climate's Deep Impact*

- Although the power of climate to sculpt the earth's surface has long been recognized, new research shows that climate may also influence geologic activity deep underground and play a role in the deformational history of mountain systems.
- An example of the climate-geology relation can be found in the feedback mechanism between the Himalaya and the annual Indian monsoon. Heavy rains erode the south flank of the Himalaya. And the erosion, recent evidence suggests, may cause a slow, steady flow of deep crustal material from underneath the southern Tibetan Plateau to the range front, rebuilding it, to more efficiently intercept the monsoon.
- Rapid and recent uplift of the terrain that receives the rains supports the idea of the flow of deep crustal material to the range front. Evidence for such uplift includes geodetic data as well as unusually steep topography and stream channels across the range front. High erosion rates, verified by thermochronologic studies, also support this scenario.

BETH WALD *Aurora Photos (preceding pages)*; JEN CHRISTIANSEN; SOURCE: KIP HODGES; THE VISIBLE EARTH/NASA *(base satellite image) (this page)* partial explanation of how these landforms arose. The earth's great mountain ranges mark regions where sections of the broken, rigid outer skin of the planet, known as lithospheric plates, are currently colliding or have done so in the past. The Himalaya were born when the Indian plate, drifting northward from the Mesozoic supercontinent of Gondwana, slammed into the Eurasian plate about 45 million years ago. A freight train or oil tanker may require many minutes to stop once the brakes are applied—the inertia of the Indian plate was so great that even today it continues to converge with the Eurasian plate at a rate of about four centimeters a year.

The advancing Indian plate, acting as what physicists would call a rigid indenter, has continually forced part of the Eurasian lithosphere out of its way. As Paul Tapponnier and Peter Molnar suggested in the 1970s when they were at the Massachusetts Institute of Technology, the indenter has apparently pushed relatively rigid blocks of lithosphere, separated from one another by curved faults, eastward toward Southeast Asia [see "The Collision between India and Eurasia," by P. Molnar and P. Tapponnier; Scientific American, April 1977].

Another consequence of plate collision was shortening and thickening of the earth's crust (the top layer of the lithosphere). The average thickness of continental crust is about 30 kilometers. At mountain ranges, however, the crust can be much thicker. And the Himalayan-Tibetan orogenic system boasts the thickest crust on the planet—more than 70 kilometers in some places. The correlation between the thickest crust and the highest mountains is a manifestation of Archimedes' principle—an object immersed in a fluid is buoyed up by a force equivalent to the weight of the fluid it displaces. Just as a thick iceberg floats higher than a thinner iceberg because it displaces more of the denser seawater, a region of anomalously thick crust, such as the Himalayan-Tibetan orogenic system, "floats" higher in the underlying, denser mantle than do adjacent regions with thinner crust.

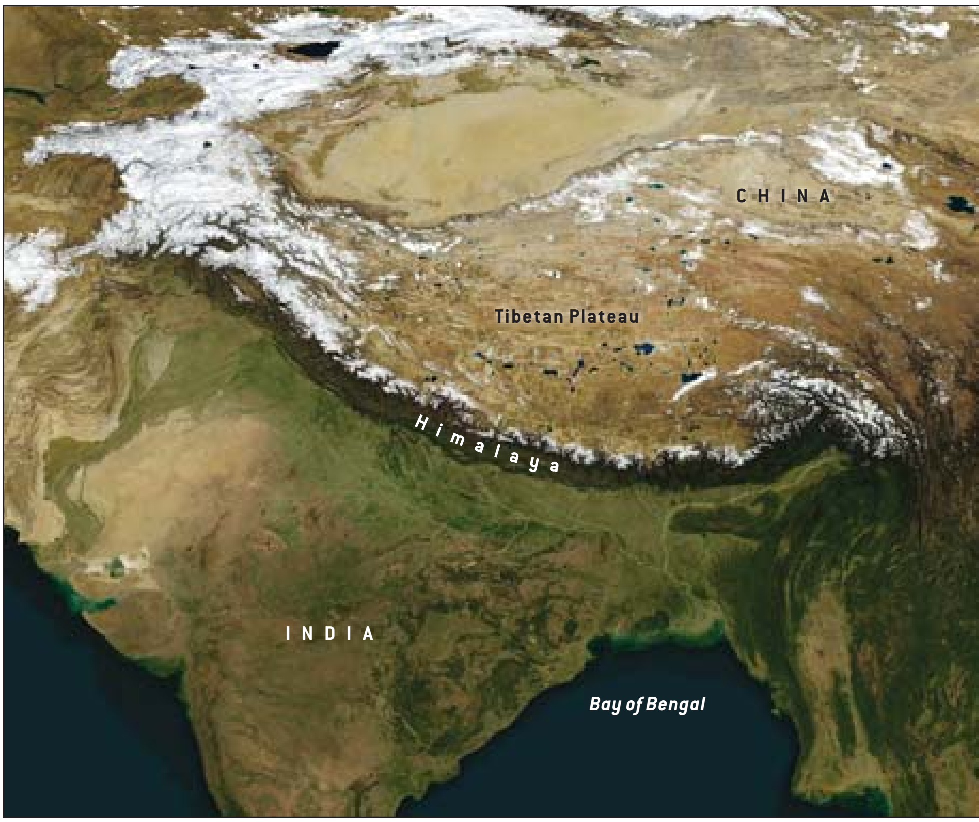
Energy Flow and Mountain Building

CRASHING PLATES help to explain the genesis of mountain ranges. But other processes, involving energy transfer, also influence mountain evolution. The great insight—and surprise—of the past few years is that monsoon rainfall appears to influence the way that energy transfer takes place at depth within the Himalayan-Tibetan system.

From the perspective of basic physics, an orogenic system is a storehouse of energy, akin to a reservoir behind a hydroelectric power plant. By impeding the flow of a river, a hydroelectric dam converts kinetic energy into gravitational potential energy stored in the form of a deep lake. Because of the elevation of the lake's surface compared with the river level downstream, a major difference in potential energy exists across a dam. Given the chance, the reservoir will rapidly dissipate its excess potential energy in an attempt to reestablish equilibrium with its surroundings—it will burst the dam. The only thing preventing this catastrophic action is the strength of the dam itself.

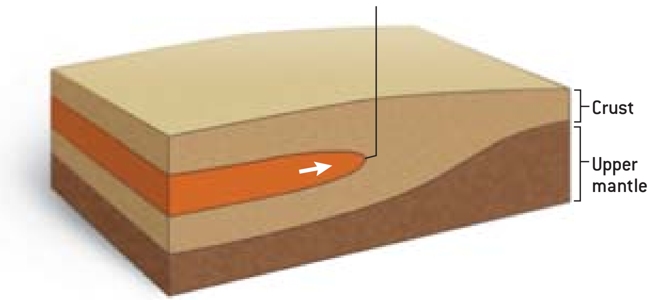
Similarly, the Himalayan-Tibetan orogenic system has a natural tendency to spread out to dissipate the excess potential energy it has by virtue of the difference between its crust thickness and that of the adjacent lowlands.

The rigid upper crust, broken into blocks by large faults, flows like sand. At deeper levels of the crust, where temperatures and pressures are much higher, the rock instead flows like toothpaste being squeezed through a tube. Theoretical models and field observations show that such deep crustal fluid channels may persist for millions of years. And given the opportunity to dissipate potential energy, these channels will slowly spread. Recent work by my M.I.T. colleagues Marin Clark and Leigh Royden supports the idea that the gentle, steady decline in elevation from the Ti-



THE VISIBLE EARTH/NASA

MONSOONS' EFFECTS on the surface are clearly visible in this satellite image of the Indian subcontinent. South of the Himalaya, the torrential rains promote the growth of vegetation, giving the area its green hue. North of the wall of mountains that intercepts the monsoon, the Tibetan plateau is relatively dry. White areas are snow-covered.



LATERAL FLOW of fluid lower crustal material underneath Tibet is theorized by some geologists. The flow would dissipate excess gravitational potential energy in the plateau region.

betan Plateau to the southeast was caused by outward flow of Tibetan lower crust [*see illustration below*] far under the eastern margin of the plateau.

In an exciting recent development, however, evidence has been found that suggests that the lower crust underneath Tibet is also flowing southward, directly toward the Himalayan mountain front. To begin to understand why, remember that water in a reservoir naturally flows preferentially to any crack in the dam. Similarly, the lower crustal channel might be expected to flow most readily in the direction of least resistance, toward the Himalayan range front, where surface erosion—driven by monsoon rainfall—is most active.

This crustal flow serves as a linchpin for emerging theories about the monsoon-mountain relation. As the rains relentlessly drive erosion along the Himalayan range front, the southern margin of the Tibetan Plateau should migrate northward. Instead, evidence suggests, the material removed by erosion is replenished by material in the lower crustal channel flowing southward and toward the surface. And to close the feedback loop, the erosion of the range front is what attracts the channel toward the surface in the first place.

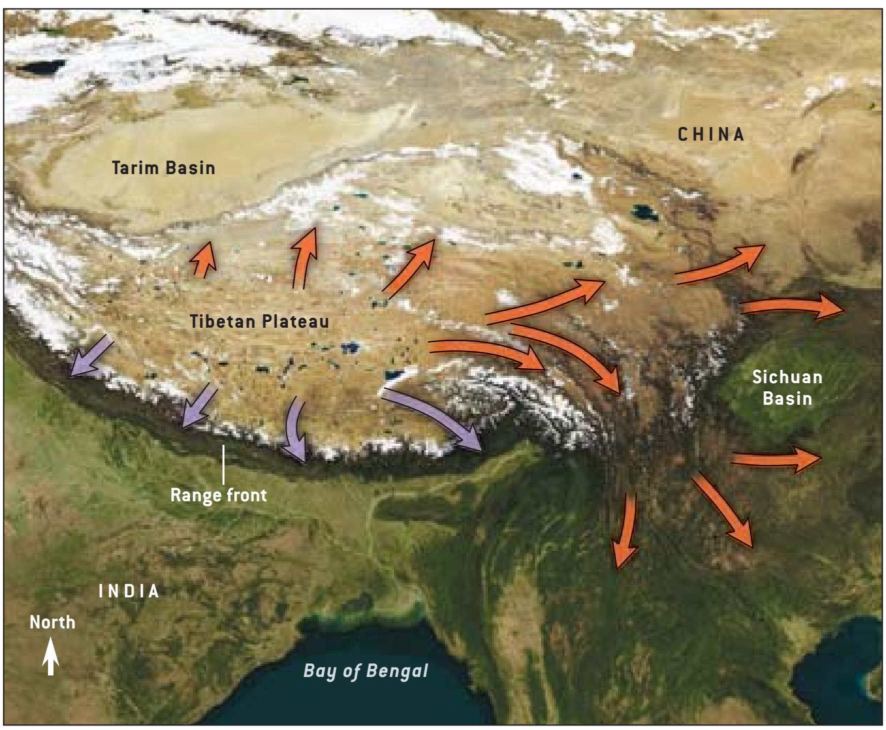
Dissecting the Flow

THIS INTERPRETATION has its roots in the study of major faults in the Himalaya. Most major faults in mountainous regions are thrust faults, the kind that typically develop during plate collisions. Those of us who have shoveled away New England snowstorms are all too familiar with thrust fault features: the shingled sheets of crusty snow that build up in front of our shovels are separated by fractures analogous to thrust faults. North-south vertical cross sections through the Himalaya show an architecture that is dominated by the Main Central, the Main Boundary and the Main Frontal thrust fault systems, all of which merge at depth to become what is known as the Himalayan Sole (as in "bottom") thrust [*see illustration on opposite page*]. Material above these faults is moving southward, carried by the conveyor belt of the thrust, relative to the material below.

Some 20 years ago, however, a research team at M.I.T. (I was a junior member then), along with investigators from other institutions, identified a second type of fault system in the Himalaya, which radically changed our view of continent-continent collisional mountain building. Lying near the crest of the Himalayan ranges, the South Tibetan fault system is characterized by faults with geometries similar to those of the thrust faults but with an opposite "shear sense": rocks above faults in this system have moved northward relative to the rocks below.

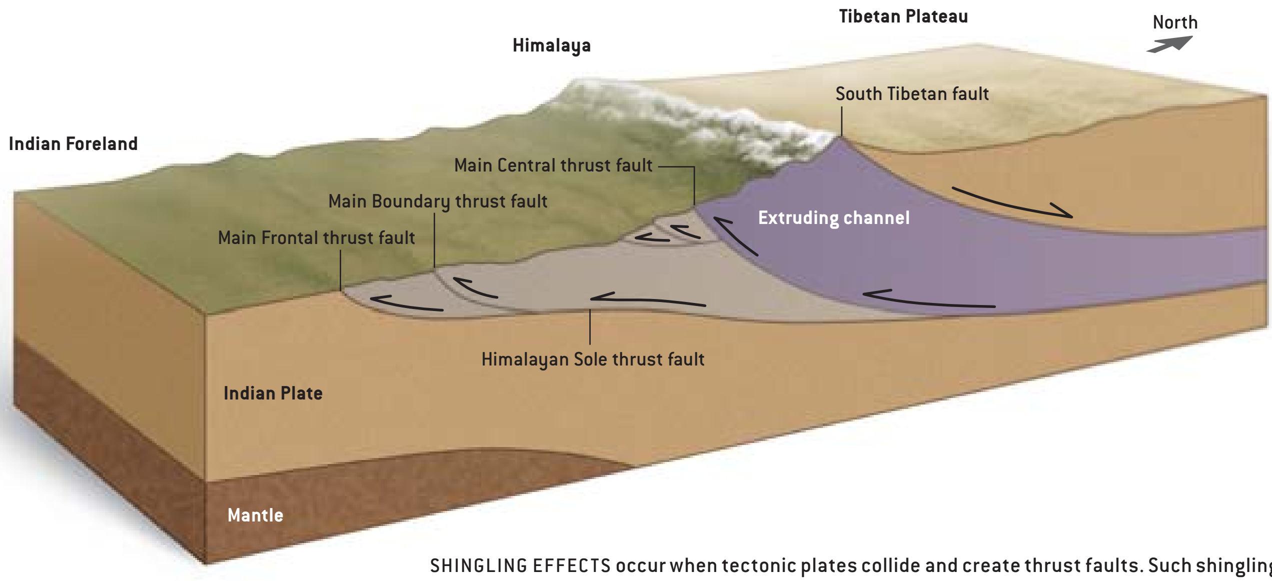
Such fault systems, which are known as detachment systems, are common in zones where the crust spreads and thins. Examples include mid-ocean ridges and the Basin and Range province of North America, which covers the area between the Colorado Plateau and the Sierra Nevada. But major detachment systems were not thought to occur in places where tectonic plates collide head-on until the South Tibetan system was discovered.

