SEXUAL EQUALITY IN 7000 B.C. • MECHANICAL HAND CALCULATOR

SCIENTIFIC AMERICAN

RFID:
The Promise and Perils of Talkative Chips

JANUARY 2004
WWW.SCIAM.COM
$4.95

Loop Quantum Gravity

A Physics Theory Shatters Space and Time

THE CANNIBALISTIC MILKY WAY

SCHIZOPHRENIA

THE RISKS OF AN EARLIER SPRING

january 2004

contents

SCIENTIFIC AMERICAN Volume 290 Number 1

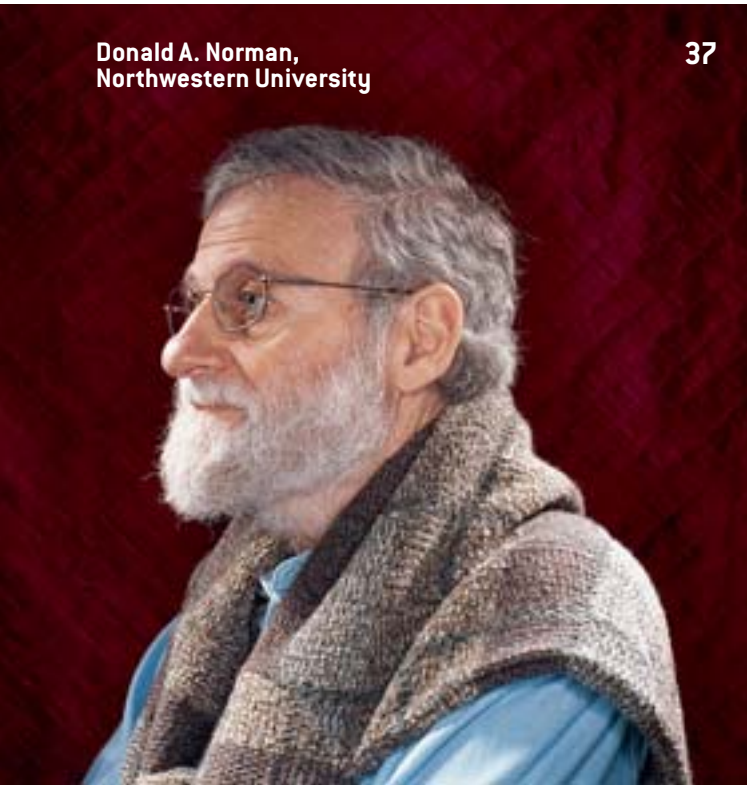
features

92 The Curta calculator

departments

30

104



Donald A. Norman, Northwestern University 37

columns

Cover image by Kenn Brown.

Scientific American (ISSN 0036-8733), published monthly by Scientific American, Inc., 415 Madison Avenue, New York, N.Y. 10017-1111. Periodicals postage paid at New York, N.Y., and at additional mailing offices. Canada Post International Publications Mail (Canadian Distribution) Sales Agreement No. 242764. Canadian BN No. 127387652RT; QST No. Q1015332537. Subscription rates: one year $34.97, Canada $49 USD, International $55 USD. Postmaster: Send address changes to Scientific American, Box 3187, Harlan, Iowa 51537. **Reprints available:** write Reprint Department, Scientific American, Inc., 415 Madison Avenue, New York, N.Y. 10017-1111; (212) 451-8877; fax: (212) 355-0408 or send e-mail to sacust@sciam.com **Subscription inquiries: U.S. and Canada (800) 333-1199; other (515) 247-7631.** Printed in U.S.A.

Can Biologists Be Trusted?

Three years ago gene-splicing biologists at the Australian National University in Canberra were seeking a contraceptive vaccine for mice to reduce the pest population. In the process, they unexpectedly transformed a virus for the rodent disease mousepox into a highly lethal pathogen that kills 60 percent of infected mice, even those that are normally immune. American researchers continuing that line of work recently reported at a conference in Geneva that they had produced a similar virus that is nearly 100 percent fatal.

The rationale for such experiments is that they might assist the authorities in preparing for bioterror attacks. The counterargument is that they might aid bioterrorists. (Fortunately, the changes that make these pox viruses so harmful also seem to render them noncontagious.) Concerns are not restricted to projects with obvious relevance to germ warfare; the broader worry is that even innocuous research might be misused. The policy question becomes: Is biology too dangerous to be entrusted to biologists?

Ever since the Manhattan Project, national security restrictions have been a fact of life for physicists. The government has, for the most part, allowed biologists to police themselves. In 1975, for instance, fears surrounding genetic engineering prompted researchers to agree that any such experiments would need to be approved by qualified Recombinant DNA Advisory Committees (RACs).

Fears of bioterrorism call for a similar response, and the biology community has already taken action. Last October the National Research Council (NRC) issued recommendations for overseeing unclassified experiments that might advance terrorists' work on biological weapons. The new guidelines recommend a multitiered regulatory approach. The responsibilities of the RACs would expand to cover all types of plausibly risky experiments, such as those aimed at disabling vaccines, conferring resistance to antibiotics, enhancing virulence, or turning cells and proteins into weapons. A new advisory board within the Department of Health and Human Services would offer direction to the RACs while encouraging dialogue between scientists and security specialists. The report also urges the establishment of an International Forum on Biosecurity to weave a consistent net of biotech safeguards in all countries.

Many researchers and defense experts have hailed the NRC proposals as sensibly balancing security and scientific freedom. But John H. Marburger, science adviser to President George W. Bush, was quoted in the *New York Times* as saying that the administration had not yet taken a position on the proposals and might ask for more restrictions.

It is only reasonable to ask whether the proposals do enough to guarantee security. Additional restrictions that might encumber inquiry and the free exchange of data pose their own dangers, however. For example, as the NRC report notes, the White House has sometimes shown enthusiasm for restricting access to information by categorizing it as "sensitive but unclassified." Such a vague label applied to research could harm national security by crippling scientific creativity.

Certain curbs on biomedical research are prudent and appropriate and can be adopted without sacrificing liberties essential to progress. Scientists themselves, in partnership with government, are best qualified to set those limits. The NRC plan for biology should be given a chance to work as it is now.

THE EDITORS editors@sciam.com

TEK IMAGE *Photo Researchers, Inc.*

How to Contact Us

EDITORIAL

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

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

SUBSCRIPTIONS

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

REPRINTS

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

PERMISSIONS

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

ADVERTISING

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

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

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

San Francisco
415-403-9030
fax: 415-403-9033

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

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

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

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

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

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

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

Sweden
Publicitas Nordic AB
+46-8-442-7050
fax: +46-8-442-7059

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

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

India
Yogesh Rao
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

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

On the Web

WWW.SCIENTIFICAMERICAN.COM

FEATURED THIS MONTH

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

GEOCHEMISTS TRACE THE ICEMAN'S TRAVELS

They say dead men tell no tales. If that was ever true, it is certainly not so in our scientific era. Case in point: Ötzi, the 5,000-year-old "Iceman" mummy discovered in 1991 by two hikers high in the Alps along the Austrian-Italian border. Affectionately nicknamed for the Ötzal region in which he was found, Ötzi has been subjected to waves of tests in an attempt to reconstruct his life and death. Now researchers have amassed evidence suggesting that the Iceman, believed to be in his mid-40s when he died, may have spent his entire life in present-day Italy, within about 60 kilometers of where he was found.

Climate Change Linked to Improved Vintages

Long hours and a lot of work go into producing a winning wine. But recent climate changes may have lent vintners a helping hand. Scientists report that most of the world's most renowned wine regions have experienced warming during their growing seasons that is associated with better overall vintages and more consistency from year to year.

Ask the Experts

How can deleted computer files be recovered at a later date?

Clay Shields, a professor of computer science at Georgetown University, explains.

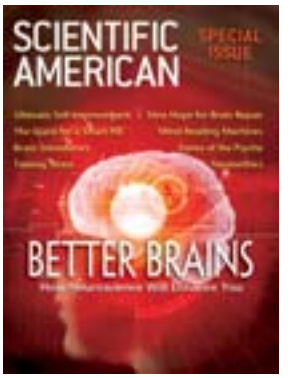
SOUTH TYROL MUSEUM OF ARCHAEOLOGY

SCIENTIFIC AMERICAN®
Established 1845

Letters

EDITORS@SCIAM.COM

EVERYBODY HAS THE RIGHT to change his or her mind. But what if the subject of change is not the mind but the brain? This thought, explored in the September 2003 single-topic issue "Better Brains," stirred a gale-force gust of letters from readers. Some were thrilled about the new possibilities for treating neurological diseases. But the moral gray area of gray matter alteration also inspired some consternation and even urgency. Several readers questioned the true impetus behind the lucrative business of brain improvement. Others raised concerns about the physiological and ethical hazards of trying to improve brains that are not actually "broken." These ideas and more fill the following pages.



ISSUES OF ENHANCEMENT

The essay "Is Better Best?" by Arthur L. Caplan, neglects to mention the influence of creativity on thought. Brain engineering may create more effective thinkers, but it has yet to be proved that the brain can be stimulated to create new ideas. That is, we may be able to help a potential Shakespeare, Einstein or da Vinci produce his ideas more effectively, but we cannot create such thinkers, with their novel ideas, "from scratch." This hurdle may be found in the quest for artificial intelligence as well. I believe that humanity has little to fear from brain engineering or artificial intelligence. Although biological enhancements may enrich our existence, diversity itself will be left to more old-fashioned methods: opportunity, coincidence and necessity.

Karmen Lee Franklin
Arvada, Colo.

Caplan notes that the essence of humanness is to "try to improve the world and oneself." In doing so, he has asserted a convenient definition of human life in one sentence, without defending his definition. Yet even if he were right, might not the manner in which we seek improvement also affect our humanity? If we turn ourselves into souped-up machines in the quest for perfection, doesn't this reveal something about our humanness? The real harm of enhancement is that it can undermine our most basic and stable ideas about identity, personality, accomplishment, virtue and dignity. Too much for a brief letter, but certainly enough to preclude a carefree rush into enhancement.

Daniel Tobey
via e-mail

Though arguing strongly in favor of brain improvement, Caplan never explains what he means by enhancing, optimizing or improving our brains—and I fear the consequences of such "improvements." To understand how that could be problematic, suppose someone wanted to do better in business and eliminated inhibitions from his brain to make himself more ruthless.

Humankind has a long and tragic history of attempted self-improvement. Chinese women bound their feet to improve their beauty; women of the former German Democratic Republic sought athletic prowess with massive doses of testosterone. Eugenics offered to better the human race, and Hitler attempted to apply its teachings. These days silicone and various dopants are used to alter appearance and athletic abilities. As a professor of physiology, I have seen nervous students who took tranquilizers to improve their performance but then became too incoherent to function.