Attempts to incorporate the detachment system into the overall dynamics of the Himalayan-Tibetan orogenic sys-



FLOWING CHANNEL in many regions (*indicated by orange arrows on the satellite image above*) does not breach the surface but simply inflates the crust as it moves away from the plateau interior (*as shown in cross-sectional diagram at top*). But in regions where intense monsoon rainfall causes rapid erosion (*indicated by purple arrows*), the channel may be drawn to the surface (*cross-sectional diagram on opposite page*).

JEN CHRISTIANSEN; SOURCE: "A SYNTHESIS OF THE CHANNEL FLOW-EXTRUSION HYPOTHESIS AS DEVELOPED FOR THE HIMALAYAN-TIBETAN OROGENIC SYSTEM," BY K. V. HODGES, IN *CHANNEL FLOW, DUCTILE EXTRUSION, AND EXHUMATION OF LOWER-MIDDLE CRUST IN CONTINENTAL COLLISION ZONES*, EDITED BY R. LAW, M. SEARLE AND L. GODIN. GEOLOGICAL SOCIETY SPECIAL PUBLICATION, LONDON (IN PRESS) (*top*); JEN CHRISTIANSEN; SOURCE: KIP HODGES; THE VISIBLE EARTH/NASA (*bottom*)



SHINGLING EFFECTS occur when tectonic plates collide and create thrust faults. Such shingling—the result of the India-Asia plate collision—has occurred in the Himalaya. Faults of a second type are found near the crest of the Himalaya, dipping northward below the Tibetan Plateau. Constituting what is known as the South Tibetan fault system, these faults share geometric similarities with the thrust faults, but rocks slip along this system in the opposite direction. This fault system may also mark the top of the fluid lower crustal channel below Tibet. New evidence suggests that northward slip along the South Tibetan fault system and simultaneous southward slip along the southern faults permit the southward extrusion of this channel toward the Himalayan range front. (Tan regions are moving north. Purple and gray regions are moving south.)

JEN CHRISTIANSEN; SOURCE: "SOUTHWARD EXTRUSION OF TIBETAN CRUST AND ITS EFFECT ON HIMALAYAN TECTONICS," BY K. V. HODGES, J. M. HURTADO AND K. X. WHIPPLE, IN *TECTONICS*, VOL. 20, NO. 6, PAGES 799–809; 2001

tem have led us to theorize that the flow of ductile lower crust continuously replenishes the range front of the Himalaya: material between the Main Frontal thrust and the South Tibetan fault system, especially the rocks above the Main Central thrust, moves southward relative to the material both above the South Tibetan fault system and below the Himalaya Sole thrust. This southward extruding zone is thought to represent the ductile lower crustal channel of Tibet as it grinds its way to the surface.

Evidence indicates that both the South Tibetan fault system and the Main Central thrust system were active during the Early Miocene epoch, between about 16 million and 22 million years ago. In the late 1980s my M.I.T. colleagues Clark Burchfiel and Leigh Royden and I proposed a simple model in which coordinated slip on these fault systems during the Early Miocene led to southward movement of the channel of rocks between the two fault systems. In our view, this process was driven by the pressure differential between the rising Tibetan Plateau and India.

Since then, several research teams have refined this so-called extrusion model. Djordje Grujic, then at the Swiss Federal Institute of Technology in Zurich, and his colleagues showed that strain patterns in rocks within and between the two fault systems were consistent with ductile deformation, analogous to the flow of fluid in a pipe. And the late K. Douglas Nelson, then at Syracuse University, hypothesized that a fluid lower crust might have existed underneath Tibet as far back as the Early Miocene—and that the rock now exposed at the earth's surface between the South Tibetan and Main Central thrust systems may represent the ductile lower crustal channel that became more rigid as it reached the surface.

Most variants of the channel extrusion model [*see illustration above*] have focused on the Miocene evolution of the Himalayan-Tibetan system, but evidence is mounting that the extrusion process is ongoing. Several research groups have used geologic chronometers—based on the radioactive decay of elements in minerals that crystallized during deformational events—to establish a better understanding of the movement histories of major Himalayan fault systems. The results show that faulting has occurred near or just south of the surface traces of the Main Central thrust and South Tibetan fault systems at various times over the past 20 million years. And the most recent deformation is extremely young, by geologic standards: my group's research indicates

THE AUTHOR

KIP HODGES investigates the development and evolution of mountain systems, integrating theoretical and laboratory methods with field studies. When he wrote this article, he had been a professor in the department of earth, atmospheric and planetary sciences at the Massachusetts Institute of Technology for 23 years. In July he became founding director of the new School of Earth and Space Exploration at Arizona State University (thus ending his snow shoveling experiments). He has an undergraduate degree in geology from the University of North Carolina at Chapel Hill and a Ph.D. in geology from M.I.T. Although much of his work has focused on the Himalaya and the Tibetan Plateau, he also has done tectonic research in Scandinavia, polar East Greenland, Ireland, the western U.S., Baja California and the Peruvian Andes. The North Carolina native is editor of *Tectonics* and is on the editorial board of *Contributions to Mineralogy and Petrology.* An avid recreational diver, Hodges spends as much time as possible below the waves, gaining new perspectives on our overwater world.

that structures that could have acted as boundaries of an extruding channel have been active in central Nepal in the past few thousand years. Extrusion of the channel may even be happening today.

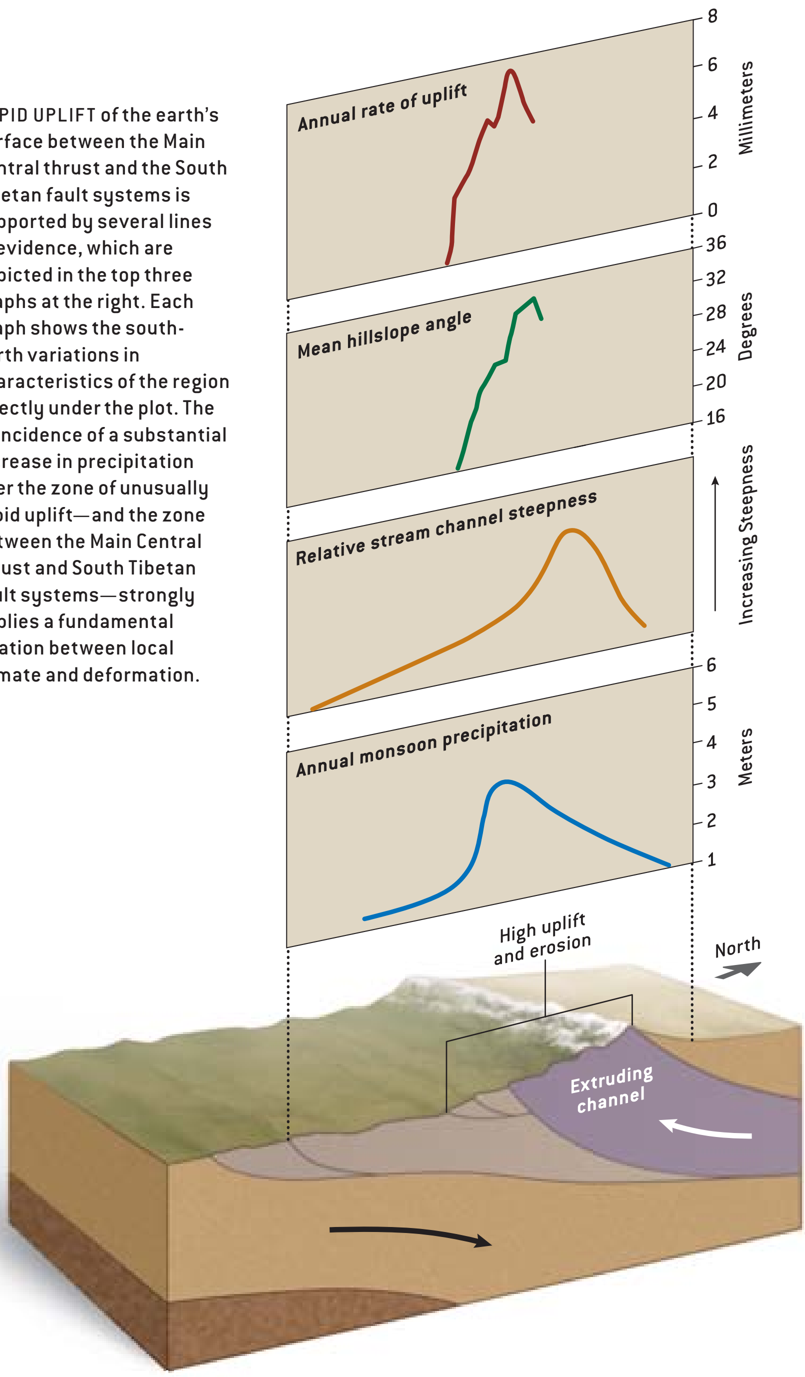
Extrusion and Landscape Evolution

IF CHANNEL EXTRUSION has played an important part in remodeling the Himalaya, evidence for unusually rapid uplift of the earth's surface in the zone between the surface traces of the fault systems should exist. Indeed, precise measurements made with traditional surveying equipment and Global Positioning System satellites over the past 30 years indicate that this zone is rising at a rate of a few millimeters a year relative to the region south of the surface trace of the Main Central thrust system.

Additional evidence for rapid uplift comes from the landscape itself. In mountain systems, rapid uplift is usually correlated with high topographic relief—a large change in elevation over a short horizontal distance—and steep river gradients. Stream profiles are in fact excessively steep across the Himalayan range front in Nepal where the streams cross the zone of proposed extrusion. Analysis of the topography across the front also shows unusually high relief there. But an important question is whether or not this phase of rapid uplift is simply a recent development over the past few decades or has persisted over geologic timescales.

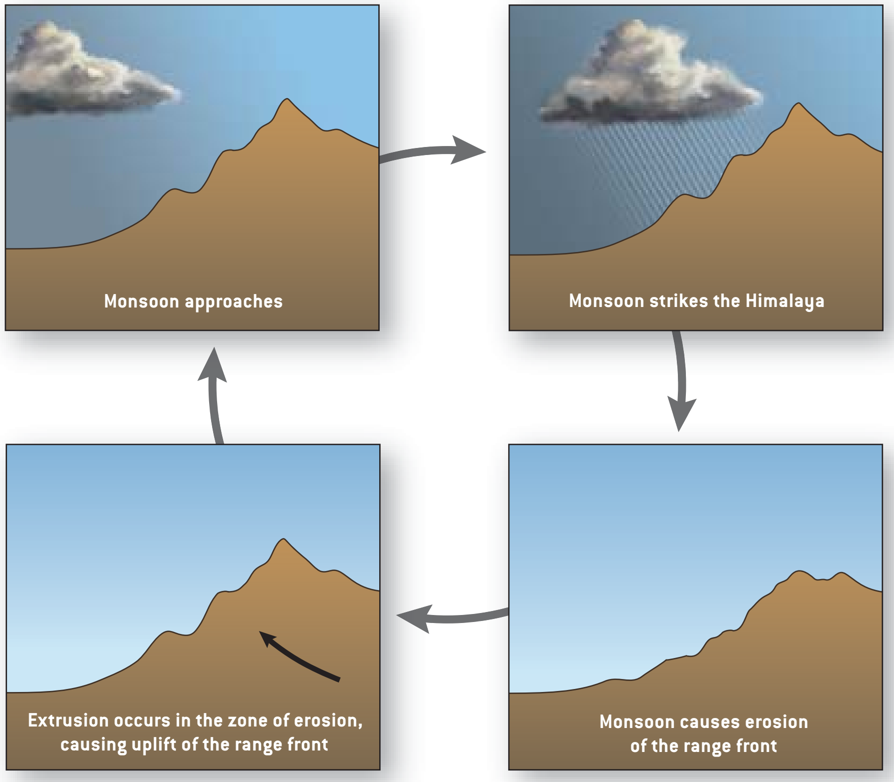
We know that erosion occurs in response to the monsoon today. But various techniques can provide estimates of erosion rates over long periods. And by knowing the history of erosion of the range front, we can get a good sense of the pace of uplift. One of the most powerful methods of estimating erosion rates in orogenic systems is based on the natural production of what are known as cosmogenic nuclides (isotopes produced by cosmic rays) in surface rock exposures. The concentration of cosmogenic nuclides, such as beryllium 10 and aluminum 26, in surface samples is proportional to the elapsed time since exposure. My M.I.T. colleagues Cameron Wobus and Kelin Whipple and I, working with Arjun Heimsath of Dartmouth College, have used this technique in central Nepal to demonstrate a more than threefold increase in erosion rate in the zone of proposed extrusion relative to the region farther south over millennial timescales.