Caplan writes that coercion will not be needed to induce people who want to "optimize" their brains, because market-driven societies encourage improvement. When baldness, impotence, facial wrinkles and cellulite are the (market-driven) scourges of civilization, whereas malar-

ia, cholera and malnutrition are largely ignored, we are a long way from understanding real improvement.

H. Peter Clamann
University of Bern
Switzerland

PLANNING A HEAD

I appreciate your commencing your special issue with reference to my views on the upcoming "marriage of the biologic and the cybernetic" ["A Vote for Neuroethics," Perspectives], despite your skepticism. I will note that a primary source of our different outlooks on the prospects for brain reverse engineering is that we are considering different time frames.

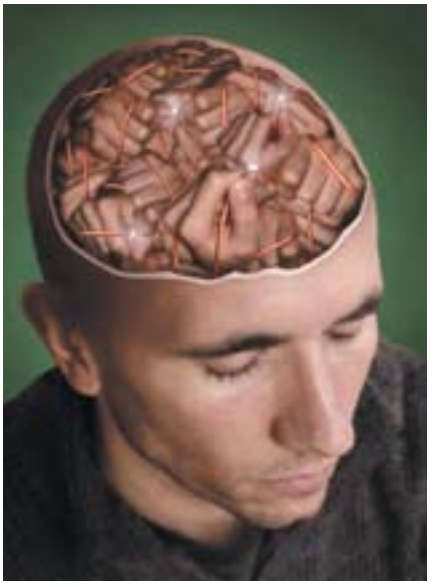
The special issue describes well some of the neuroscience advances now in development, innovations that we can expect to benefit from during this coming decade. We need to ask: What happens after that?

Progress will not only continue, but its pace will continue to accelerate. The reason for the acceleration is that each stage of progress in a given technology creates more powerful tools to enable the next stage.

Consider, for instance, that spatial and temporal resolution of brain-scanning technologies is clearly improving at an exponential pace. One of many examples is the in vivo scanning system being developed at the University of Pennsylvania, which is designed to resolve individual neurons in a cluster of up to 1,000 simultaneous cells with submillisecond temporal resolution, a dramatic improvement over current systems.

According to my models, we are doubling the paradigm shift rate (the rate of technical change) approximately every decade, so we can reasonably anticipate a dozen generations of technology over the next three decades. Scientists are trained to be conservative in their outlook and expectations, which translates into an understandable reluctance to think beyond the next step of capability. When a generation of technology was longer than a human generation, this orientation served society's needs well enough. With the rapid acceleration of progress, however, a short-term look ahead is no longer sufficient. The public has a legitimate interest in informed opinion that looks forward to 20 to 30 years from now.

When we consider the implications of multiple generations of technology, the availability over the next several decades of enormous increases in the capacity of our computational and communication tools, the advent of molecular nanotechnology, and far greater insight into the principles of operation of the human brain, I believe that our perspectives will converge.



NEW TECHNOLOGIES could lend a hand to brain function and repair. But should they?

Ray Kurzweil
Kurzweil Technologies
Wellesley Hills, Mass.

PHARMACEUTICAL COSTS

With regard to your entire September issue, and in particular the article "Diagnosing Disorders," by Steven E. Hyman, I am surprised that you did not mention the extra costs required to subsidize the neurological treatments discussed. For example, in a table indicating the percentage of individuals suffering psychiatric trouble, the author suggests that roughly 20 percent of individuals contend at any one time with a serious affliction. Assume that medical costs for each one amount to $1,000 a year (in reality, the figure would be much higher). With some 20 percent of 300 million people in the U.S. alone to choose from, that means a total of at least $60 billion in potentially new medical care.

This vast incentive might explain why drug companies fund this research. Once the research is legitimized, the health care industry extracts the costs back from society, to the current tune of 13 percent of the GDP. It is relatively easy (and profitable) to germinate a new crop of "illnesses." It is not so easy (and hardly as profitable) to ascertain the true reasons behind today's social dissatisfactions.

Richard Borbely
Simi Valley, Calif.

LAST HURRAH

This issue has the finest, most even-handed and in-depth writing and editing I have read on a series of very subtle topics. So often, even in professional journals, articles describe only a part of issues—neuronogenesis, as if the entire brain and spinal cord can regrow; or cognitive techniques for correcting dyslexia, aphasia, "left brain" thinking and so forth. Your issue includes necessary caveats about controls, "off-label" drug uses, multiple points of view, preliminary results and more, while recognizing the thrilling potentials and advances in brain research. My congratulations on excellent conceptualizations, elegant writing and editing, and fascinating reading.

Sidney Werkman
Department of Psychiatry
Georgetown University School of Medicine

ERRATUM In "Data Points" [News Scan], the distance between Jupiter and the sun should have read 778 million kilometers, not 778 billion kilometers. The distance between the new planet and HD70642 should have read 494 million kilometers, not 494 billion kilometers.

MELISSA SZALKOWSKI

50, 100 & 150 Years Ago

FROM SCIENTIFIC AMERICAN

Stone Age Treasure ▪ Air Age Optimism ▪ Petrochemical Light

JANUARY 1954

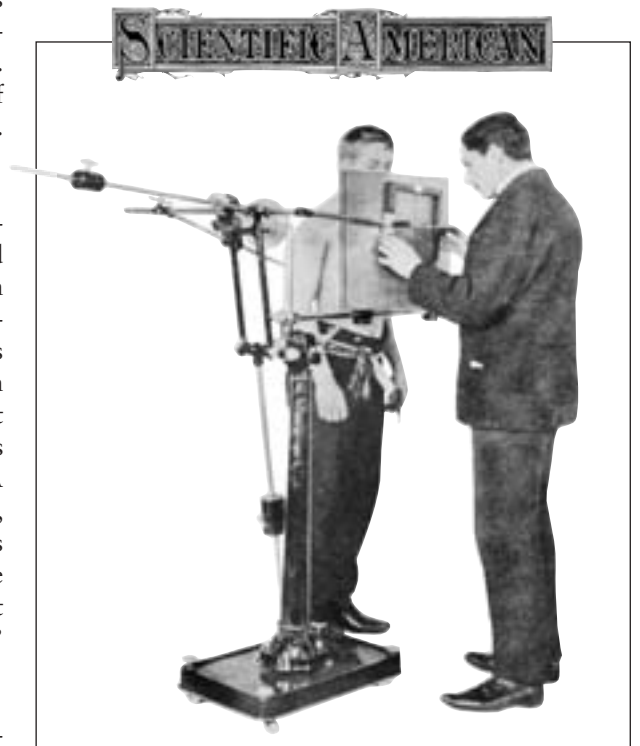
OLDUVAI GORGE—"This canyon in Tanganyika Territory in East Africa has yielded the most complete sequence of early human tools ever discovered, and along with these a great wealth of remains of the now extinct animals that Stone Age man hunted. In the successive deposits in the Gorge is written some 400,000 years of man's cultural history—from the Middle Pleistocene to about 15,000 years ago. They cover almost the whole span of man's hand-axe phase, known to archaeologists as the Chelles-Acheul culture. Olduvai Gorge, so rich in the relics of human settlements, seems an ideal place to look for the remains of hand-axe man himself. The conditions for fossilization of his bones there were excellent. —L.S.B. Leakey"

"LINEAR B" CRACKED—"An important ancient script which had defied translation for more than half a century has just been deciphered. The writing, known as 'Minoan Linear B,' was in use in the Cretan maritime empire that flourished more than 2,500 years ago, long before Homer's time. A British architect, Michael Ventris, working on the problem in his spare time, solved the puzzle. The writing was found in 1896 at Cnossos in Crete on clay tablets."

JANUARY 1904

THE AIRPLANE AGE—"The successful flight of a motor-driven aeroplane built by the brothers Orville and Wilbur Wright is an event of supreme importance in the history of aeronautics. This feat marks the commencement of an epoch in the history of the aeroplane; for now that an aeroplane has been built that can fly, the work of gathering experimental data will proceed with a rapidity which was impossible when aeroplane flight, at least on a full-sized scale, had never gone beyond the theoretical stage."

THE BIRDS AND THE SEEDS—"There may seem little in the migration of the summer birds to furnish data for scientific deductions; but the modern student of our native birds sees in these annual flights material for reflection and observation of the greatest importance. The problem of weed destruction is, for instance, intimately wrapped up in the migratory habits of the millions of our summer birds. Many of our most noxious garden and field weeds produce in a single season as many as one hundred thousand seeds. There is only one effective agency that keeps in check these prolific weeds. When the seeds of the weeds ripen in the late summer and fall, the millions of migratory birds begin their journey southward, devouring the weed seeds. We have always supposed that the birds started southward as soon as the chill of autumn approached. But they are not weather prophets at all, but simply hungry little creatures following in the footsteps of ripening seeds."



X-RAYS: Apparatus for pinpointing internal organs, 1904

X-RAY ACCURACY—"The orthodiagraph, just brought out by the Berlin Allgemeine Elektricitäts-Gesellschaft, is a Röntgen apparatus allowing of the true image of any object being obtained in any desired position of the drawing plane. The luminous screen, which also carries the drawing stylus, is rigidly connected with the Röntgen bulb by a U-shaped frame made up of jointed sections [*see photograph*]. When a drawing is to be made directly on the body, the bristol-board is removed from the drawing frame, and a dermatograph stylus should be inserted into the drawing stylus instead of a pencil."

JANUARY 1854

PARAFFIN CANDLES—"If all the reports which have come to us recently from abroad, with respect to new discoveries in making candles, are true, all our whaling ships will soon be laid up in port or converted into *coal grunters*. In a quarry about twelve miles to the west of Edinburgh, Scotland, rests a thick bed of dark-colored shale. A few years ago some one thought of distilling shales. Some of them are exceedingly rich in an inflammable substance, resolvable into gas and tar, and which has received the name of parafine. Of this substance, beautiful candles are made, in no degree inferior to those of wax. In Ireland, peat is thrown into huge retorts and there distilled."

BIOINFORMATICS

NECTAR for Your Health

REVAMPING U.S. MEDICAL RESEARCH MEANS UNIFYING DATA **BY DANIEL G. DUPONT**

On September 30, 2003, the director of the National Institutes of Health announced a long-awaited restructuring of government-funded medical research. The Research Roadmap, Elias Zerhouni stated, would position the NIH—by far the largest source of money for medical investigators—to take better advantage of recent advances, such as the mapping of the human genome, and to overcome barriers that limit researchers' ability to access and share data.

The plan calls for "new pathways to discovery," greater interdisciplinary research through new collaborations, and a "reengineering" of clinical research, according to the NIH. A key component largely lost in the flurry of promise and proposal outlined last fall was an information network initiative—critical in making the road map complete and in revolutionizing the methods by which medical data are collected, stored and shared. The effort, called the National Electronic Clinical Trials and Research Network, or NECTAR, will unite vast and disparate databases into one massive pool—and ultimately help to turn research data into therapies more effectively.

DIGITIZING paper records would be essential for creating a planned giant data pool.

ER PRODUCTIONS *Corbis*

The way things work today is considered wildly inefficient, notes Daniel R. Masys, director of biomedical informatics at the University of California at San Diego. "As an institution, or perhaps as a drug company, you have a scientific question in mind, consult with biostatisticians and determine the number of people needed and specifications to answer the question, write the forms for the questions and data, hire people to type the data into databases, and then at the end you publish a paper," Masys explains. The primary data, however, remain the property of the institution. "You keep your own data, and the next trial, you do it all over again," he says.

NECTAR will change all that, states Stephen I. Katz, director of the National Institute of Arthritis and Musculoskeletal and Skin Diseases and an important figure in the

road map's development. It will consolidate data in a user-friendly, Internet-based system. In this way, it would eliminate "the need to develop an entirely new infrastructure for every new major study," Katz remarks.

The NIH has already begun reviewing existing technologies, databases and networks to see what can be part of NECTAR, says Amy Patterson, director of the NIH Office of Biotechnology Activities, who is helping to steer NECTAR. Among the networks to be studied are some, such as those of the Veterans Administration, that deal primarily with health care information and others that contain clinical data. Over the next two years the NIH will solicit the input of biomedical researchers and information technology experts in an attempt to construct a handful of pilot projects that will extend existing networks and allow the concept of a global research network to be tested in miniature.

Along the way it will develop software to standardize and simplify the authoring of study protocols, and it will collaborate with agencies such as the Food and Drug Administration to ensure that medical events—in particular, adverse reactions—are described in uniform fashion. Uniformity is essential: if researchers do not speak the same language (and today they do not), then their data cannot be pooled.

Another imperative: less paper. "Eighty percent of the battle is getting America using digital medical records," Masys says. Otherwise, paper records would have to be converted to digital or left out of the database.

The NIH also plans to ensure the privacy of medical information by complying with the requirements for "electronic transmission and privacy of health data" laid out in the Health Insurance Portability and Accountability Act of 1996, Patterson says. Although the NIH does not have to comply with the act, it has "for decades protected the privacy of patient data because of other federal regulations," she adds, noting that many of the participating institutions must comply with the act.

In five years, a broad prototype effort is supposed to be up and running; in another five, the NIH expects to have in place "the fabric of a national network of networks," Patterson states. NECTAR is a monumental undertaking, and she does not expect it to be easy. For that reason, she explains, the NIH will involve the institutions that will be part of the network as the plan is developed. That part seems to have gotten off rather slowly; two months after the road map was announced, Masys said he and others in the research and informatics communities were still largely in the dark. But he applauds the NIH for its vision, which he calls "exactly the right thing to do on a national scale."

Patterson promises that NECTAR will soon pick up steam with the issuance of solicitations for pilot projects. Judging from the feedback received already, she believes a "rational and highly communicative" strategy can lead to the forging of the necessary partnerships. "There's a real hunger out there to have some uniformity and some collaboration among research centers," she observes. Satisfying that hunger, the NIH hopes, will more quickly transform research findings into drugs and therapy that people can use.

Daniel G. Dupont edits the online news service InsideDefense.com.

NECTAR: SWEET FOR EVERYONE

The National Institutes of Health is trying to get public and private institutions and pharmaceutical companies on board with its plan to develop a national database network called NECTAR. Everyone has good reason to go along. "Big pharma," for one, would benefit by having access to more data. Because the firms must submit their research data to obtain drug approval, the government has tremendous leverage in enforcing common standards and creating the data pool that will be at the heart of NECTAR.

BIOTERROR

Uncertain Threat

DOES SMALLPOX REALLY SPREAD THAT EASILY? BY GUNJAN SINHA

"**Biological terrorism** is our future, and smallpox is a serious threat," insists Ken Alibek, who headed the former Soviet Union's biological weapons program. Now vice chairman of Advanced Biosystems, based in Alexandria, Va., Alibek was one of 200 epidemiologists and tropical disease experts who gathered in Geneva last October to discuss how nations should prepare for an outbreak. The U.S. has already outlined its plan—a voluntary regimen that aims to vaccinate a total of 10.5 million people in phases.

Some scientists, however, see little data to

support such widespread vaccination. The plan is partly based on mock scenarios and mathematical models that attempt to predict the magnitude of an outbreak. One major problem is that they must use data on smallpox transmission gathered from pre-1977 Africa, where the last smallpox case occurred. The virus might behave completely differently in today's unvaccinated cosmopolitan societies. And all models rely on assumptions that by their nature are inaccurate.

The most grave outbreak scenario is "Dark Winter," to which U.S. Secretary of Defense Donald H. Rumsfeld has referred a number of times. It predicts that simultaneous attacks in three shopping malls could balloon to as many as one million dead and three million infected.

OVERBLOWN FEAR? Smallpox—here, from a 1973 Bangladesh case—may not spread that easily.

But many scientists find the scenario too extreme. What is most contentious is the infection rate. Dark Winter assumes that each infected person will transmit the virus to 10 others and even to descendants for several generations. This is not, however, what epidemiologists have observed in the field. Rarely was smallpox transmitted to more than two or three people, if at all, says J. Michael Lane, former director of the smallpox eradication program at the Centers for Disease Control and Prevention, and most were infected by prolonged exposure. What is more, the virus is not transmissible until physical symptoms appear. By that time, Lane states, the person usually feels "so awful" that they are bedridden. And even though the virus may not behave the same way today, Dark Winter assumes that the sick are not effectively isolated, which is "totally unrealistic," he adds.

So instead of vaccinating millions, Lane would prefer to vaccinate a core group of first responders—around 40,000 people—and then to vaccinate only people who come into contact with an infected person (the vaccine is also effective for up to four days after infection). His plan more closely reflects what has actually transpired in terms of vaccination numbers [see "Spotty Defense," News Scan; Scientific American, May 2003].

Proponents of mass vaccination also cite a few exceptional cases in which smallpox spread easily. In 1970 a young engineer returned to his home in Meschede, Germany, after spending some time in Pakistan. Soon after, he checked himself into a hospital with flulike symptoms. Doctors quickly diagnosed him with smallpox, but during his stay 19 other people also became ill. The most bizarre case was the infection of a person who had briefly walked into the hospital lobby, discovered he was lost and left. The sick engineer had a cough, a highly unusual symptom but one that nonetheless made the virus highly transmissible. No one knows whether the smallpox strain was unusually hardy or the patients uncharacteristically weak.

Another outbreak occurred in 1963 when a young man, who had spent some time in India, came down with smallpox on returning to his home in Poland. By the time health authorities figured out he had smallpox several weeks later, 99 other people became ill. To contain the outbreak, authorities vaccinated eight million people, even though the population had been vaccinated as infants. (The illness tends to be less severe in vaccinated people, however.) Scary as they are, these stories are isolated cases and clearly do not represent how the virus behaved in the majority of outbreaks. "Surveillance and containment strategies were key components of the smallpox eradication program," Lane notes. "We must not lose sight of that."

But supporters of more widespread vaccination are sticking to their guns. Although everyone agrees that an attack is unlikely, any outbreak, however small, would be "economically and psychologically devastating," Alibek states. In his view, widespread vaccination would help preempt the chaos likely to follow. (His company, Advanced Biosystems, conducts research on therapeutics to counter biological weapons.) Countries hoping to defend against a smallpox attack, it seems, will have to strike the balance between science and fear.

Gunjan Sinha is based in Frankfurt, Germany.

PUTTING SMALLPOX INTO THE WIND

Some observers argue that the smallpox virus could be engineered to be more deadly, and one anecdote suggests the pathogen can easily be aerosolized. In 1971 a fisheries research ship was floating a little too close to Vozrozhdeniye Island in the Aral Sea, where the Soviets tested biological weapons. A scientist, the only person on deck, came down with smallpox a few days after returning to shore. Although the nature of how the woman was infected is controversial (the Russian government has apparently been cagey about the matter), some scientists are convinced that the Soviets unleashed aerosolized virus that was hardy enough to float nine miles downwind.

REUTERS NEWMEDIA, INC. *Corbis*

ACOUSTICS

String Theory

A WEAK SUN MAY HAVE SWEETENED THE STRADIVARIUS **BY LAURA WRIGHT**

STRADIVARIUS VIOLINS may have benefited from colder than average temperatures.

Myriad proposals have surfaced in the past several centuries to explain how Antonio Stradivari imbued his now priceless wares with transcendental sound. Some have suggested that Stradivari used beams from ancient cathedrals; others argued that he gave his wood a good urine soaking. The latest theory proposes that the craftsman should thank the sun's rays—or lack thereof.

Stradivari could not have known that his lifetime coincided almost exactly with the Maunder Minimum—the 70-year period (from 1645 to 1715) of reduced solar activity that contributed to colder temperatures throughout western Europe during what is called the Little Ice Age.

Stradivari and the Maunder Minimum "began life a year apart," says Lloyd H. Burckle, a paleobiologist at the Lamont-Doherty Earth Observatory of Columbia University. "Which means that during his later years, the golden period, he had to build violins out of wood that grew during the Maunder Minimum." The reduced radiation from the sun would have slowed the movement of the warm air over the Atlantic Ocean to western Europe, setting off a decades-long period of colder, drier climate. Such conditions would have been especially harsh for a tree adapted to temperate climes, such as the Norway spruce, Stradivari's favorite for making soundboards. The result was slower, more even tree growth, which would yield a stronger and denser wood—positive attributes for violin crafting.

A changing climate probably didn't act alone in the Alpine forest of northern Italy, where Stradivari is said to have harvested trunks, Burckle notes. But when coupled with a unique amalgam of environmental factors—such as the regional geology, soil chemistry and moisture and slope and direction of the mountainside on which chosen trees grew—the altered climate becomes a more viable player. Burckle presented his hypothesis to Henri Grissino-Mayer, a tree-ring scientist from the University of Tennessee who has studied the influence of the Maunder Minimum on trees in western Europe, and the pair published the idea in the summer 2003 issue of the journal *Dendrochronologia*.

If indeed the Maunder Minimum led to the superlative sounds of the Stradivarius instruments, then it might appear that future violins would never produce similarly dulcet tones. "If you say it's the climate and it will never return, that makes it all seem hopeless," remarks Joseph Nagyvary, a chemist and violin maker at Texas A&M University. But having studied for three decades how various wood treatments can enhance the sound of instruments, Nagyvary thinks Stradivarius-like quality is achievable without an ice age: "We can now make the sound just as good."

Laura Wright is based in New York City.