Different methods can be used to explore erosion patterns on even longer timescales. Most people are familiar with the isotopic "clocks" used to determine the age of geologic or biologic materials. (A favorite example is dating based on the isotope carbon 14.) But not all geologic chronometers measure the crystallization age of a rock sample. Some instead indicate the time when a rock that initially crystallized at a high temperature deep underground cooled through a narrow temperature range as erosion allowed the sample to reach the earth's surface. By applying such thermochronometers to samples collected from the Himalayan front in the Annapurna range of central Nepal, my col-



RAPID UPLIFT of the earth's surface between the Main Central thrust and the South Tibetan fault systems is supported by several lines of evidence, which are depicted in the top three graphs at the right. Each graph shows the south-north variations in characteristics of the region directly under the plot. The coincidence of a substantial increase in precipitation over the zone of unusually rapid uplift—and the zone between the Main Central thrust and South Tibetan fault systems—strongly implies a fundamental relation between local climate and deformation.

JEN CHRISTIANSEN; SOURCE: KIP HODGES



FEEDBACK LOOP between monsoons and the Himalaya is shown here in simplified form. The summer monsoons track northward but are blocked by the ramparts of the Himalaya and drop enormous amounts of rain. The rain causes erosion, enabling the fluid lower crust of Tibet to extrude toward the range front. One effect of this extrusion is uplift, which counterbalances the tendency of rapid erosion to tear down the range front. By maintaining a steep range front, the extrusion process focuses monsoon rainfall, perpetuating the cycle.

JEN CHRISTIANSEN

league Ann Blythe of the University of Southern California and my student Kate Huntington of M.I.T. have shown that erosion has increased in the zone of proposed extrusion for at least the past few million years. Collectively, the evidence from landscape studies, cosmogenic nuclides and thermochronometry supports the idea of rapid uplift sustained over very long periods.

Running water and flowing glacial ice are recognized as principal agents of physical erosion. Higher rates of precipitation, such as those associated with monsoons, would lead to higher rates of erosion along the Himalayan range front. If the proposed feedback loop between range front erosion and southward extrusion of the crustal channel is viable, we would expect the pattern of high monsoon precipitation to coincide with the pattern of sustained surface uplift. Several groups are defining these precipitation patterns using data collected from remote sensing satellites and networks of weather stations.

For example, Ana Barros of Duke University has been leading a multi-institutional study aimed at defining monsoon precipitation patterns in the Annapurna range since 1999. The results in fact indicate that monsoon rainfall is highest in the zone of rapid erosion and uplift defined by other studies. More important, the broad correspondence of the region of high monsoon precipitation to the zone of proposed extrusion is consistent with the hypothesis that the channel extrusion is caused by range-front erosion in response to the Indian monsoon. The extrusion activated by rapid erosion does indeed appear to build the wall that intercepts the monsoon on its northward trek. Intense rainfall along the front leads to the increased erosion that activates extrusion, thereby completing the loop.

Theoretical models confirm that such relations are to be expected in evolving orogenic systems. Chris Beaumont and his colleagues at Dalhousie University in Halifax, Nova Scotia, have been using such models to explore the possible impact of erosion on systems with mechanically weak lower crustal channels. With a model constructed to represent the essential physics of the Himalayan-Tibetan system in the Miocene, they have shown that a weak lower crustal channel propagates (or "tunnels") laterally in regions with low surface erosion but toward the surface in regions of high erosion—such as the modern Himalayan range front.

The existence of feedbacks between climatic and geologic deformational processes is a novel concept. Researchers who have been concerned only with their specific discipline—for example, tectonics—find themselves in a more exciting and more deeply interconnected world. Just as solid-earth scientists, atmospheric scientists and hydrologists are already joining forces in this effort, we should soon see the integration of biological studies into the "new" tectonics—if climate can influence the growth of mountain ranges, perhaps ecosystems can as well. SA

MORE TO EXPLORE

Himalayan Tectonics Explained by Extrusion of a Low-Viscosity Crustal Channel Coupled to Focused Surface Denudation. C. Beaumont et al. in *Nature,* Vol. 414, pages 738–742; December 13, 2001.

Southward Extrusion of Tibetan Crust and Its Effect on Himalayan Tectonics. K. V. Hodges et al. in *Tectonics,* Vol. 20, No. 6, pages 799–809; 2001.

Has Focused Denudation Sustained Active Thrusting at the Himalayan Topographic Front? Cameron W. Wobus et al. in *Geology,* Vol. 31, No. 10, pages 861–864; October 2003.

Monitoring the Monsoon in the Himalayas: Observations in Central Nepal, June 2001. Ana P. Barros and Timothy J. Lang in *Monthly Weather Review,* Vol. 131, No. 7, pages 1408–1427; 2003.

Quaternary Deformation, River Steepening, and Heavy Precipitation at the Front of the Higher Himalayan Ranges. Kip V. Hodges et al. in *Earth and Planetary Science Letters,* Vol. 220, pages 379–389; 2004.

A Great Leap in Graphics

The quality of 3-D computer graphics is poised for a quantum jump forward, thanks to speedier ways to simulate the flight of light

By W. Wayt Gibbs

For those of us who frittered our formative years away blasting blocky space invaders, video games today can widen the eyes and slacken the jaw. The primitive pixelated ape of Donkey Kong has evolved into a three-dimensional King Kong of startling detail. Some newer Xbox 360 games render their lead characters from an intricate mesh of more than 20,000 polygons, each tiny patch drawn dozens of times a second with its own subtle texture, shading and gloss.

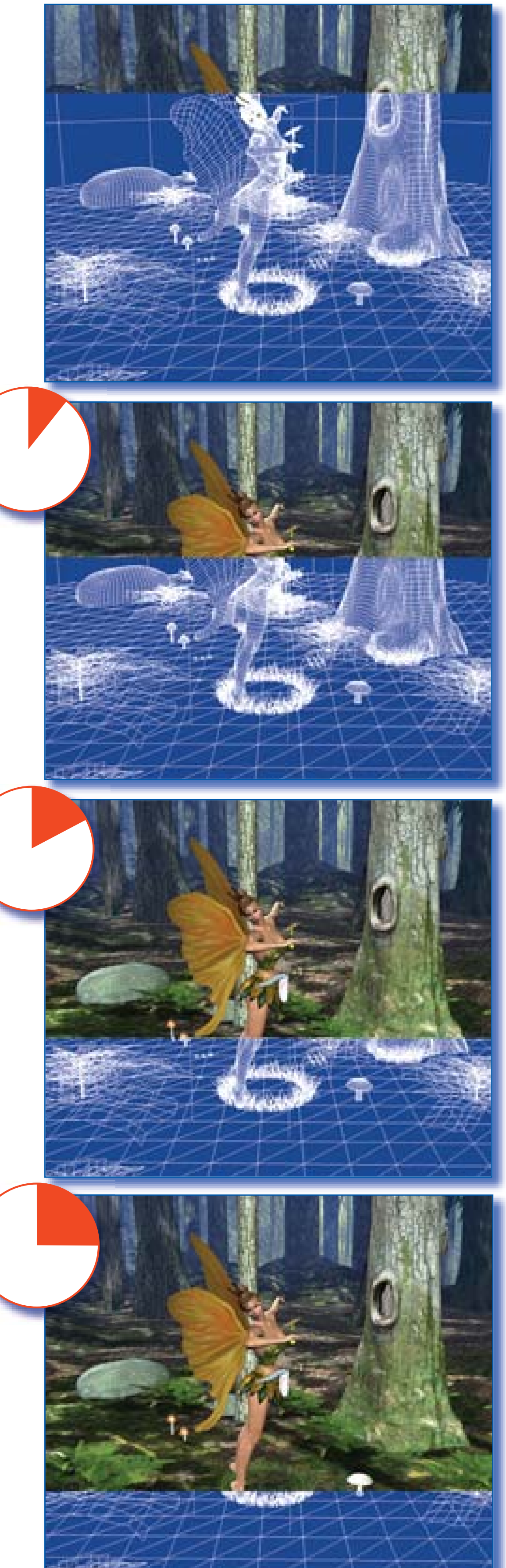
Beyond the booming game industry, the evolution of graphics has lifted interactive software for design, engineering, architecture, medical imaging and scientific visualization to new heights of performance. Much of the credit belongs to advances in graphics processing units (GPUs), the microchips at the heart of computer video cards that transform 3-D scenes into 2-D frames at speeds faster than a trigger twitch. As the rendering capabilities of GPUs soared, so did the revenues of ATI, NVIDIA and Intel, which make the most popular models.

A wide gulf of verisimilitude, however, still separates interactive graphics from feature film effects and photography. And some experts say the only way that personal computers will ever cross that gulf—to reach the nirvana of computer graphics in which synthetic scenes display all the fluid motion and subtle shadings of reality—is through a basic change in how machines render 3-D models.

This change, from the so-called rasterization method that GPUs use to a more scientific approach known as ray tracing, was long dismissed as infeasible for interactive, rapidly changing scenes. But advances in both software and hardware have recently propelled ray tracing to within range of the consumer PC market.

Ray tracing, which originated with Renaissance painter Albrecht Dürer and

W. WAYT GIBBS; MODELS COURTESY OF DAZ PRODUCTIONS AND INGO WALD



was first reduced to computer code in the 1970s, is now in the early stages of its own renaissance, says Philipp Slusallek, a computer science professor who leads a ray-tracing research group at Saarland University in Germany. The timing is fortuitous. "GPUs are running up against the wall," he asserts.

Ray Tracing Reborn

THREE PROBLEMS VEX GPUS as they try to render a 3-D scene by rasterizing it. All three weaknesses stem from the initial step in rasterization: dissecting the virtual world into a montage of flat polygons such as triangles, each of which is treated as if it were independent of the rest. The first issue is that most real objects have curves and look unnatural when approximated by facets.

When modelers can afford the labor involved, they add more polygons to round out bulges and concavities. But that exacerbates the second problem, which is that raster systems are designed to process every polygon in the model, even those hidden from view. "Until it has rendered the very last polygon, the system won't know whether that final shape covers up every other object," explains Gordon Stoll, a graphics researcher at Intel. The computational cost of a raster rendering thus rises in direct proportion to the geometric complexity of the scene. So every time the amount of detail doubles, the rate at which a GPU churns out frames of video falls by about half.

A third and more important problem, Slusallek argues, is that "shadows, reflections and other effects simply cannot be done right on GPUs." The reason, Stoll explains, is that "it is just physically incorrect to assume that polygons are independent. The appearance of every object in a scene depends on every other object, because light bounces around." Computer scientists refer to this phenomenon as the problem of global illumination. Newer GPUs can reprocess the scene multiple times to approximate indirect lighting. But that work-around chews up memory, clogs internal data channels and still falls far short of photorealism.

Ray tracing avoids these problems by simulating the ricochets of light rays throughout a scene. Where raster engines rely on tricks, approximations and hand-tweaked artistry, ray-tracing programs instead lean on the laws of optics—and with physics comes fidelity. "Rays of light really *are* independent," Stoll notes, so in ray-traced images, reflections, shadows and smoke appear as they should.

"And when I turn on these effects in a ray tracer, they automatically compose properly, so I can get a reflection of a shadow of a puff of smoke—that's not true for raster graphics," he adds. So far only ray tracers have been able to arrive at near-perfect solutions to the problem of global illumination.

When Pixar Animation Studios set out to make *Cars,* an animated film that opened in June, artists there found that the metallic bodies of the vehicular characters did not shine quite right unless they added ray tracers to the usual raster rendering system. On all its previous movies, the studio had resisted using ray tracing for the same reason that makers of games and other interactive software have: the intensity of computation that the physics demands has always brought microprocessors to their knees. Even with Pixar's fast network of 3,000 state-of-the-art computers, each second of finished film took days to render. Film producers may tolerate such delays, but gamers, engineers and doctors generally will not.

Ray tracing now seems set to leap the speed barrier, however. Better algorithms and custom-designed hardware have accelerated ray-tracing renderers by more than two orders of magnitude. As ray tracing goes "real time," fast-moving computer imagery should become both dramatically more realistic and easier to create.

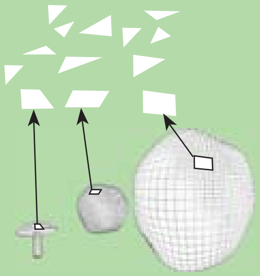
By 2003, Slusallek was so convinced real-time ray tracing was ready for mar-

FLEET-FOOTED ALGORITHMS and custom microchips for ray-tracing 3-D models can now render hundreds of frames of a complex and rapidly changing scene (*opposite page*) in less time than conventional ray-tracing software takes to draw a single frame (*this page*).