MIGUEL VILLAGRAN *EPA*

NEUROLOGY

Aching Atrophy

MORE THAN UNPLEASANT, CHRONIC PAIN SHRINKS THE BRAIN **BY LISA MELTON**

An occasional headache is a nuisance, but severe, unrelenting pain can blight one's existence. Scientists have now learned that chronic pain, which often leads to anxiety and depression, can also effect neurological changes. It can shrink the brain and impair one of the most valuable mental functions: the ability to make good decisions.

Pain is a defense system that indicates when something is wrong, comments Marshall Devor, a pioneer in pain research at the Hebrew University of Jerusalem. "When

there is a persistent tissue disorder or there has been injury to the nerves, it's like an alarm that is broken. Pain becomes a disease in its own right," Devor points out.

Pain signals originate at the site of injury but soon lay siege to the entire nervous system. When pain is unremitting, dramatic changes follow: spinal cord neurons become hypersensitive and start firing in response to weak stimuli. This hyperexcitability ratchets up all pain responses, which explains why people with diseases such as arthritis, cancer and diabetes or with nerve trauma caused by surgery sometimes experience widespread pain from even the lightest touch.

"Pain always travels to the brain" and could cause damage, surmises A. Vania Apkarian, a bioelectrical engineer and physiologist at Northwestern University. To test his hypothesis, Apkarian turned to magnetic imaging. Zooming in on the brain chemical *N*-acetyl aspartate—the amount of which correlates with the density of neurons—he identified a striking difference in the prefrontal cortex. Pain was apparently triggering brain atrophy there.

Apkarian compared the overall volume and regional gray matter density in patients who had chronic back pain with those features in nonsuffering control subjects. The preliminary results were revealing: the average atrophy was greater in those with lower back pain than was normal. "The difference is highly significant," he states.

Because the prefrontal cortex is crucial for emotional decision making, Apkarian wondered if constant pain might be clouding people's judgment. He asked 26 people who had suffered lower back pain for more than one year and 29 normal volunteers to play a gambling card game called the Iowa Gambling Task. The test was originally developed by neuroscientist Antonio R. Damasio of the University of Iowa and his colleagues to study decision making in risky, emotionally laden situations.

The game involves selecting cards from decks with different potential cash payouts and penalties. Normal subjects

PUTTING UP WITH PAIN

Percent of Americans reporting chronic or recurrent pain in the past year: 57

Percent reporting pain "all the time": 40

Percent reporting constant pain in the U.K.: 14

European average: 19

Most common type: back pain

Estimated U.K. health costs related to pain syndromes, 1998: £1.6 billion

Cost when informal care and productivity losses are factored in: £10.7 billion

Annual cost to U.S. employers, estimated as lost productivity: $61.2 billion

SOURCES: Research!America 2003 telephone survey of 1,004 adults; Pain Society of the U.K.; European Federation of IASP (International Association for the Study of Pain) Chapters; Journal of the American Medical Association, *November 12, 2003.*

STRAIGHT TO THE BRAIN: Chronic pain, such as that from a bad back, not only saps the joy of living but also shrivels neurons.

COMPUTER IMAGE BY ROBERT S. WINTER *Photo Researchers, Inc.*

learned to optimize their choices, tending to select cards from decks that made them money. But participants with a pain history tended to select cards randomly: they seemed to lack a master plan, which resulted in 40 percent fewer good choices compared with those made by nonsufferers. What is more, the amount of suffering correlated with how badly they played. "Chronic pain is driving these people to make poor judgments," concludes Apkarian, who presented these findings at a Novartis Foundation symposium last fall in Tsukuba, Japan.

Yet other cognitive abilities remained intact. "None of these patients are dramatically impaired," says Apkarian, who, to avoid confounding factors, excluded from the study people with high depression or anxiety. "This study raises the question of whether these people are making appropriate decisions in everyday life," speculates Apkarian, who found similar effects with sufferers of chronic complex regional pain syndrome, a nerve disorder that may follow injury to the arms or legs.

"These are very interesting results, but we need to know more about what these changes really mean. Are they reflecting changes in brain metabolism," or do they indicate "true nerve cell loss?" wonders Anthony Jones, director of the human pain research group at the University of Manchester in England. "It seems unlikely that a strong sensory input would cause brain damage, since we know the brain is so good at protecting itself," he adds. If the loss is real, then the next step would be to determine if the damage can be reversed—and compensate for painful choices.

Lisa Melton is based in London.

PHYSICS

Seeing Single Photons

A SUPERCONDUCTING WAY TO SPOT PHOTONS ONE BY ONE

BY GRAHAM P. COLLINS

Charge-coupled devices, or CCDs, have become commonplace in modern consumer electronics. They are used in digital cameras and camcorders and in document scanners. Introduced in the late 1970s, they have become the workhorse light detector for astronomers. But CCDs have a number of limitations. In particular, they do not detect the wavelength (and hence the color) of light. Digital cameras get around this by having red, blue and green filters over individual pixels or over three separate CCD arrays. Filters, however, reduce the sensitivity and are of no use for measuring wavelengths with any precision. Now a group of researchers at the Jet Propulsion Laboratory and the California Institute of Technology, led by Peter K. Day of JPL, has demonstrated a detector based on superconducting technology that can detect individual photons and identify their wavelength. Best of all, the detector seems well suited to being engineered into a large array like a CCD.

The heart of the detector is made out of a thin film of aluminum on a sapphire substrate. The aluminum is etched by standard photolithographic processes to form a meander-

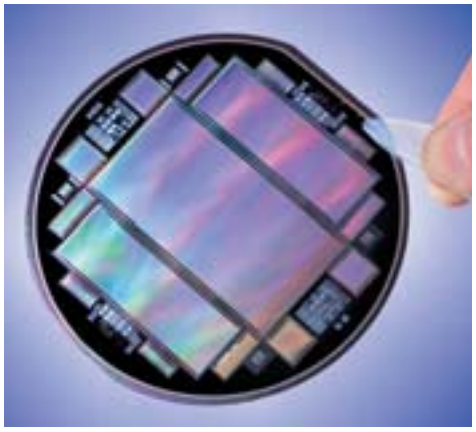
ing strip. When cooled to near absolute zero (less than one kelvin), the aluminum becomes superconducting. Like the vibrations of a tuning fork, current in the aluminum strip oscillates at a resonant frequency.

So how does all this detect a photon? In a superconductor, electrons form loosely bound pairs called Cooper pairs. It is those electrons that flow without resistance, and the ease with which they flow affects the strip's resonant frequency. When a photon strikes the strip, it breaks up some of the Cooper pairs, making the superconductor more "sluggish," which shifts the strip's resonant frequency and also diminishes the strength of the resonance. The photon's energy, which depends on its wavelength, determines the number of pairs that are broken and, therefore, the degree of change in the resonance. Amplifiers and other circuitry complete the detection process. The JPL-Caltech group tested a prototype with x-ray photons emitted by a radioactive isotope of iron, but the general design could be adapted for any wavelength from the submillimeter (microwave) range to gamma rays.

The JPL-Caltech sensor has an advantage over some competing designs that require a large number of output wires and a separate preamplifier for every pixel. By having each pixel operate at a slightly different resonant frequency, a large array of pixels could potentially all share one preamplifier and a single output wire.

Highly sensitive single-photon detectors have a wide variety of uses, including astronomical observations ranging from submillimeter wavelengths to gamma rays, x-ray analysis of materials, fluorescence microscopy of single molecules, and telecommunications. They have even been used to look for faults in integrated circuits by observing the infrared light emitted by transistors when they switch.

Before the JPL-Caltech device can join the ranks of other single-photon detectors, however, certain problems still remain to be worked out. In particular, noise levels are higher than expected. The detector's sensitivity "is good enough for some ground-based astronomy," JPL's Day says, "but an improvement of at least a factor of 10 is needed for the space-borne telescope applications we are interested in." The source of the noise that compromises the sensitivity will have to be identified and eliminated before the new detector is completely ready for prime time.



CHARGE-COUPLED DEVICES, such as this one from the Keck Telescope in Mauna Kea, Hawaii, are a mainstay in modern imaging. Superconducting detectors offer the promise of single-photon sensitivity.

DETECTING ON THE EDGE

Another type of highly sensitive photon detector is known as the transition edge sensor. It uses a patch of superconductor kept right on the "edge" of its superconducting transition—that is, at exactly the temperature at which its electrical resistance plummets to zero. A tiny change in temperature—such as that caused by absorption of a single photon—results in a large change in the sensor's resistance, which can be monitored by the output circuitry. A group at NIST in Boulder, Colo., is developing eight 1,600-pixel arrays to be deployed at the James Clerk Maxwell Telescope in Hawaii for observations in the submillimeter wave band (at present, the telescope is served by arrays totaling 128 pixels). Every pixel requires a dedicated superconducting quantum interference device (better known as a SQUID) serving it as an amplifier.

LAWRENCE BERKELEY NATIONAL LABORATORY *SPL*

OIL RECOVERY

Planning for *Prestige*

HOPE FOR GETTING THE OIL OUT OF A SUNKEN TANKER **BY LUIS MIGUEL ARIZA**

Some 14,000 of its 77,000 metric tons of heavy oil remain in the tanker *Prestige*, which sank off the coast of Spain in November 2002 and now rests below 3,800 meters of water. The spill immediately following the tanker's breakup caused upward of $1 billion in damages to Spain's shoreline and fisheries, and officials worry that the remainder may seep and periodically contaminate the coast. An attempt last October to retrieve some of the remaining oil has given engineers hope that they might be able to remove the infamous cargo safely.

The test took months of planning by the Spanish oil company Repsol YPF, which recruited engineers from various industries spe-

NEED TO KNOW: LOOKING FOR OIL

Determining the amount of fuel left in the *Prestige* meant using a neutron log. This device, used by the oil industry, relies on a radioactive source that emits neutrons, which are absorbed by hydrogen atoms. On absorption, gamma rays are emitted. Because water and oil give different gamma-radiation signatures, engineers could determine the amount of fuel in the tanker: about 13,100 metric tons in the bow, 700 in the stern.

IN THE BLACK: Fuel coated the rocks on Spain's northwest coast shortly after the breakup of the tanker *Prestige* in November 2002. Nearly 14,000 tons of oil remain in the sunken vessel; officials fear that fuel could wash ashore in the future if it is not removed soon.

cializing in deepwater operations. Ramon Hernan, technical director of the Repsol team, notes that until now, "there was no successful attempt to recover oil from a ship beyond a depth of 150 meters" and that "no robots had worked successfully at almost 4,000 meters." Retrieving the oil demanded modified deep-sea equipment and remotely operated vehicles (ROVs). "When you talk about 4,000 meters deep, there is no commercial activity there. Few industries push beyond 3,000 meters," explains Massimo Fontolan, a managing director of SonSub, the Italian firm that built one of the ROVs used in the operation.