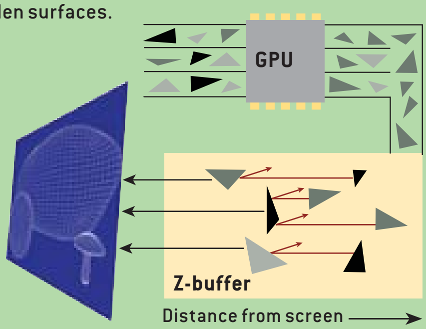
FROM FAST RASTERS TO REAL-TIME RAYS

RASTER GRAPHICS

Raster graphics systems in today's computer video cards and home console game machines start with a 3-D scene that the software has divided into objects and then subdivided into polygons (typically triangles).



The polygons stream through "pipelines" in the graphics processor (GPU), which work on many pieces in parallel to transform their geometry and to compute their shading. A "z-buffer" then sorts polygons by distance from the viewpoint, selects the frontmost for display, and tosses out the computed results for hidden surfaces.

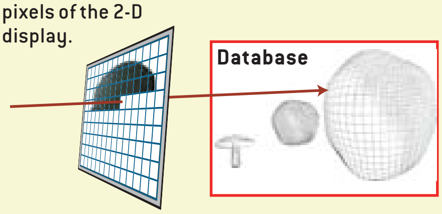


The polygons often must pass through the pipelines multiple times for each frame of video so that the system can calculate the shading, textures, translucency, reflections and other effects that add realism to the scene.

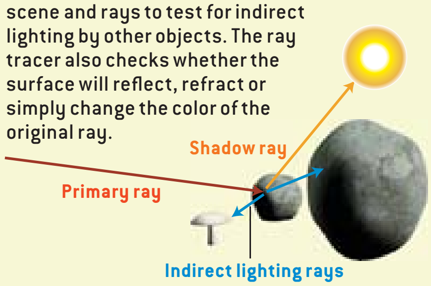
TRADITIONAL RAY TRACING

Ray tracing renders a 3-D scene by shooting virtual rays through the pixels of the 2-D display.

The scene is stored as a database of objects, which can include curved as well as flat surfaces.



When a ray hits an object, the system launches "shadow" rays toward each light source in the scene and rays to test for indirect lighting by other objects. The ray tracer also checks whether the surface will reflect, refract or simply change the color of the original ray.



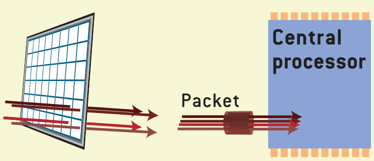
Thanks to its recursive nature, this method can render a scene accurately in just one pass.

REAL-TIME RAY TRACING

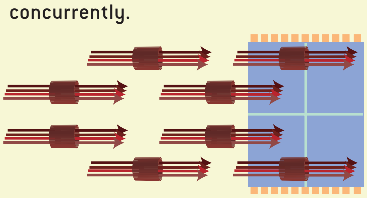
Real-time ray tracing is already in commercial use on high-end servers and is becoming increasingly feasible on consumer-level computers. Three kinds of advances have cut rendering time from hours to a fraction of a second per frame.

1 Running rays together in parallel now happens at several levels within real-time ray tracers on desktop PCs.

Programs group similar rays into "packets," then march all the rays in a packet in lockstep through the same set of computations.

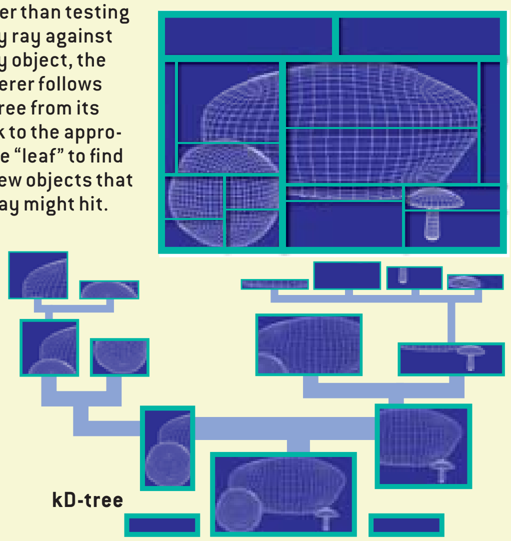


Each packet is processed by a separate program thread. Newer multicore CPUs can run more than a dozen such threads concurrently.



2 Acceleration structures split the 3-D scene into a hierarchy, called a kD-tree, organized so that each section carries roughly equal computational cost.

Rather than testing every ray against every object, the renderer follows the tree from its trunk to the appropriate "leaf" to find the few objects that the ray might hit.



3 Customized microchips built last year at Saarland University run at a mere 66 megahertz in prototype versions yet can render some ray-traced scenes more quickly than a 2,600-megahertz Pentium-4 system.

Once it is commercialized, the "ray processing unit," or RPU, should run roughly 50 times as fast—more than speedy enough for interactive software.

Prototype RPU is tested by its creators at Saarland University.

W. WAYT GIBBS; MODELS COURTESY OF DAZ PRODUCTIONS; PHILIPP SLUSALLEK *Saarland University* (*photograph*)

EVERY BOLT, KNOB AND CABLE is included in Boeing's 3-D model of its 777 jet. A scene of this complexity overwhelms traditional raster rendering approaches, because they attempt to process every one of its 350 million polygons when drawing a frame of video. Ray-tracing systems, which by their nature ignore any objects that are hidden from view, handle such detail with aplomb.

PHILIPP SLUSALLEK *Saarland University*; MODELS PROVIDED BY AND USED WITH PERMISSION OF BOEING COMPANY

ket that he and his colleagues at Saarland spun off a company to commercialize the technology. They founded inTrace, using first-generation software that required a cluster of high-powered servers to render full-resolution, photorealistic images at 10 frames a second or higher. BMW, Volkswagen, Airbus and other firms snapped up inTrace systems despite the price and now use them to evaluate the sight lines, interior reflections and curb appeal of vehicle designs still on the drawing boards.

Boosting the Speed of Virtual Light

RAY-TRACING ALGORITHMS have meanwhile improved so much that they can achieve interactive speeds on a single high-end PC. In 2004 Slusallek and his co-workers Carsten Benthin and Ingo Wald demonstrated a way to use ray tracing to rapidly render scenes crowded with curved, free-form shapes, without first carving them into polygons the way GPUs and "classic" ray-tracing programs do. And within the past two years Slusallek, Stoll and Wald (who is now at the University of Utah) have each demonstrated significantly faster ways to identify which object, if any, a virtual photon hits as it traverses a scene from screen to light source. (To save on computation, the rays are traced in reverse; the physics is the same.)

Testing a ray for collisions with an object, Slusallek says, is like finding a specific book at a library. "You don't start looking at the top of the first shelf each time. You use an index." Building an index to the database of 3-D objects is easy. The hard part is figuring out how to rebuild it within a few milliseconds every time something changes in the model. "It's as if the shelves in the library are constantly moving," he says.

Stoll and his Intel colleague Jim Hurley are working on a system that improves on the classical approach, which is an index known as a kD-tree. A kD-tree chops the 3-D space into pieces of various sizes and then arranges them into a treelike 3-D hierarchy. To save time, Intel's Razor system selects the branches of the tree it needs for each frame and rebuilds only those. Razor's kD-tree also represents the scene at multiple levels of detail, so that it can quickly draw a distant castle, for example, without having to fetch data on every one of its bricks.

"Razor is our most ambitious, over-the-horizon work," Hurley says. "It can handle explosions, splashes and the full range of lighting effects. It's not very fast yet, but by optimizing the code, we expect to boost performance by a factor of 50 to 100." The optimizations will exploit the ability of the latest "multicore" central processors to run a dozen or more program threads simultaneously.

Last year the Saarland group designed a new chip that can run many ray-tracing calculations in parallel. In tests the "RPU"—or ray processing unit—is able to churn out tens of ray-traced frames a second. To demonstrate its abilities, Slusallek had two students create a virtual island with 40 million polygons and ocean waves that reflect stars and fires on the beach. "We did all this in a couple of months," he reports. "With ray tracing, you don't need all the artistic work that goes into making rasterized environments look realistic; you just make the model, press a button, and it looks great." Slusallek is now lining up investors to commercialize the design as his team crafts software drivers and compilers for the RPU.

"The technology is finally there to handle highly detailed, realistic environments in real time," Slusallek says. "It isn't clear yet which platform will be the one that allows ray tracing to really take off into the mainstream: it could be big multicore CPUs, customized processors like our RPU, or ray-tracing features added incrementally by the GPU makers." What does seem clear is that a disruptive shift has begun that signals a new burst of evolution in computer graphics. SA

W. Wayt Gibbs is senior writer.

MORE TO EXPLORE

Ray Tracing Goes Mainstream. Jim Hurley in *Intel Technology Journal*, Vol. 9, Issue 2; May 2005.

RPU: A Programmable Ray Processing Unit for Realtime Ray Tracing. Sven Woop, Jörg Schmittler and Philipp Slusallek in *Proceedings of ACM SIGGRAPH, 2005.*

Razor: An Architecture for Dynamic Multiresolution Ray Tracing. Gordon Stoll et al. University of Texas at Austin, Department of Computer Sciences Technical Report #TR-06-21, 2006.

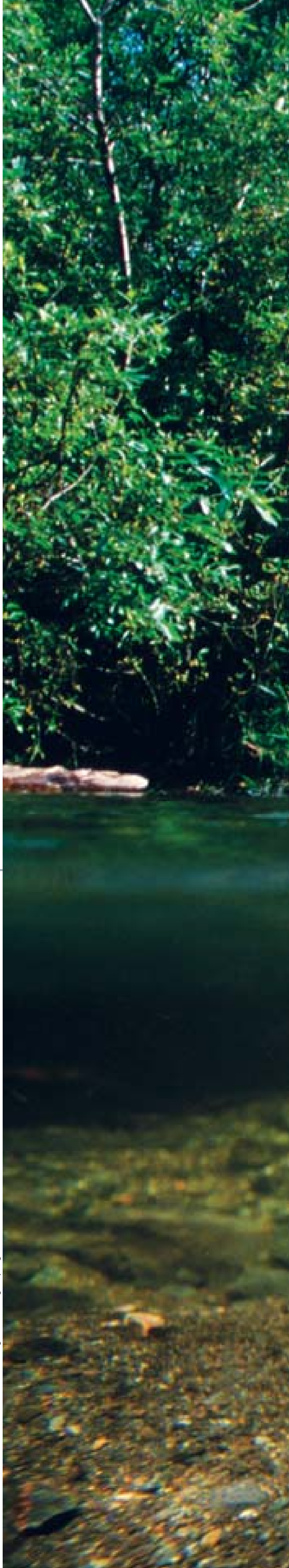
the fish & the forest

By Scott M. Gende and Thomas P. Quinn

Salmon-catching bears fertilize forests with the partially eaten carcasses of their favorite food

Few wildlife spectacles in North America compare to the sight of bears gathered along streams and rivers to scoop up spawning salmon. The hungry bears have long attracted attention, particularly from fishery managers, who in the late 1940s proposed their broadscale culling in Alaska to reduce the "economic damage" the predators might be wreaking on salmon populations. In fact, several sensationalized reports implied that Alaska might fall into "financial and social collapse" unless the bear populations were controlled.

Fortunately, common sense came to the rescue, and the bear cull never came about. Scientific interest in the interaction between bears and salmon died down. Recently, however, researchers have discovered a new facet of this relationship, and the finding has radically changed notions about how the salmon, the streams and the bordering woodlands affect one another—and, naturally, notions about how they should be managed.



SOCKEYE SALMON swims past a foraging brown bear in a small stream in Alaska. The fish turn bright red with a pale green head as they prepare to spawn in freshwater.

JOHN HYDE *Wild Things Photography*

Our own contributions to this work have spanned more than a decade. During this time we have walked hundreds of kilometers along salmon streams, examined tens of thousands of salmon carcasses, and had too many close encounters with agitated bears. Our findings surprised us: bears actually fertilize the forests, nourishing them by discarding partially eaten salmon carcasses. Not intentionally, of course, but the end result is that these large predators bring valuable marine-derived nutrients, in the form of salmon tissue, to the streamside woodlands, where the uneaten fish provide sustenance for an array of animals and plants. The flow of nutrients from ocean to streams to woodlands is an unexpected, even unprecedented, uphill direction for resources to travel. A close look at the life history of the predator and its favorite prey helped us and other scientists piece together how this unusual transfer system operates.

The Nutrient Express

PACIFIC SALMON—including chinook, coho, chum, pink and sockeye—vary in abundance, size and behavior, but all members of this genus (*Oncorhynchus*) share the same general life cycle. Young salmon emerge from the gravel in streams or lakes in spring and then, over various periods of time, migrate to the ocean. After living at sea between one and four years, they return to their natal streams to spawn and die. The young salmon are quite small when they leave freshwater, weighing from less than one gram to about 20 grams, and they are quite big when they return, ranging from two to 10 kilograms or more. Consequently, even though most juveniles die at sea, the return migration and death of adult salmon produce a large net flow of nutrients and energy from the ocean to stream and lake ecosystems.

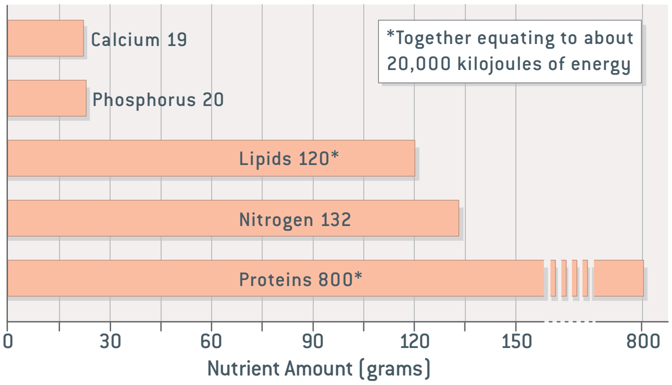
This influx of energy from the ocean can have an extraordinary effect on freshwater systems because the nutrient composition of the fish and their densities are so great. For example, an adult male chum salmon on the spawning grounds contains an average of 130 grams of nitrogen, 20 grams of phosphorus and more than 20,000 kilojoules of energy in the form of protein and fat. Multiplying the average nutrient composition of salmon by the total number of returning fish, we found that a 250-meter reach of stream in southeastern Alaska received more than 80 kilograms of nitrogen and 11 kilograms of phosphorus in the form of chum salmon tissue in just over a month.

The behavior of the bears that feast on the salmon—brown bears (*Ursus arctos*, also known as grizzlies at inland locations) and black bears (*U. americanus*)—is the other part of the equation. Salmon are a crucial resource for the bears because the survival and reproductive success of these large mammals depend on the amount of fat they can deposit in the late summer and fall. Once bears enter their dens in the early winter, they neither eat nor drink for up to seven months. Yet bears are not true hibernators—their body temperature stays above ambient levels—so they must produce body heat to maintain metabolic functions throughout the cold months. In addition, females give birth and lactate during this time.

Because the bears' survival and reproductive success are closely tied to their physical condition in the autumn, natural selection favors those that get the most nourishment out of the fish they eat. And bears exhibit two behaviors to this end.

NUTRIENTS IN SALMON

Based on adult male chum salmon



BEAR PREDATION

Creek	Average no. of salmon	Average no. killed by bears	Average percent killed
Bear	3,907	1,183	32
Big Whitefish	786	342	48
Eagle	818	399	53
Fenno	5,228	666	12
Hansen	6,229	2,450	49
Hidden Lake	2,010	671	43
Little Whitefish	173	93	58
Pick	5,837	1,949	35

SALMON contain such valuable nutrients (*above*) and their population densities (*below*) are so great that the fish have a huge impact on freshwater systems. The authors calculated that as a result of the foraging bears along several streams in Alaska the total amount of nitrogen and phosphorus provided by salmon carcasses equals or exceeds recommended concentrations of commercial fertilizer for plants in northern forests.

LUCY READING-IKKANDA

Overview/*Sea to Shore*

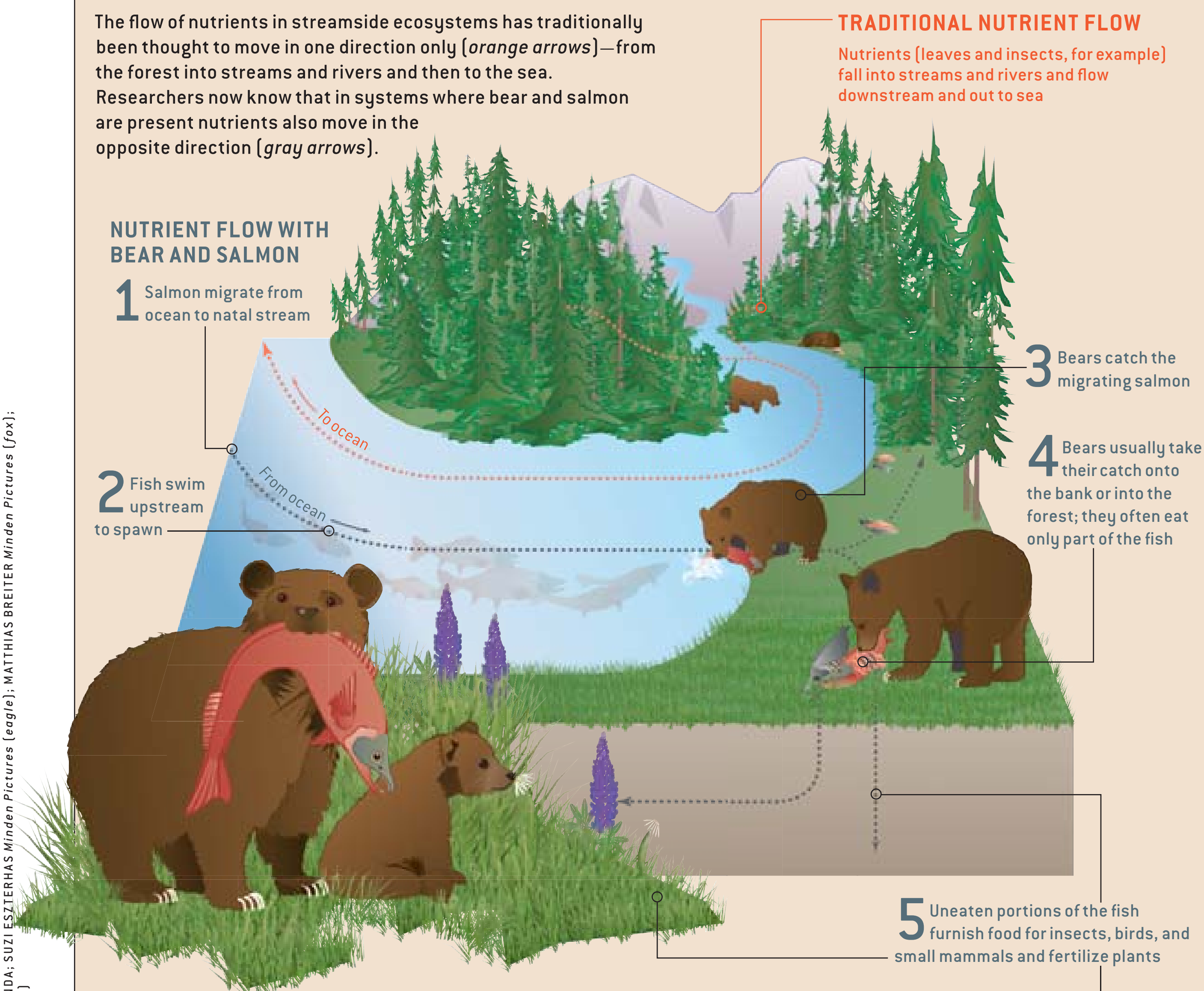
- To avoid confrontations, bears often carry their salmon catch onto the streamside bank or into the forest.
- Once safely alone, they usually eat only the most nutritious part of the fish and discard the rest, which still contains many valuable minerals and calories.
- These rich remains support a hierarchy of animals and plants.
- The dependence of the ecosystem on the salmon carcasses has captured the attention of fishery and forest managers, who now recognize the importance of both bears and salmon to the system.

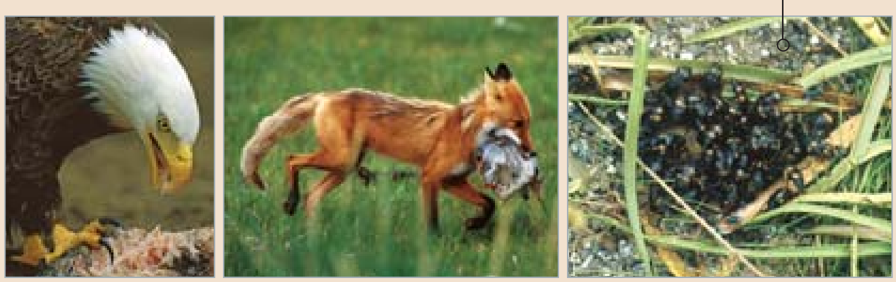
First, to avoid interference from other bears, they often carry captured salmon to the stream bank or into the riparian (streamside) forest before eating. Bears are relatively solitary for most of their adult lives, save for a few weeks of courtship in spring and summer. When they aggregate in riparian areas to feed on salmon, they become aggressive. These confrontations can be relatively benign, resulting in one bear stealing a fish from another, or violent, ending in serious injury or death to a bear or its offspring. Carrying the carcass into the forest out of sight of other bears is a way to forestall confrontation.

The second important behavior is that the bears often eat only the most nourishing part of the salmon. If the salmon densities are high, it takes a bear less than a minute to capture a fish in a small stream, and under these conditions of plenty,

ENERGY FROM THE SEA

The flow of nutrients in streamside ecosystems has traditionally been thought to move in one direction only (*orange arrows*)—from the forest into streams and rivers and then to the sea. Researchers now know that in systems where bear and salmon are present nutrients also move in the opposite direction (*gray arrows*).





CREATURES that scavenge the discarded salmon include (*left to right*) bald eagles, red foxes and carrion beetles.

ILLUSTRATION BY LUCY READING-IKKANDA; SUZI ESZTERHAS *Minden Pictures* (*eagle*); MATTHIAS BREITER *Minden Pictures* (*fox*); COURTESY OF SCOTT M. GENDE (*beetles*)

THE AUTHORS

SCOTT M. GENDE and *THOMAS P. QUINN* have long shared an interest in the interactions between bears and salmon. Gende is a coastal ecologist with the National Park Service in Juneau, Alaska. He has focused much of his research on the ecological consequences of salmon in aquatic and terrestrial ecosystems. Quinn has been a professor in the School of Aquatic and Fishery Sciences at the University of Washington since 1990. He is author of *Evolution and Behavior of Pacific Salmon and Trout* (University of Washington Press, 2005).

the bears rarely eat the whole fish. An analysis of more than 20,000 carcasses revealed that bears consumed about 25 percent of each captured salmon, selectively eating only the parts highest in fat content, such as the eggs. In fact, it is common for bears to carry a carcass to the stream bank and not eat a bite after finding it is a male salmon or a female that has already spawned her eggs. (Salmon do not feed once they enter freshwater, so their body fat, which is quite high initially, is progressively depleted—by 90 percent or more—as they migrate and spawn.)

After consuming choice morsels, bears abandon the carcass and return to the stream to spear another fish. Thus, bears kill far more salmon than they eat. At a small stream in southeastern Alaska, for instance, we observed a 200-kilogram female brown bear capture more than 40 chum salmon during several foraging bouts over the course of eight hours. She removed over 143 kilograms of salmon (70 percent of her body weight!) from the stream but consumed only a small fraction of this bounty.

Special Delivery

WHY IS THIS UNUSUAL feeding behavior important for the vitality of the ecosystem? After all, in the absence of bears, the salmon would still die following spawning, and their carcasses would be scavenged by birds, fishes and insects in the streams, decomposed by microbes and flushed out to the ocean. By killing many of the fatter salmon, carrying the nutrient-loaded fish to the forest, and abandoning the carcass with most of the biomass remaining, bears make a tremendous amount of food and nutrients available to streamside plants and animals that would not otherwise have access to this resource. The bears are truly ecosystem engineers: they deliver marine-derived nutrients to the riparian system.

The spread occurs because many different animals make use of the protein and fat in the abandoned fish. Flies, beetles, slugs and other invertebrates colonize the carcasses almost immediately and deposit their eggs there. Gulls, ravens, crows, jays, magpies, mink, marten, and other species of birds and mammals readily and often quickly make a meal of the carcasses. (We once observed a bear capture a fish and walk into

Why Some Bears Prefer Berries

Perched on a small wooden platform nearly 30 meters high in a streamside tree, the two of us have spent more than 1,000 hours watching bears spear salmon. We soon noticed that a loose social structure among bears forms at even the smallest streams. In general, larger bears win confrontations or are avoided by smaller bears, regardless of whether the contestants are male or female. Subadults and small females, particularly those accompanied by young cubs, tend to be the most subordinate.

Dominant bears forage more often and for longer periods than other bears; they capture more salmon in each foraging bout; and they carry the carcasses shorter distances from the stream. They also consume less from each captured fish. Subordinate bears kill fewer fish per foraging bout, carry the carcasses much farther from the stream, and eat more tissue from each captured fish.

One ramification of this behavior is that small streams may have an upper limit to the number of bears that will feed there. As the bears become more numerous and aggressive interactions increase, subordinate bears may actually have greater success feeding on lower energy foods, such as berries or grasses, than on salmon. Such pressures may explain why we often see bears in upland or alpine areas far from salmon streams, even when salmon are spawning. *—S.M.G. and T.P.Q.*

CONFRONTATIONS can drive subordinate bears away from salmon-laden streams to feed in more peaceful upland meadows.

STEVEN KAZLOWSKI *Peter Arnold, Inc.*

a high grassy meadow, where it began to feast on its catch. When it was distracted by another bear, a mink darted out of the high grass, grabbed part of the carcass and scampered back into the forest.) In Washington State, researchers have compiled a list of more than 50 species of terrestrial vertebrates nourished by salmon carcasses.

A creature does not have to consume the salmon directly to benefit from the ocean's largesse. The insects that colonize carcasses are devoured in turn by wasps, birds and other insectivores, including small mammals such as voles and mice, that eat not only the insects but the carcasses themselves. We have found that densities of insectivorous songbirds can be higher along salmon streams than along waterways that do not support spawning salmon, suggesting that the bird communities respond to the abundance of insects produced by the harvest of salmon carcasses.