The SonSub ROV did the bulk of the work, including patching various cracks in the *Prestige*. Other equipment was essential to perform specialty functions—including drilling a 70-centimeter-wide hole through the tanker and installing a double valve. The SonSub ROV positioned a plastic bag eight stories tall and 2.5 meters wide over the hole. Once the valves were opened, the fuel escaped out of the tanker, rising up into the collection bag as a stiff, straight column thanks to frigid temperatures and 380 atmospheres of pressure.

Whether the same procedures and technology will work for the rest of the cargo is not clear. (Repsol has been tight-lipped about details and would not permit project engineers from other firms to speak independently.) The fuel came out because of gravity: oil is lighter than water, so it rose up the water column and into the bag. It took 18 hours for 100 tons of oil to collect in the bag. The oil, however, may be too thick for all of it to come out by itself.

The team of Críspulo Gallegos, a chemical engineer at the University of Huelva in Spain, simulated the behavior of the fuel at 150 and 400 atmospheres. The researchers discovered that the viscosity of the oil depends on the flow rate, which is expected to decrease as the level in the *Prestige* tanks drops. As a result, the remaining fuel will thicken and have a harder time getting out. More holes could be drilled, but there is a limit to the number and the diameter of the holes that can be bored into the tanks. Engineers could try to direct seawater into the tanks to help flush out the fuel. The trick would be pumping in the water without compromising the structural integrity of the tanks.

The giant bag used to capture the oil may also need a redesign. Though consisting of several tough polymer layers, it broke after it was hoisted into a pool on board a surface ship. Fortunately, none of the oil escaped out to sea. Despite the challenges, Hernan is confident about getting the remaining oil. "The important fact here is that on October 11, there were 100 tons in the *Prestige*" that were removed, he remarks, "and a week later, the fuel was in El Ferrol, the Spanish port, for processing." The rest of the sunken oil, Spain hopes, could be retrieved for processing by this spring.

Luis Miguel Ariza is based in Madrid.

REUTERS NEWMEDIA, INC. *Corbis*

BY THE NUMBERS

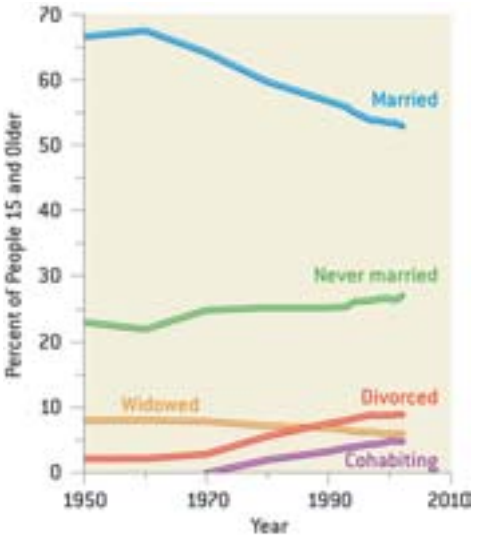
Living Together

IN THE U.S., COHABITATION IS HERE TO STAY

BY RODGER DOYLE

In some parts of the world, such as Latin America, cohabitation is a widespread and ancient tradition, but in the U.S. and other Western countries, it barely existed three decades ago. In America today, for every 100 married couples, there are 10 unmarried couples living together. But this statistic understates the prevalence of the practice, as a third of all married women younger than 45 have cohabited at some point in their lives.

Although an increasing number of older Americans do so, cohabitation takes place mostly among those 40 or younger and is generally short-lived, with two thirds of unions lasting two years or less. Those who cohabit tend to have less income and education, are less religious, and are more likely than noncohabitants to come from broken homes. Living together, more so for whites than blacks, is usually a prelude to marriage, although about one out of four cohabiting women do not see their unions as a first step to marriage but simply as an alternative to being single or dating. Some cohabitants live with parents, relatives or housemates. Cohabitation tends to be most prevalent in New England, Florida and the West and least prevalent in the South, the most conservatively religious region in the U.S.



SOURCE: U.S. Census. Data on cohabitants from Philip N. Cohen, University of California, Irvine.

Little evidence supports the popular notion that cohabitation is good training for marriage. Indeed, some research shows that those who cohabit are more likely to suffer marital discord and divorce. Other research, however, suggests that cohabitation as such is not to blame; rather it is the behavior of cohabitants, some of whom are prone to violence and excessive drinking. Compared with those in married-couple families, children in cohabiting families tend be poorer, are not as well fed and are not read to as frequently; they also have more behavioral problems. Children living with married parents fare better, although their advantage over children living with cohabiting parents may reflect race, ethnicity and their parents' education. The number of cohabiting couples with children is growing rapidly, and as things stand now, one in four children can expect to live in a cohabiting family sometime during childhood.

According to one theory, cohabitation thrives because women, with their growing financial independence, no longer feel the urgency of finding a husband to support them but nonetheless want to enjoy the benefits of a live-in partner. There is little evidence, however, that affluent women find marriage less desirable. Another possible explanation is that women, particularly young women, may be more inclined to cohabit because of a growing disenchantment with marriage, which they often see as a situation in which the wife takes on most of the domestic work. Still another explanation is that the need for an arrangement short of marriage was always there but remained unfulfilled until the 1970s, when feminism, oral contraception, more individualistic attitudes and social activism combined to loosen the bonds of convention.

The rise in cohabitation has only partially made up for the decline in marriage. Fewer than 60 percent of those who cohabit have never married, and thus the increase in the never-marrieds in the past three decades results only in part from rising cohabitation.

Rodger Doyle can be reached at rdoyle2@adelphia.net

CHARACTERISTICS OF COHABITANTS

Family Status	Percent
Never married	58
Divorced, separated	37
Widowed	4
Children	43
No children	57

Age	Percent
Under 20	2
20–29	36
30–39	29
40–54	25
55–64	5
65+	3

Race	Percent
White (non-Hispanic)	70
Black	13
Hispanic	14
Asian and Pacific Islanders	2
Native Americans	1

FURTHER READING

Changes in Family Structure and Child Well-Being: Evidence from the 2002 National Survey of America's Families.
Gregory Acs and Sandi Nelson. Urban Institute, 2003.
www.urban.org

Unions 2003.
The National Marriage Project.
http://marriage.rutgers.edu

Center for Family and Demographic Research. See papers by Wendy Manning, Susan Brown and Pamela Smock. www.bgsu.edu/organizations/cfdr/main.html

RODGER DOYLE

DATA POINTS: CONJOINED

The successful separations last October of two sets of twins joined at the head (one in Dallas and one in Rome) belie the long-standing surgical challenge. The earliest recorded separation traces to Constantinople circa A.D. 945, when doctors attempted to save the life of one Armenian twin after his brother, to whom he was joined at the abdomen, died. (After separation, the living twin died after only three days.) Most conjoined twins don't survive past their first day after birth. —*JR Minkel*

Recorded number of conjoined twins, to November 2003: 1,279
Conjoined twins per 100,000 births: 1 to 2
Per 200 identical twins: 1
Percent stillborn: 40 to 60
Percent surviving one day: 35
Ratio of female to male survivors: 3:1

Percent of conjoined twins joined at the:
Side: 28
Front of the chest: 19
Navel: 18
Hip, spine or sacrum (rump): 19
Head: 16

Number of recorded separation attempts: 245
Percent success rate when twins were joined at the:
Navel: 82
Hip: 63
Sacrum: 68

Number of attempts to separate twins joined at the crown: 33
Individual survivors: 34

SOURCES: www.twinstuff.com; Jonathan Muraskas, Loyola University; Entwined Lives: Twins and What They Tell Us about Human Behavior, *by Nancy L. Segal (Plume, 2000);* Conjoined Twins: Developmental Malformations and Clinical Implications, *edited by Rowena Spencer (Johns Hopkins University Press, 2003).*

MEDICINE

Gut Feeling

YOU'RE FEELING . . . BETTER: Hypnotherapy could produce long-lasting benefits.

Hypnosis is commonly thought of as a parlor trick, but it has some uses in treating medical conditions. One such sickness is irritable bowel syndrome. The ailment troubles up to 58 million Americans, causing abdominal pain, bloating, constipation and diarrhea. Special diets and drug therapies often fail to work well. Nearly 20 years ago researchers first tried hypnotherapy for the disorder. More recently, British researchers followed 204 patients for up to six years and found that 12 weekly one-hour hypnotherapy sessions significantly improved symptoms 71 percent of the time. Of these, 81 percent maintained gains for years after stopping hypnosis. These patients also said they took fewer drugs and saw doctors less frequently. Although hypnotherapy can be expensive, the investigators suggest the long-term benefits offset the cost. Fewer than one in 10 patients attempted alternatives after completing hypnotherapy. The findings appear in the October issue of the journal *Gut*. —*Charles Choi*

BIOLOGY

Holding in Suspense

Suspended animation sounds like science fiction, but recently biologists uncovered genetic mechanisms that actively coordinate this process—at least for an oxygen-starved *Caenorhabditis elegans* worm. The scientists used a technique called RNA interference to disrupt the activity of specific genes. When *C. elegans* embryos lacked a functional copy of the gene *san-1* or *mdf-2*, they were more likely to succumb to a lack of oxygen than their normal peers, which can maintain suspension for days. These genes are key to coordinating the motions of cell contents during cell division. When oxygen-starved, embryos with knocked-out *san-1* or *mdf-2* failed to sort chromosomes properly as they grew. The researchers note that these kinds of genes are highly conserved, indicating that a code for suspended animation could be found throughout the animal kingdom. Indeed, invertebrates, fish and mammals can at times enter suspended animation to survive extreme oxygen deprivation. The study is discussed in the November 7 *Science*. —*Charles Choi*

PHYSICS

Slip and Slide

FRICTION is inevitable when two objects come together—or maybe not.