In the longer term, the foraging of all these animals, together with leaching by rain and microbial activity, breaks down the carcasses, making the nitrogen, phosphorus and other nutrients available to riparian plants. Plant growth in northern forests is often limited by either nitrogen or phosphorus, and thus the bears' foraging activities may influence growth rates of many plant species in these areas. Along several streams in Alaska, we have calculated that the total amount of nitrogen and phosphorus provided by the carcasses equals or exceeds recommended concentrations of commercial fertilizer for similar plants in northern forests. In some cases, up to 70 percent of the nitrogen in the foliage of streamside shrubs and trees is of salmon origin. Not surprisingly, one study found that growth of Sitka spruce, the dominant streamside tree in the area, was three times greater along salmon streams than along nonsalmon streams. In several studies, researchers correlated the amount of salmon-derived nitrogen or carbon directly with the movements of bears, providing further evidence that their foraging behavior is the mechanism that delivers the salmon nutrients to riparian plants.

Managing the Nutrient Express

AS A RESULT of this new understanding, scientists are redefining how these ecosystems function and thus how they could be managed. Traditionally, it was thought that nutrient flow in such systems moved in one direction only, driven by gravity: nutrients, in the form of leaves, invertebrates and other material, fell from the forest into rivers and creeks, flowed downstream and moved out to the ocean. We know now that they also move in the opposite direction: nutrients, in the form of migrating salmon, travel from the ocean to freshwater and then, carried by foraging bears, to land. Any management action that reduces the number of salmon or bears will affect the nutrient flow and the many creatures that depend on it.

Commercial fishing rates, for example, are generally based on the maximum number of salmon that can be captured without threatening the viability of the population; the "excess" individuals are captured in the fishery. Salmon managers have begun to reconsider these rates to incorporate the

MATT WALLIS *Skagit Valley Herald*

SALMON CARCASSES are being dispersed by helicopter into areas where populations of bears and salmon have been severely reduced or eliminated, such as this drop over the Baker River in Washington State.

needs of other species in the ecosystem. In areas where salmon runs are seriously reduced or wiped out, state agencies are now transporting salmon carcasses—dropping them from helicopters or dispersing them from trucks—to riparian systems as a restoration effort intended to mimic natural processes until salmon runs return to their historical levels. The new knowledge has even sparked entrepreneurial enterprises: one company in Alaska, recognizing the fertilization qualities of the marine-derived nitrogen and phosphorus, is exporting compost soil made of wood chips and salmon carcasses.

We have come a long way since the 1940s in teasing out the ecological ramifications of the fishing bears, and undoubtedly we will learn much more as research continues. What is clear now, however, is that bears and salmon are key components in these ecosystems and that both have been severely depleted or exterminated in many of their historical areas. It remains to be seen whether the greatest challenge lies in understanding the full extent of this relationship or in restoring it where it once flourished. SA

MORE TO EXPLORE

Balancing Natural and Sexual Selection in Sockeye Salmon: Interactions between Body Size, Reproductive Opportunity and Vulnerability to Predation by Bears. T. P. Quinn, A. P. Hendry and G. B. Buck in *Evolutionary Ecology Research,* Vol. 3, pages 917–937; 2001.

Pacific Salmon in Aquatic and Terrestrial Ecosystems. S. M. Gende, R. T. Edwards, M. F. Willson and M. S. Wipfli in *BioScience,* Vol. 52, No. 10, pages 917–928; October 2002.

Nutrients in Salmonid Ecosystems: Sustaining Production and Biodiversity. Edited by John G. Stockner. American Fisheries Society, 2003.

Magnitude and Fate of Salmon-Derived Nutrients and Energy in a Coastal Stream Ecosystem. S. M. Gende, T. P. Quinn, M. F. Willson, R. Heintz and T. M. Scott in *Journal of Freshwater Ecology,* Vol. 19, No. 1, pages 149–160; March 2004.

WORKINGKNOWLEDGE

PERPENDICULAR RECORDING

Going Vertical

Since IBM introduced the hard disk drive 50 years ago, the density of data storage has increased by 65 million times—with much of that rise coming in the past decade. Each data bit in the ferromagnetic layer that coats the tiny disks in computers, video game consoles and iPods has gotten ever smaller—now a mere 30 nanometers across—and closer to its neighbors.

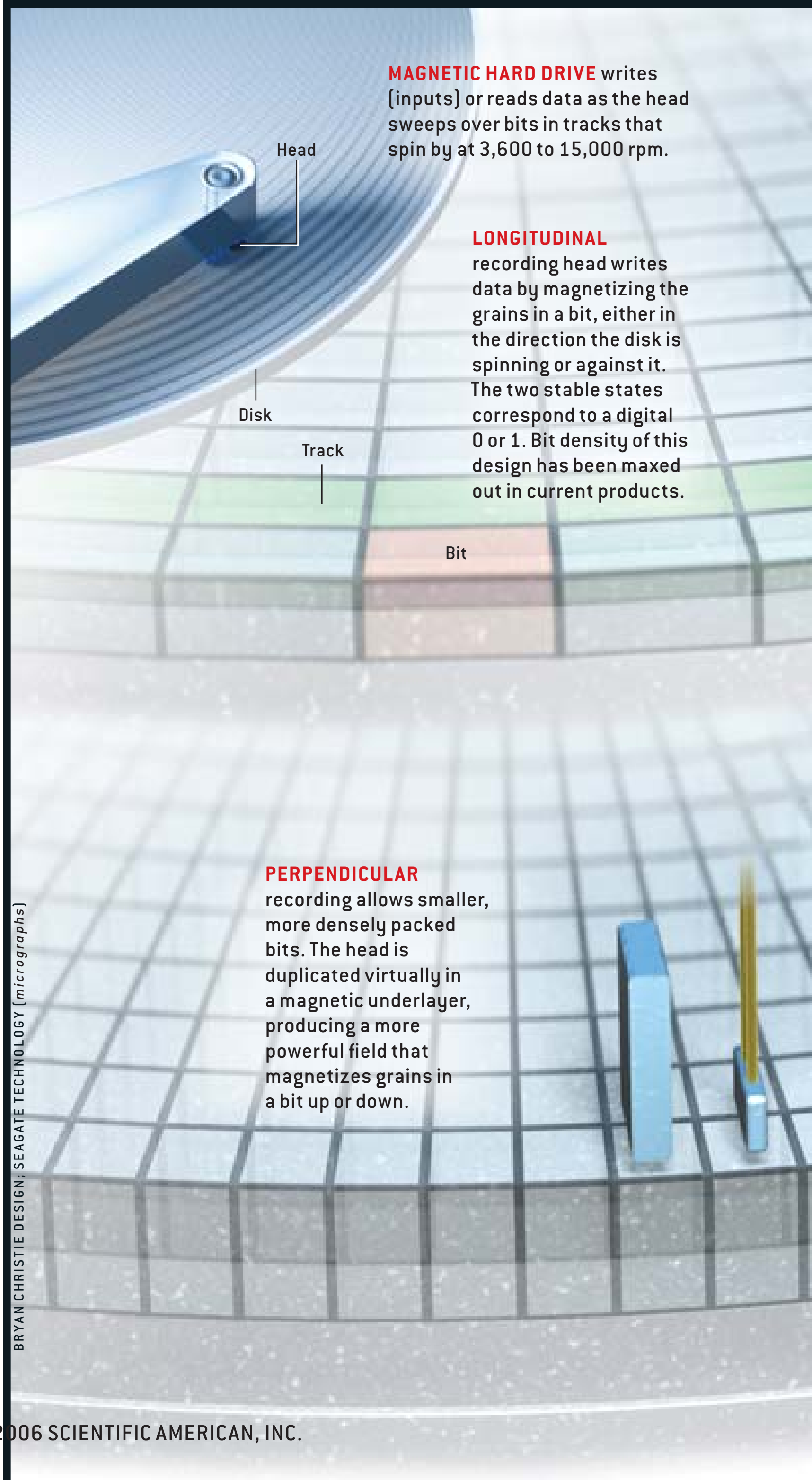
Designers have predicted for years that the miniaturization will hit a fundamental limit determined by the superparamagnetic effect: as the bit becomes more minuscule, the atomic energy holding the magnetic orientation of the bit's atoms in place (which defines a digital state of 0 or 1) becomes so small that ambient thermal energy can destabilize it, thus corrupting the data. That day has finally come. The big drive makers, such as Seagate Technology and Hitachi, are already shipping hard drives with the new architecture.

The scheme—known as perpendicular, or vertical, recording—replaces the venerable longitudinal (horizontal) recording. A bit is magnetized perpendicular to the plane of the disk rather than in line with it, allowing even smaller bits to be packed many times tighter, thereby raising density up to 10-fold [*see illustrations at right*]. Manufacturers say that the new design would allow a video iPod to hold 1,500 hours of video instead of 150.

Longitudinal drives have maxed out at a density of about 100 gigabits per square inch. The first perpendicular drives shipped sported 130 gigabits per square inch. Ironically, notes Mark Kryder, chief technical officer at Seagate in Scotts Valley, Calif., "perpendicular recording will probably only get us to 500 to 700 gigabits per square inch. Then we will hit the superparamagnetic effect for *that* design." The improvement translates into four or five years of new products, he says.

The next iteration, according to Kryder, will most likely be "heat-assisted magnetic recording," in which a laser heats a bit as it is magnetized [*see bottom illustration at far right*]. This approach, he says, will allow a leap to one terabit (1,000 gigabits) per square inch.

And after that? Kryder says, "We are already experimenting with two radically different approaches," which bear names like patterned media and probe storage. To infinity and beyond. —*Mark Fischetti*



MAGNETIC HARD DRIVE writes (inputs) or reads data as the head sweeps over bits in tracks that spin by at 3,600 to 15,000 rpm.

LONGITUDINAL recording head writes data by magnetizing the grains in a bit, either in the direction the disk is spinning or against it. The two stable states correspond to a digital 0 or 1. Bit density of this design has been maxed out in current products.

PERPENDICULAR recording allows smaller, more densely packed bits. The head is duplicated virtually in a magnetic underlayer, producing a more powerful field that magnetizes grains in a bit up or down.

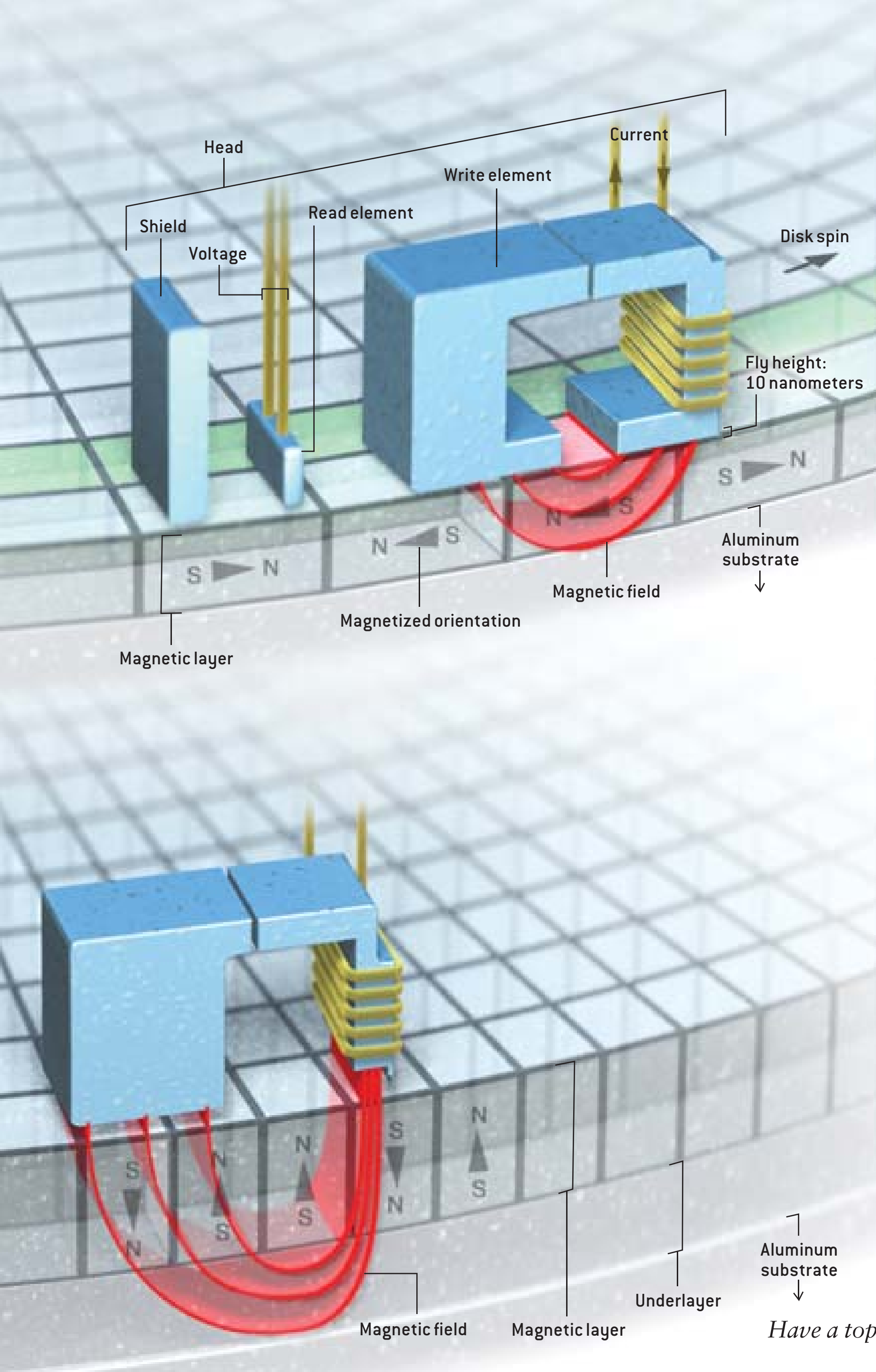
BRYAN CHRISTIE DESIGN; SEAGATE TECHNOLOGY (*micrographs*)

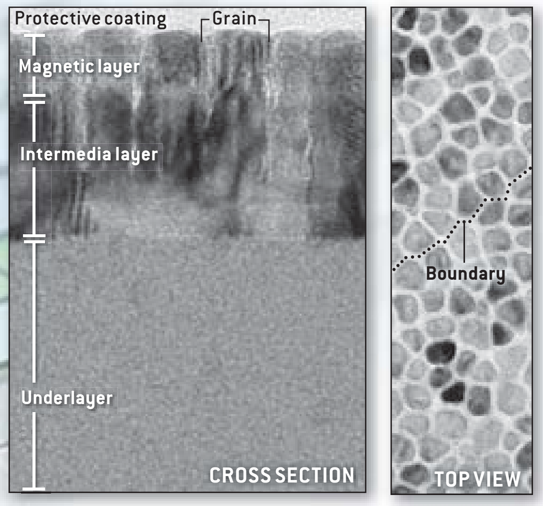
DID YOU KNOW...