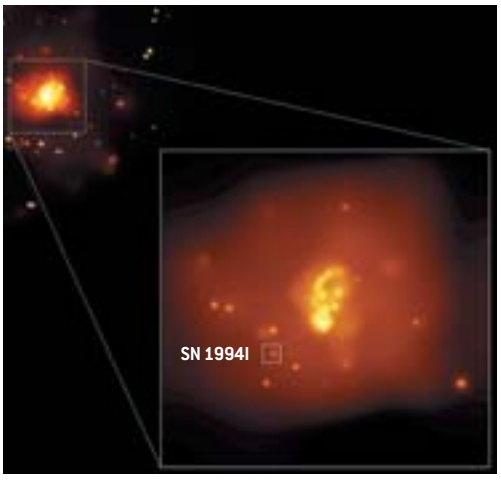
Negative friction, which would cause molecules sliding past one another to speed up rather than slow down, might be possible. Behind the theory is the van der Waals force, which normally causes molecules to weakly attract one another. Electrons vibrate, and when close together, they jiggle in sync, thereby generating an electric pull. Negative friction could result from modified jiggling. The effect would take advantage of the Doppler shift, in which each molecule sees its neighbors' electrons vibrating at slightly higher frequencies as the molecules approach and at lower frequencies as they drift apart. Physicist Adam E. Cohen of Stanford University and physical chemist Shaul Mukamel of the University of California at Irvine say it should be possible to change how electrons vibrate (through light or heat, for instance) and to tune the frequencies so that molecules attract one another as they approach but repel as they move apart. The theory will be published in *Physical Review Letters*. —*Charles Choi*

JOHNER *Photonica (top);* DiMAGGIO/KALISH *Corbis (bottom)*

ASTRONOMY

Blasts, Bursts and Flashes

Astronomers long suspected that bursts of gamma rays were related to exploding stars but remained unsure how to categorize the events with respect to other celestial blasts. A gamma-ray burst that reached the earth on March 29, 2003, however, suggests that most such occurrences are produced by the same type of cosmic blast. The burst, the closest ever recorded—at 2.6 billion light-years—enabled astronomers to measure in detail the energy produced. A comparison to previous bursts, x-ray flashes and rare, type Ic supernovae revealed that such events release nearly the same amount of energy (roughly equal to that produced by the sun in its lifetime). Hence, they probably share a common origin, most likely the death of a massive star. Edo Berger of the California Institute of Technology, who studied the burst, says that what differs between the explosions is the "escape route" the energy takes. The research appears in the November 13 *Nature*. —*Chris Jozefowicz*



WHIRLPOOL GALAXY, seen in x-rays, contains a supernova called SN 1994I. Such rare, type Ic blasts may be at the root of gamma-ray bursts and x-ray flashes.

MATERIALS SCIENCE

A Pulse for Magnetic Memory

The time needed to toggle between magnetic states, which sets the top speed of magnetic memory chips, may have just seen a big improvement. Researchers knew that laser light shining on the ferromagnetic element gadolinium vibrated its atoms and in turn rocked their magnetic spin states relative to one another, but they thought the transmission of energy between vibration and spin had to be messy and random, making memory storage impossible. Now German physicists have observed that chopping the laser light into 30-femtosecond (10^{-15}-second) pulses causes gadolinium atoms and their spins to wobble in lockstep at three terahertz—1,000 times as fast as conventional magnetic memory systems. The scientists speculate that combining pulses may produce magnetic bits suitable for short-lived buffer memory, although incorporating ultrashort laser pulse technology into computers would be tricky, to say the least. The findings were to have appeared in a November issue of *Physical Review Letters*. —*JR Minkel*

DETECTORS

Snoop Tube

Existing detectors for pollutants and chemical and biological agents sense only relatively high particle densities. Although vibrating devices can concentrate aerosols into low-pressure nodes, current designs are hard to align and consume lots of power. Now a pipe made of piezoelectric crystal has shown it can concentrate particles up to 40 times using a mere 0.1 watt, making it suitable for battery-powered, handheld detectors, according to Los Alamos National Laboratory scientists. They vibrated tubes several inches wide and long in and out (oscillating the tube's diameter) to produce an internal standing pressure wave in which particulates could be trapped. The tube generated three narrow streams aligned with the axis, at airflows of up to 250 liters of air per minute, as reported at a November meeting of the Acoustical Society of America. —*JR Minkel*

VIBRATIONS trap aerosols along three nodes (*white dots*).

BRIEF POINTS

- **A compound based on a mutant form of the HDL cholesterol molecule found among rural Italian villagers reduced years' worth of plaque buildup in coronary arteries after just five weeks of treatment.**

  Journal of the American Medical Association, November 5, 2003

- **Researchers effectively created the genome of the bacterial virus Phi-X174 from scratch in 14 days. Previous efforts took years, and the resulting synthetic organisms harbored genetic defects.**

  Proceedings of the National Academy of Sciences USA (in press)

- **Lemmings don't follow one another in a suicidal jump into the sea, but they do follow boom-and-bust population cycles—apparently because of a combination of predators (foxes, owls and others), rather than from shortages of food or space.**

  Science, October 31, 2003

- **The shape of beverage containers influences how much people pour and drink. They will pour more into a short, wide glass than into a tall, narrow one, even though they think they do the opposite.**

  Journal of Consumer Research, December 2003

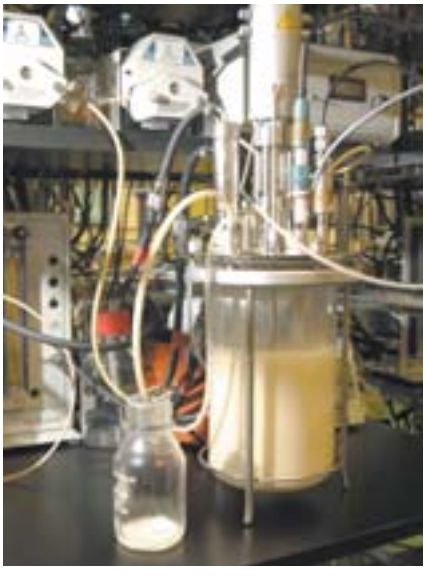
NASA/CXC/UMD/A. WILSON ET AL. (*top*); COURTESY OF LOS ALAMOS NATIONAL LABORATORY (*bottom*)

Supercharging Protein Manufacture

A career deviation leads to a dynamic approach to producing biotech drugs By GARY STIX

Tillman U. Gerngross came to Dartmouth College in the late 1990s as a tenure-track professor who wanted to study "green" plastics derived from plant-derived sugars. His first major project centered on performing an analysis of the costs and benefits of these supposed materials of the future.

In 1999 he published a paper in *Nature Biotechnology* that detailed the results of a life-cycle analysis of bioplastics manufacturing. It showed that making these purportedly eco-friendly products required more fossil fuels than fabricating petroleum-based plastics. "We have spent literally hundreds of millions of dollars developing these technologies to make green polymers. And at the end of the day, the net impact is going to be marginal," Gerngross says [see "How Green Are Green Plastics?" by Tillman U. Gerngross and Steven C. Slater; SCIENTIFIC AMERICAN, August 2000].



FERMENTER: A GlycoFi vessel is where protein-based drugs will be made in yeast.

The 1999 paper got a lot of attention. But it also spurred the Austrian native's decision to look for another line of research. "In one publication, I essentially divested myself of that [green plastics] work," he says. For a while, Gerngross became what he describes as a "poster boy for the debunking movement." But more scientist than pundit, he realized that he could not spend decades belaboring this one idea. The experience taught him that before taking on any new research endeavor, he should examine whether the scientific problem he had chosen to go after was really worth solving.

At the time, the Human Genome Project was entering its final stages. Trained as a chemical engineer and molecular biologist, Gerngross started to take a close look at all the steps leading from gene identification to the coding and making of proteins—the staple therapeutics in biotechnology. In particular, the manufacture of proteins caught his eye. "I realized that this is fairly medieval. It is a lengthy process that we can't control well, that has all sorts of shortcomings and that there ought to be a better way of making proteins." Today's standard method involves inserting a gene into Chinese hamster ovary cells or other mammalian cells, which then express the human protein; it can take two to three weeks to produce relatively small amounts of a protein-based drug.

Gerngross wondered whether generating human proteins in yeast might produce better results: "Yeast can make boatloads of protein, but they can't put the right sugars on the protein." Among other things, the sugars ensure that the protein folds properly and that it is thermodynamically stable. As he talked to colleagues, Gerngross realized that devising a production process for glycosylated proteins—ones with the desired sugars added—would meet his criterion for pursuing worthwhile research. "People said to me, 'This is a hard problem, but if you solve it, this would be a big deal.' "

Typically an academic applies for a government grant and sets to work with a few graduate students. As a newcomer to the field of glycobiology, Gerngross knew he had little chance of getting support through traditional funding routes—and if he did receive the money, it would take years to achieve substantive results: "By that time, the boat would have left and someone else would have picked this up."

Charles E. Hutchinson, a teaching partner and former dean of the engineering school at Dartmouth, was intrigued. He told Gerngross that the only way to proceed quickly would be to launch a company. A veteran of multiple start-ups, Hutchinson helped to interest a

GLYCOFI

venture-capital firm, Polaris Ventures in Waltham, Mass., in providing $600,000 in 2000. In exchange for an equity stake in the newly formed GlycoFi (short for "glycosylation fidelity"), the university agreed to let the two men use Gerngross's laboratory space on campus to get started. With a call extending a job offer to Stefan Wildt, a former postdoctoral colleague from the Massachusetts Institute of Technology, the company had become more than what Gerngross calls "a postal box and a cute idea."

Gerngross and his colleagues set about reengineering the glycosylation pathway of several yeasts, initially focusing on *Pichia pastoris,* which is widely used in the production of industrial enzymes. First, it was necessary to knock out the genes in yeast that encoded enzymes that would place the wrong sugars on a human protein, making it an immediate target for disposal by immune cells. Deleting genes was by far the simplest task.

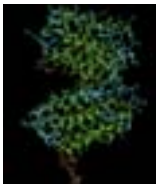
The biggest challenge, and one that had foiled other investigators, came next: to create an assembly line of enzymes needed to put the appropriate sugars on a human protein being manufactured in the yeast cell. Kirin Brewery, for one, had inserted the human gene for a critical glycosylation enzyme in yeast, but little had happened. The GlycoFi team reasoned that for the enzyme to work, it would have to get to the right place in the yeast cell. The researchers attached a peptide, a small chain of amino acids, to the enzyme. This peptide zip code then directed the enzyme to either the yeast cell's endoplasmic reticulum or its Golgi apparatus.

In addition to helping the enzymes find their way in the cell, GlycoFi began a cross-species search to locate the best enzymes to perform the diverse reactions required to sugarcoat the human proteins. The enzymes were not always culled from human cells; rat, worm, plant or yeast enzymes sometimes carried out the reactions needed to glycosylate a human protein better than their human counterparts did. The genes for the best enzymes, whether rat or human, were engineered to express the correct peptide zip codes and then inserted into the yeast.

This sugar assembly line has functioned better than anyone expected. For reasons no one yet fully understands, the yeast does not appear to be weakened by this fiddling with its internal workings. The most recent report on GlycoFi's research—announcing the first production of a human protein decorated with complex sugar molecules—was published in the August 29, 2003, *Science*.

More still needs to be done before GlycoFi can offer a complete industrial platform that will compete with Chinese hamster ovary cells. The yeast must be engineered further to add the sugar sialic acid to a protein. But the possibility of making human proteins in yeast cells looms as a formidable technology. Gerngross notes that fermentation times in yeast may take three days, compared with two to three weeks in hamster cells. And both the amount of protein produced and the uniformity of the product show the promise of the technology. Lowered production costs from these improvements in manufacturing could potentially bring down the cost of biotechnology drugs.