TIGHT TRACKS: Narrower, more closely spaced tracks of bits can increase storage density. Cutting-edge drives squeeze in about 150,000 tracks per inch; each track is about 110 nanometers wide, separated by a 60-nanometer gap. The write head matches the track width, yet it can generate small magnetic fields at the edges that can potentially overwrite bits in adjacent tracks. Manufacturers use various head shapes and shields to minimize the risk. Such disruption is more problematic for perpendicular recording, because the common underlayer can couple the field to tracks farther away.

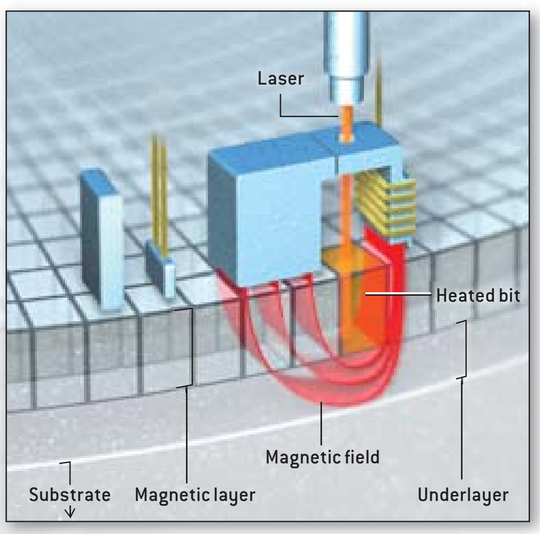
MORE BUT LESS: Because perpendicular drives have an added underlayer, manufacturing costs would in principle be higher than for longitudinal drives. But Mark Kryder of Seagate says the yield (salable parts versus rejects) in production is proving to be greater than for comparable conventional drives, offsetting the rise.

CHALLENGER: Flash memory made with semiconductors—like that for digital camera cards and memory sticks—is also getting denser quickly. In products requiring relatively small capacity, such as the iPod Nano, flash memory is competitive with hard drives; it costs slightly more but is smaller and better resists shock and vibration. Magnetic drives, however, offer lower cost per gigabyte and rewrite data faster and more reliably, attributes critical to most computer and video applications.





ONE BIT comprises 50 to 100 grains of ferromagnetic material magnetized in a roughly rectangular zone (*left*). The boundary is a nonmagnetized zone, a jagged plane along grain lines (*right*).

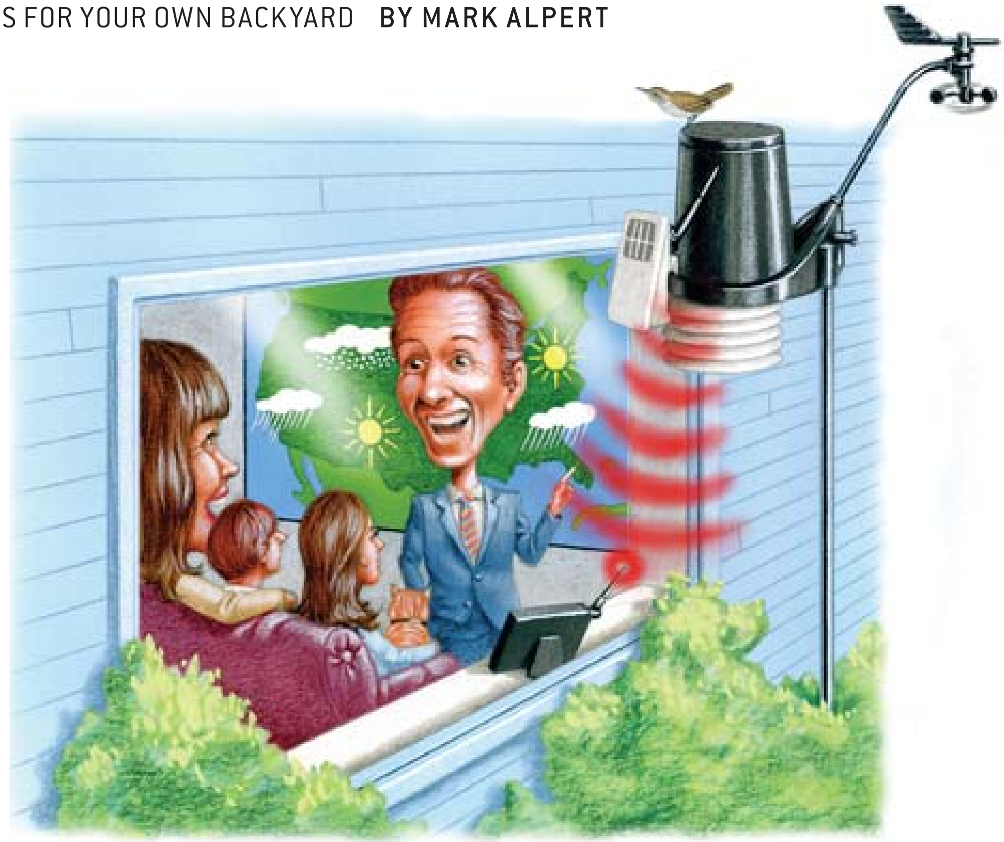


HEAT-ASSISTED recording, still experimental, allows even smaller bits of novel material, which is harder to magnetize. When the head applies a field, a laser heats the bit, momentarily lowering its coercivity so it will magnetize more easily.

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

Weather Gets Personal

NEW DEVICES CAN MAKE FORECASTS FOR YOUR OWN BACKYARD **BY MARK ALPERT**



AMATEUR METEOROLOGISTS can get accurate measurements of temperature, humidity, rainfall, barometric pressure and wind speed from the Vantage Pro2 weather station. The integrated sensor suite wirelessly transmits data to the display console every 2.5 seconds.

JOHN FRASER

My mother loves watching the weather news on television. When I was a kid, she insisted on total silence during the five minutes near the end of the evening broadcast when the meteorologist stood in front of a map of New York City and pointed at the various temperature readings. Her face rapt, she gazed at the radar images of storm clouds sweeping through the tristate area. I asked her once why she loved the weather news so much, and she said with a shrug, "I want to know what to wear tomorrow." But this explanation never satisfied me. Her obsession seemed to go way beyond a simple desire to know whether to put on a coat or carry an umbrella the next morning.

Although I don't watch much television nowadays, I've inherited my mother's fascination with weather. When I read the *New York Times*, the first thing I turn to is the newspaper's weather report, specifically the graph that shows how the recorded and forecasted temperatures compare with the normal highs and lows for the week. So I was intrigued when I learned that Davis Instruments, a company based in Hayward, Calif., had introduced a personal weather station that could wirelessly send me continual updates on the conditions outside, even in the crowded canyons of Manhattan. Although similar devices have become quite common among weather hobbyists, the Vantage Pro2 is the first to combine powerful transmission capabilities with proprietary software than can make accurate forecasts for the microclimate around your home.

The weather station has two parts, an integrated sensor suite that can be installed on your roof or in your backyard, and a display console that can be placed on a desk or mounted on a wall for easy viewing. The sensor suite includes an anemometer—the familiar pinwheel of cups that measures wind speed—as well as a wind vane, a rain collector, a thermometer and a humidity gauge. (The barometer is in the console.) The thermometer and humidity sensor are housed in a slotted plastic chamber that lets air in but shades the instruments from sunlight; for even more accurate measurements, some models come with a fan that circulates air within the chamber. (The Vantage Pro2 costs $495 for the cabled station and $595 for the wireless version; the fan adds another $200.)

As I assembled the sensor suite, I took a special interest in the rain collector, which funnels the drops into a mechanism called a tipping bucket. This ingenious little gadget looks like a seesaw with a scoop at each end. First, the water collects in the upraised scoop until one hundredth of an inch of rain has fallen. Then the full scoop tips downward, spilling its water to the ground and raising the other scoop, which begins to fill with rain. Each time the device seesaws, a magnet hanging from the middle of the lever swings by a switch that records the event, adding another 0.01 inch to the rainfall tally.

VANTAGE PRO2 STATION requires some assembly. The rain collector (*black cylinder in center*) sits atop the slotted plastic chamber holding the thermometer and humidity sensor.

COURTESY OF DAVIS INSTRUMENTS

The power for the sensor suite comes from a small solar panel, which generates enough electricity to charge a capacitor that keeps the device running at night. (A lithium battery provides backup power for weather stations located in especially dark places, such as Alaska during the winter.) The wireless transmitter that sends the weather data to the display console consumes about eight milliwatts. Although this amount doesn't sound like a lot of juice, it is significantly more than that used by previous models, which were constrained by Federal Communications Commission regulations that limit the power of radio signals sent on a single frequency.

The Vantage Pro2 avoids those limits by transmitting spread-spectrum signals that hop from one frequency to another. (Believe it or not, movie star Hedy Lamarr co-invented the technique during World War II as a way to prevent the jamming of radio-guided torpedoes.) As a result, the sensor suite can send data easily through windows and walls; the transmission range is at least 1,000 feet when there is an open line of sight to the display console. I tested the device in the headquarters of *Scientific American* by hauling the sensor suite into an office on the other side of the floor. The signal successfully traversed doors, corridors and all kinds of partitions to arrive at the console sitting on my desk. Then I took the weather station downstairs in the elevator and set it up on the opposite side of Madison Avenue, where it continued to send data to the 12th floor.

In addition to displaying the sensor readings, which are updated every 2.5 seconds, the console's screen can also show graphs of weather conditions over the past 24 hours. A software package called WeatherLink allows users to download the data to their PCs, where they can analyze and graph the information to their hearts' content and post it on the Web if they wish. Best of all, the console can provide custom-tailored forecasts using a software algorithm that predicts the local weather by carefully weighing all the station's measurements.

Less expensive weather stations—some bare-bones machines run as low as $30—typically rely on barometric pressure alone to determine their forecasts, but the Vantage Pro2 also considers wind speed and direction, temperature and humidity, as well as the station's geographic location. (The user inputs latitude and longitude when setting up the device.) For example, if a station on the East Coast detects falling pressure and winds from the north and west, the software will probably predict that a storm is on the way. But the forecast may differ if the winds are coming from the south and southeast. And a station in California will predict rain less often than a station reporting the same conditions elsewhere in the country, because the algorithm takes into account the Mediterranean climate on the West Coast.

Russell Heilig, director of sales for Davis Instruments, says the typical buyer of the Vantage Pro2 is an amateur meteorologist who gets a kick out of beating the weather forecasts on television. But the customers also include farmers who add soil-moisture and leaf-wetness sensors to their stations. Winemakers, in particular, can use the information to determine when to spray for fungi and other pests. Perhaps the most unexpected customers are NASCAR race teams, who need to know the precise temperature, humidity and barometric pressure at the track to maximize the performance of their car engines.