"We hope to be able to produce longer-lasting and better drugs," Gerngross says. "You may not have to administer as much as you would with another drug to get the same therapeutic effect." GlycoFi might also make drugs that simply cannot be produced in mammalian cells. Gerngross points out that yeast, for example, can manufacture high concentrations of the properly glycosylated form of the protein alpha-1 antitrypsin, a deficiency of which can cause liver and lung disease. Creating the protein in hamster cells is impractical because of low yields.



The challenge that had foiled others was to create an assembly line of enzymes that put the appropriate sugars on a human protein made in yeast.

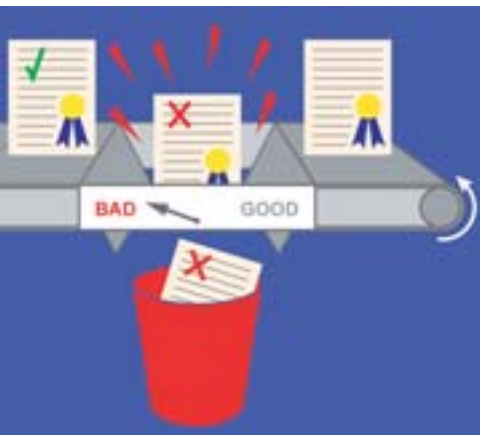
GlycoFi, now with 37 employees, has grown beyond the confines of Gerngross's college lab. Its new headquarters in Lebanon, N.H., was a presidential campaign stop for Senator Joseph Lieberman of Connecticut last July. The company is now closing its third round of venture financing, having brought in nearly $18 million since its inception. Moreover, it has already received some revenue from drugmakers such as Biogen Idec and Baxter Healthcare, which have each supplied a gene; in return, GlycoFi is providing the specified protein. So far, discerning how to put sugars on human proteins made in yeast looks like a problem well worth solving. SA

GLYCOFI (*glycosylated protein*)

In Search of Better Patents

How to get rid of bad filings without costly lawsuits By GARY STIX

Federal courts and the U.S. Patent and Trademark Office have expanded the scope of patenting to include areas—such as genes, software and business methods—that were once thought to be of questionable validity. For instance, in 1998 the patent office granted an application for a method of charging more for a product for which demand fluctuates little in response to a price change.



The business-method patent in question triggers worries about the requisite qualifications of the patent examiners involved. If the evaluators had consulted the chapters on imperfect competition in any number of economics textbooks, they might have thrown out the application as violating the standards that a patent should be both new and nonobvious (inventive). Poor decision making during the examination process leaves patents open to being overturned by the courts—at a cost that can range into the millions of dollars for a legal proceeding. Because of the widening breadth of what can be patented, and the seeming inability of examiners to stay up to date, some analysts have proposed providing a means to invalidate a patent short of a lawsuit.

The current process allows reexamination of patents in only a few narrowly construed circumstances. And it is usually employed by a patent holder to broaden the claims of an existing patent, not to question its validity. Two scholars—Jonathan D. Levin, an economics professor at Stanford University, and Richard C. Levin, president of Yale University, a son-and-father team—recently made the case for changes to the status quo in a National Research Council report entitled "Patents in the Knowledge-Based Economy." The Levins created an economic model that showed the benefits of a system similar to one in Europe that would provide a simple and inexpensive administrative procedure that lets a patent be rescinded when it fails to meet such basic criteria as being new, useful and nonobvious.

A new kind of "postgrant" review would encourage the adoption of innovative technology by eliminating uncertainties about whether a patent would be overturned or upheld. Others would then know whether they would have to license the technology or would be free to pursue its use unhindered. In Europe, the estimated cost of undertaking a patent opposition is less than $100,000 for each party, although the adjudication proceeding takes nearly three years as a result of long deadlines for filing claims and counterclaims.

A streamlined version of this process in the U.S. could dispose of cases more quickly. The Federal Trade Commission, in fact, recommended last October instituting a revamped postgrant review procedure, and, separately, the patent office is contemplating improved reviews as part of a major internal overhaul plan. Putting in place a new type of evaluation, however, might encourage more challenges to existing patents because of the lower costs of undertaking such an action. In Europe, more than 8 percent of biotechnology and pharmaceutical patents were opposed, compared with a 1 percent litigation rate in the U.S.

But the Levins argue that other benefits may accrue beyond just saving on the cost of litigation. Better postgrant review would help ensure that government confers the monopoly privilege of a patent only on truly innovative inventions. Moreover, many of those who file an opposition may do so because of their detailed knowledge of a highly technical area, such as genetics or software, supplying valuable lessons to patent examiners who struggle to stay current with the state of the art. A decision from an opposition proceeding will provide a good reading on the critical determinants of whether an invention really succeeds in living up to its name. SA

JENNIFER KANE

Bunkum!

Broad-mindedness is a virtue when investigating extraordinary claims, but often they turn out to be pure bunk By MICHAEL SHERMER

Those of us who practice skepticism for a living often find ourselves tiptoeing politely around the PC police, who think that all beliefs and opinions are equal. Thus, when asked, "Are you a debunker?" my initial instinct is to dissemble and mutter something about being an investigator, as if that will soften the blow.

But what need, really, is there to assuage? According to the Oxford English Dictionary, to debunk is to "remove the nonsense from; to expose false claims or pretensions." Bunk is slang for "humbug," and bunkum is "empty claptrap oratory." Here is some bunk that merits no brook.

Ear coning cleans your ears and mind. The idea is to lie down on your side with your head on a pillow. Then place a long, narrow, cylindrical cone of wax into your ear canal until a tight seal forms. Light the open end of the cone on fire. The negative pressure created will not only remove undesirable earwax, according to Coning Works in Sedona, Ariz., but also provide "spiritual opening and emotional clearing, realignment and cleansing of subtle energy flows, sharpening of mental functioning, vision, hearing, smell, taste and color perception." The technique "acts as a catalyst to clear out debris from nerve endings allowing for clear vibrational flow to corresponding areas of mind, body and spirit." Why pay $25 to $75 to have your ears cleaned by your doctor, asks another ear-cone seller, Wholistic Health Solutions, "when you can easily do it at home?"

Well, for starters, according to a 1996 study conducted by physicians at the Spokane [Wash.] Ear, Nose and Throat Clinic and published in the journal *Laryngoscope,* "Tympanometric measurements in an ear canal model demonstrated that ear candles do not produce negative pressure," and thus there was no removal of wax in the eight ears tested. Worse, a survey of 122 otolaryngologists (ear, nose and throat docs) identified 21 ear injuries from ear coning. If one is inclined toward such self-mutilation (or a good chortle), however, I recommend a quick stop at the satirical buttcandle.com, which touts a "gentler alternative to laxatives, enemas and anti-flatulence pills" in the form of a carefully (and gently) placed hollow candle that when burning creates a vacuum that draws out impurities. Best of all, it's "100% soluble and septic-safe."

Laundry balls clean clothes. These spherical, toroidal or spiked balls contain no chemicals and yet are purportedly reusable indefinitely in the washing machine to clean, deodorize, sterilize, bleach and soften clothes. But they do not "ionize," "structure," "cluster" or "magnetize" water, as various manufacturers claim. They all work on the same principle: washing clothes in soapless warm water does have some cleansing effect, particularly for nongreasy garments mainly soiled by dust, dirt and sweat. But with laundry balls costing from $25 to $75, golf balls are just as effective and a lot cheaper.

A counterfeit pen can detect counterfeit bills. Containing tincture of iodine that reacts with the starch in recycled paper to create a black streak, the pen only works to catch counterfeiters who are brainless enough to use cheap paper, thus creating a false sense of security. Meanwhile clever counterfeiters who use high-quality fiber or linen paper containing no starch or whitening agents continue to fleece their marks. Merchants beware: after warning law-enforcement agencies—who ignored him—fellow skeptic James Randi periodically applies commercial spray starch on $50 and $100 bills for recirculation into the economy in the hopes that false pen positives will force the bunkum squads into action.

No one talks bunkum with a better vocabulary than those who lace their hokum with scientistic jargon.

To "buncomize" is to "talk bunkum," and no one does this with a better vocabulary than pseudoscientists, who lace their hokum narratives with scientistic jargon. (One laundry ball manufacturer claims that it "works on 'Quantum Mechanics' (Physics), not chemistry, with a method called 'Structured Water Technology.'" Another uses "infra-red waves that change the molecular structure of the water.") To "do a bunk" is to "make an escape" or "to depart hurriedly," a wise move when skeptics arrive on the scene fully armed with steel-jacketed science and armor-piercing reason. SA

Michael Shermer is publisher of Skeptic *(www.skeptic.com) and author of* How We Believe *and* In Darwin's Shadow.

BRAD HINES

Why Machines Should Fear

Once a curmudgeonly champion of "usable" design, cognitive scientist Donald A. Norman argues that future machines will need emotions to be truly dependable By W. WAYT GIBBS

Slowly and with care, Donald A. Norman refills his teacup, but the tea drips down the pot anyway. I look down at the small puddles of green tea on the restaurant table and back up at Norman. Here it comes, I think, bracing myself for a classic Norman fulmination on how basic design flaws in ordinary objects are the true sources of most "human error." After all, such cantankerous critiques in his 1988 book *The Psychology of Everyday Things* were what brought him international fame outside the narrow field of cognitive science.

JEFF SCIORTINO

DONALD A. NORMAN: EMOTIONAL DESIGNER

- **First design project: ham radio station built during childhood from military surplus parts.**
- **Characteristic obsession: finding out the purpose of the notch in a *cuillère à sauce individuelle,* a spoonlike utensil in fancier restaurants in Europe.**
- **Typical job: scientific consultant to firms such as Evolution Robotics in Pasadena, Calif., which has developed a prototype personal robot named ER2.**
- **Some favorite designs: Cooper Mini automobile; Alessi *Te ò* tea strainer, which "hugs" the cup; the Ronnefeldt tilting teapot, which holds the leaves on a shelf, immersed when steeping but out of the water when serving.**

But Norman calmly wipes his napkin over the spill without comment. Although he still calls himself a user advocate, these days he focuses less on the failures of modern technology and more on its potential, envisioning a world populated by well-performing, easy-to-use and even emotive machines.

"This is the new me," the 67-year-old professor at Northwestern University announces the next day in his keynote address to a large human-computer interaction conference in Ft. Lauderdale, Fla. "The old me was critical, always finding fault with things that didn't work." In June 2002, for example, the journal *Computer* published his excoriation of the consumer electronics industry for the absurd "living-room rocket science" needed to get high-end home theater components to function together.

But in writing *Emotional Design,* his latest work (due out in January from Basic Books), Norman seems to be attempting a metamorphosis from gadfly to oracle. "The new life is about emotion and positive things. So I only say nice things," he avers. "Or rather, I try."