But what about my mother? Would she buy one of these personal weather stations? I called her the other day to get her opinion, but I made the mistake of mentioning the device's price right away. "That's a lot of money!" she exclaimed. "Who's going to buy this? This is for multimillionaires!" I informed her that most of the customers weren't actually millionaires, but she sounded skeptical. "Well, I hope so for the manufacturer's sake," she said. "I feel bad for him already. I like to know the weather, but not to this extent." Then my mother added as an afterthought: "Your father would probably like it." SA

Reality Show

PROBLEMS WITH SCIENCE ON STAGE AND SCREEN **BY ANDREW HODGES**

SCIENCE ON STAGE: FROM *DOCTOR FAUSTUS* TO *COPENHAGEN*
by Kirsten Shepherd-Barr
Princeton University Press, 2006 ($29.95)

MAD, BAD AND DANGEROUS? THE SCIENTIST AND THE CINEMA
by Christopher Frayling
Reaktion Books/University of Chicago Press, 2005 ($35; $19.95, paperbound)

ACTOR: Darling, suppose there are only finitely many prime numbers. Multiply them all together and add one. The resulting number must be a prime, but it can't be on the original list. A contradiction, so—
EUCLID'S GHOST: *No!* The resulting number is either prime, or it has a prime factor, which—
PLAYWRIGHT: Oh, that line didn't *work*. We cut it.
EUCLID'S GHOST: If students get it wrong, they fail. How can you—
PLAYWRIGHT: We're not *students!*
EUCLID'S GHOST: Okay (*exiting*), it's only show business, it's not worth arguing.
ACTOR: I could do my nutty professor act. "*Either* it's prime *or*—aha!—"
PLAYWRIGHT: No, we'll cut that scene. We'll go straight to the revolutionary Reality Show. The audience will vote on who's the fittest, like Darwin. That makes them like the observers in quantum theory, creating parallel worlds—non-Euclidean parallels—
REVIEWER (*entering*): That dialogue is enough to illustrate some dangers in dramatizing science. The book *Science on Stage,* by Kirsten Shepherd-Barr, is a comprehensive survey championing this genre but raising many serious questions. Such plays aren't trying to teach science, but are they doing a service by engaging audiences? Or are they exploiting science to advance an art form? By citing scientific terms, drama may suggest a claim to scientific accuracy and power, while in fact only using those terms impressionistically.
PLAYWRIGHT: But there are so many deeper elements to the artistry: for instance, in contriving textual structure and dramatic form to illustrate content. That transcends mere facts.
REVIEWER: You wouldn't like it if I got the facts wrong about your play. Even in truly thoughtful productions like Stoppard's *Arcadia,* or Frayn's *Copenhagen,* the criterion of "working" dramatically is irrelevant to the exact scientific concepts involved. So is the pleasure of playing with form and content. After all, advertising can likewise be clever, artistic and "work" successfully but yet be completely misleading.
PLAYWRIGHT: What a poor analogy! The point is that the artist must meet the challenge posed by science; that's a central problem of our culture.
REVIEWER: Is it? Maybe it just hurts male writer egos that there's material they don't understand: Is that why they perversely choose abstract subjects that can't be communicated on the stage? Shepherd-Barr draws attention to women writers dealing sensibly with well-explained biological issues.
SHEPHERD-BARR (*entering*): That's putting words in my mouth!
REVIEWER: That's dramatization for

STEVEN SPIELBERG directs a dinosaur during filming of *Jurassic Park: The Lost World* in 1997.

you! But in your favor, science plays do suggest the radical potential of science to audiences. They may be more science fiction than science, but they follow the lead of playwrights throughout history—Marlowe, Goethe, Shaw—in evoking quite new possibilities for humanity. The moral dramas arising from scientific discovery are also entirely genuine. Shepherd-Barr's book will be a valuable resource for teachers and writers relating science and human life.

PLAYWRIGHT: That's not enough. In the postmodern world, one must penetrate and criticize the language and representations of science itself; it is the new mythology of our time.

REVIEWER: You could certainly explore how science *is* a sort of show business. Research has beginnings, crises, resolutions, explosive consequences. Theory is theater: it is setting and symbolic process. Scientists rehearse and present ideas in public speech. But individual achievement in science is rooted in the collective, then absorbed and superseded, in a way that is quite unlike a work of art. Your format demands miraculous individualism, such as in Brenton's play *The Genius,* where a world-shattering bomb is invented in one "eureka" instant by a young student. If that's a metaphor for 20th-century physics, it's a very crude one.

SHEPHERD-BARR: Didn't you notice that I point out how, in *Copenhagen,* Frayn carefully shows that "common culture" of scientific discussion?

REVIEWER: He has a passage where various scientists are named. It doesn't *show* slow and careful learning, problem solving, testing and arguing logically, objectively—

SCREENWRITER (*entering in a rush*): Objective? You must be joking! Science is about getting your name in lights. Citation counting—worse than the Oscars!

REVIEWER: Which leads to another book: *Mad, Bad and Dangerous,* by Christopher Frayling, about the Scientist and the Cinema. Frayling conveys a deep love and knowledge of film, but he readily admits that his medium has produced totally ridiculous science stories. He is less concerned than Shepherd-Barr with literary theory. Instead he uses a wide historical knowledge to suggest how film can help shape events. His centerpiece is a narrative running from *Metropolis* to *Dr. Strangelove,* indicating that the power of film images affected first German rocketry and then, through the von Braun story, American space technology. In discussing the more recent era, it is harder to separate science from countless political and economic issues. But Frayling hopes that film can serve science better, eliminating "mad, bad" images, educating the public in the reality of what science does, and showing scientists as ordinary people. Frayling points approvingly to the image of the mathematician in *Jurassic Park,* but this raises problems: Do scientists want to look ordinary or to look cool? You might say that scientists are *not* ordinary, any more than serious actors or writers are.

SCREENWRITER: You might also say that scientists love acting like starlets on television news and documentaries, blogging, and self-promoting as much as anyone in movies.

PLAYWRIGHT: A synthesis of science and art! That's the future of science performance: a new third culture.

REVIEWER: But there are basic things dramatic performances don't show. Science education and career structure. Rational discussion of everyday but complex health questions. The giant databases and computational methods involved in climatology. The software and hardware that the arts themselves now depend on—

SCREENWRITER: And the global weapons industry, depending on scientific infrastructure.

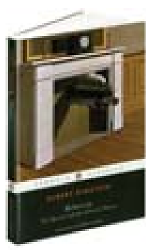
REVIEWER: Both books have a center of gravity in the 1940s and in the mid-Atlantic. I wonder how their discussion looks to today's postdocs—nomadic, transnational, quite likely Asian, who barely remember the cold war. But both spotlight questions that are erased in formal papers and narrow research-group training. They are full of interest for anyone in the business of showing reality. SA

Andrew Hodges, a mathematician at the University of Oxford, is author of Alan Turing: The Enigma, *which served as the basis for Hugh Whitemore's 1986 play,* Breaking the Code.

THE EDITORS **RECOMMEND**

RELATIVITY: THE SPECIAL AND THE GENERAL THEORY

by Albert Einstein, with an introduction by Nigel Calder. Penguin Classics, 2006 ($10)



"The present book is intended," Einstein wrote in 1916, "as far as possible, to give an exact insight into the theory of Relativity to those readers who, from a general scientific and philosophical point of view, are interested in the theory, but who are not conversant with the mathematical apparatus of theoretical physics.... In the interest of clearness, it appeared to me inevitable that I should repeat myself frequently, without paying the slightest attention to the elegance of the presentation. I adhered scrupulously to the precept of that brilliant theoretical physicist L. Boltzmann, according to whom matters of elegance ought to be left to the tailor and to the cobbler."

But it is elegant, in part because of the 1920 translation, by Robert W. Lawson, a British physicist who had polished his German while a prisoner of war in Austria. The introduction, by science writer Nigel Calder, guides the reader through the work section by section, even giving advice on which sections to skip, or at least not to worry about, if you can't "accompany Einstein through the forest of tricky ideas contained in this slim volume." Okay, this book isn't easy—again, in the master's elegant words, it "lays no small claims on the patience and on the power of abstraction of the reader"—but it is well worth the try.

Monumental Error

APPARENTLY, TO PARAPHRASE GERTRUDE STEIN, WHEN YOU COME HERE, THERE'S NO HERE HERE **BY STEVE MIRSKY**

On May 31 the Department of Homeland Security announced their 2006 antiterrorism funding grants to U.S. cities.

The New York Hall of Science. The New York Academy of Sciences. The New York Academy of Medicine. Albert Einstein College of Medicine. Columbia University College of Physicians and Surgeons. Weill Medical College of Cornell University. Mount Sinai School of Medicine. New York University School of Medicine. State University of New York Downstate Medical Center College of Medicine.

New York City had its funding cut by 40 percent from last year.

Nobel laureate Eric Kandel. Nobel laureate Richard Axel. Nobel laureate Roderick MacKinnon. Nobel laureate Harold Varmus. Nobel laureate Paul Greengard. Nobel laureate Joshua Lederberg. Nobel laureate Rosalyn Yalow.

Homeland Security officials said that they determined the amount of each grant based on a formula.

Ellis Island. The Brooklyn Bridge. The Throgs Neck Bridge. Linda Evangelista's neck. The Triborough Bridge. The 59th Street Bridge. Paul Simon. The Verrazano-Narrows Bridge. The George Washington Bridge.

The formula counts the presence of "national monuments or icons." Homeland Security officials determined that New York City had no such national monuments or icons.

The Statue of Liberty. The angel statue in Central Park. Belvedere Castle in Central Park. Central Park. Prospect Park. Van Cortlandt Park. Park Avenue. "Don't even THINK *about parking here." Opera in the Parks. The New York Philharmonic. Derek Jeter. Grand Central Terminal. The Museum of Modern Art. The American Museum of Natural History. The Metropolitan Museum of Art. The Metropolitan Opera. The Mets. The rest of the Yankees. The rest of Linda Evangelista. Yankee Stadium. Madison Square Garden. Oh, okay, the Knicks and Rangers. Okay, even Shea Stadium.*

Coney Island. The bull statue on Wall Street. Wall Street. The Prometheus statue at Rockefeller Center. The skating rink at Rockefeller Center. Rockefeller Center. The Rockefeller University. Columbia University. New York University. The City University of New York. The New School. The Cooper Union for Advancement of Science and Art. The American Academy of Dramatic Arts. The Bronx High School of Science. Scientific American. Stuyvesant High School. Brooklyn Technical High School. Hunter College High School. The Juilliard School. The Fiorello H. LaGuardia High School of Music & Art and Performing Arts. Lincoln Center. Carnegie Hall. The Ziegfeld. David Letterman. Kurt Vonnegut. Ground Zero.

Scientists also develop formulas.

LGA. JFK. The FDR. The IRT. The IND. The BMT. The U.N.

When a formula produces results that are obviously nonsense, scientists examine the data.

The Empire State Building.

Responding to a hurricane of criticism, Homeland Security Secretary Michael Chertoff explained that the Empire State Building was purposefully placed in the "large office building" category rather than the icon category because that designation helps to generate a higher risk score. Here's a flash for the folks gaming their own formula: the Empire State Building is a large office building *and* an icon.

The Chrysler Building. The Flatiron Building. The Woolworth Building. St. Patrick's Cathedral. The New York Botanical Garden. The Brooklyn Botanic Garden. The New York Aquarium. The Queens Zoo. The Prospect Park Zoo. The Central Park Zoo. The Bronx Zoo.

When a formula produces results that are obviously nonsense, scientists may also scrutinize, modify or even discard the formula itself. It's not easy, but somebody has to do it. Seriously, somebody *has* to. SA

ILLUSTRATION BY MATT COLLINS; PHOTOGRAPH BY FRANK VERONSKY

ASK THE EXPERTS

How do space probes navigate large distances with such accuracy?

—T. Storm, Broadbeach, Queensland, Australia

Jeremy Jones, chief of the navigation team for the Cassini Project at the NASA Jet Propulsion Laboratory, replies:

Accurate navigation in space depends on three factors: estimating the probe's current position and velocity (the trajectory), predicting its future trajectory and adjusting the trajectory to achieve the mission objectives.

Estimating the current trajectory requires a model of the forces acting on the probe and measurements of the distance and speed of the probe. Because gravity is dominant, the primary force models are the planetary and satellite ephemerides, which specify the location of all major solar system bodies at any given time. Current planetary ephemerides are accurate to within less than a kilometer for the inner planets and less than 10 kilometers for Jupiter and Saturn. The orbits of Saturn's moons have been determined to within an accuracy of less than one kilometer for Titan and less than 10 kilometers for the other satellites.

Measurements of the range and speed of the probe are used to locate the probe relative to Earth. For all U.S. interplanetary probes, the antennas of the Deep Space Network provide the measurements. These antennas transmit radio signals to a probe, which receives these signals and returns them to the ground station. Navigators compute the difference between the transmitted and received signals to determine a probe's distance and speed. We can fix its radial velocity to an accuracy of 0.05 millimeter per second and its range to three meters. By using measurements spread over days to weeks, we can estimate the location of the probe to an accuracy of better than one kilometer.

The estimation process also results in improvements in the force models. The resulting estimates of the current position and velocity of the probe and the improved force models help to predict the probe's future locations. These predictions are then compared against the mission objectives to determine what corrections are required. Once the predicted attitude error is known, the last step is to determine when to use the probe's propulsion system to make the required correction. A typical mission will schedule a sequence of course adjustments to minimize propellant usage and maximize accuracy. Cassini schedules three corrections between Saturn satellite encounters and achieves a flyby accuracy of a few kilometers or less.

What causes ringing in the ears?

—J. Giambrone, Center Moriches, N.Y.

James B. Snow, Jr., a physician, emeritus professor at the University of Pennsylvania, and former director of the National Institute on Deafness and Other Communication Disorders, offers this explanation:

Tinnitus ("to ring like a bell" in Latin), the medical term for ringing in the ears, is thought to occur when the brain areas involved in hearing spontaneously increase their activity; it is associated with virtually all disorders of the auditory system. It is not limited to ringing; it may also be perceived as whistling, buzzing, humming, hissing, roaring, chirping or other sounds.

The most common form of tinnitus arises from damage to the inner ear, or cochlea, caused by exposure to high volumes. Other causes include drugs such as aspirin, quinine and aminoglycoside antibiotics, cancer chemotherapeutics and other ototoxic agents, and infections and head injuries.

If the inner ear is damaged, input decreases from the cochlea to the auditory centers in the brain stem, such as the dorsal cochlear nucleus. This input loss may lead to increased spontaneous activity in the nucleus neurons, as if some inhibition had been removed. Imaging studies confirm increased neural activity in the auditory cortices of tinnitus sufferers. Their brains also show increased activity in the limbic structures associated with emotional processing. Other symptoms that sometimes appear alongside tinnitus—such as emotional distress, depression and insomnia—may have a common basis in some limbic structure, such as the nucleus accumbens. SA

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

MATT COLLINS