The one picture of a teapot that Norman includes in *Emotional Design,* for instance, is there to illustrate "why lovely things really do work better." The particular teakettle shown has a melodic whistle on its spout, so it blows a harmonious steamy chord when ready to serve. Probably few would argue with the notion that phones and computers would be improved if their bleats and whirrs were less noisome.

But Norman's point goes much deeper. "The cognitive sciences grew up studying cognition—rational, logical thought," he notes. Norman himself participated in the birth of the field, joining a program in mathematical psychology at the University of Pennsyl-

vania and later helping to launch the human information–processing department (now cognitive science) at the University of California at San Diego. "Emotion was traditionally ignored as some leftover from our animal heritage," he says. "It turns out that's not true.

"We now know, for example, that people who have suffered damage to the prefrontal lobes so that they can no longer show emotions are very intelligent and sensible, but they cannot make decisions." Although such damage is rare, and he cites little other scientific evidence, Norman concludes that "emotion, or 'affect,' is an information processing system, similar to but distinct from cognition. With cognition we understand and interpret the world—which takes time," he says. "Emotion works much more quickly, and its role is to make judgments—this is good, that is bad, this is safe."

The two systems are intertwined at a biological level, Norman points out. "The affective system pumps neurotransmitters into the brain, changing how the brain works. You actually think differently when you are anxious than when you are happy. Anxiety causes you to focus in on problems; if something doesn't work, you try it again, harder. But when you're happy, you tend to be more creative and interruptible." So if only for purely utilitarian reasons, devices and software should be designed to influence the mood of the user; they will be more effective because they are more affective.

USER FRIENDLIER: Personal robots, such as the ER2 prototype made by Evolution Robotics, might work better if they are programmed to feel fear, pride and frustration, Donald A. Norman argues.

The idea is more controversial than it may seem. Even Jakob Nielsen, a former user-interface expert at Sun Microsystems who joined with Norman to form a consulting firm five years ago, notes that "there is always a risk that designers will misinterpret this kind of analysis," taking it as carte blanche to elevate form above function.

The problem is that taste varies. Watches, for instance, are designed largely for their visceral, sensual appeal, and for that very reason they come in myriad varieties. Aside from the big hand and little hand, however, there is no standard interface. The more complicated functions of any given watch—its calendar, stopwatch, alarm, countdown timer, and so on—can be maddeningly difficult to operate. Mastering one watch is of scant help in using a different model. So as mobile phones, PDAs and other gadgets continue to morph from tools to fashion accessories, an inherent conflict may arise between the diversity of designs needed to appeal to all customers and the consistency of operation that makes devices easy to use.

On that question, "I think Don is an optimist," Nielsen says. Nielsen has studied the usability of Web sites, and the results in that realm are not encouraging. "In many ways, we still don't have the basics settled. Most people can't write a good headline for their Web site, let alone get the information architecture right."

Norman argues, moreover, that machines should not only evoke emotional responses in their owners but should also in some sense feel emotions themselves. Here he parts company with many of his colleagues in human-computer interaction.

"I'm not saying that we should try to copy human emotions," Norman elaborates. "But machines should have emotions for the same reason that people do: to keep them safe, make them curious and help them to learn." Autonomous robots, from vacuum cleaners to Mars explorers, need to deal with unexpected problems that cannot be solved by hard-coded algorithms, he argues.

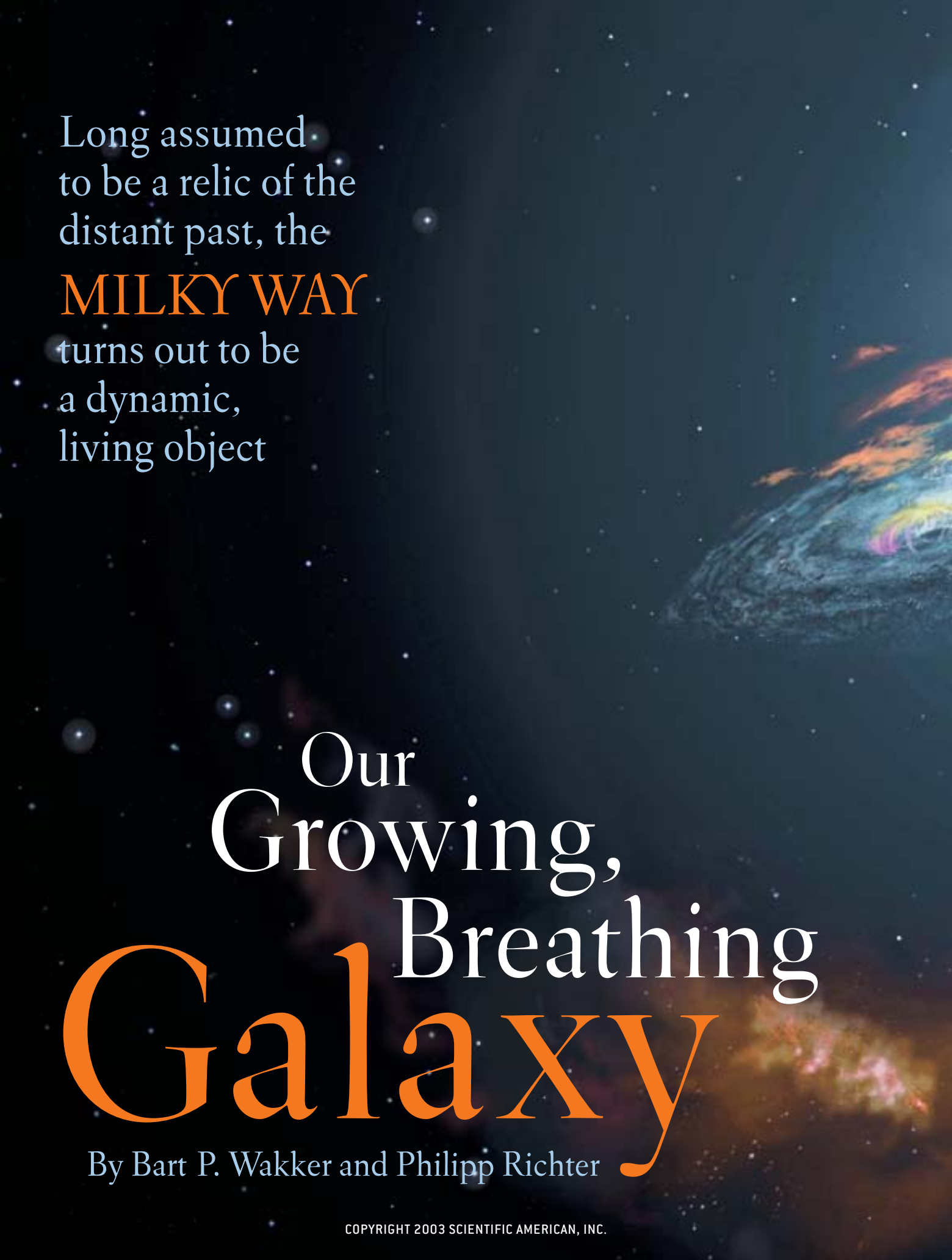
What they need are "weak methods." "Boredom," Norman explains, "is a weak method for getting out of a rut. Curiosity is a weak method for exploring an unfamiliar space. I want my automatic vacuum cleaner to fear heights so that it doesn't fall down the stairs." And, he maintains, if machines have a way of expressing their emotions—grimacing when they are frustrated, say—that would give people a useful glimpse into their internal operation.

Judging by the thousands of designers and researchers who turned out to hear his address at the Florida conference, his ideas carry weight. Yet as Norman held forth on the exhibit floor about the importance of making machines with feelings, Ben Shneiderman of the University of Maryland displayed a clear emotion of skepticism.

"My feelers come out when people use the language of people to talk about machines," he rebutted. "I think that leads down the wrong path." B. J. Fogg, whose research at Stanford University centers on the emotions that users inevitably attribute to their computers, observes that programming pseudoemotions into machines "might make the interaction with users go better. But there is an ethical question: it is a kind of deception."

And in any case, how could emotions be reduced to source code and circuitry? Such technical details are nowhere to be found in Norman's books and speeches, a limitation of which he is quite conscious. "All my life I have tried to develop frameworks, ways of looking at questions that current theories don't address. People say: that's very nice, but how do we realize this vision? You don't give us tools and measures. I guess that criticism is on the mark." SA

EVOLUTION ROBOTICS



Long assumed to be a relic of the distant past, the MILKY WAY turns out to be a dynamic, living object

Our Growing, Breathing Galaxy

By Bart P. Wakker and Philipp Richter

GULPING DOWN GAS and cannibalizing its smaller neighbors, the Milky Way galaxy is still in the process of forming. For a key to this image, see page 41.

Sometimes the hardest things to understand

are the things you are most familiar with.

We may know our hometowns intimately, yet visitors or young children may still point out things we have never noticed before. They may not be as attuned to all the minutiae, but they often see the big picture better than longtime residents can. A similar situation faces astronomers who study the Milky Way: we are so deeply embedded in our home galaxy that we cannot see it fully. When we look at other galaxies, we can discern their overall layout but not their detailed workings. When we look at our own, we can readily study the details but perceive the overall structure only indirectly.

Consequently, we have been slow to grasp the big picture of the Milky Way's structure and history. Astronomers were not even sure that the galaxy was a distinct object, only one of many billions, until the 1920s. By the mid-1950s they had painstakingly assembled the picture that most people now have of the Milky Way: a majestic pinwheel of stars and gas. In the 1960s theorists proposed that our galaxy formed early in cosmic history—by the most recent estimate, 13 billion years ago—and has remained broadly unchanged ever since.

Gradually, though, it has become clear that the Milky Way is not a finished work but rather a body that is still forming. Like the earlier discoveries, this realization has relied heavily on observing other galaxies and bringing the lessons back home. Most galaxies are now assumed to result from the merging of smaller precursors, and in the case of the Milky Way, we can observe the final stages of this process. Our galaxy is tearing apart small satellite galaxies and incorporating their stars. Meanwhile gas clouds are continually arriving from intergalactic space. No longer can researchers speak of galaxy formation in the past tense.

The evidence for the continuing accretion of gas by the Milky Way involves high-velocity clouds, or HVCs—mysterious clumps of hydrogen, up to 10 million times the mass of the sun and 10,000 light-years across, moving rapidly through the outer regions of the galaxy. HVCs were discovered 41 years ago, but only in the past five years have new data and new ideas provided the evidence that some of them represent infalling gas. HVCs also show that the galaxy is breathing—pushing out gas and then pulling it back in, as if exhaling and inhaling. In addition, the properties of HVCs suggest that a gigantic sphere of hot, tenuous plasma surrounds the galaxy. Astronomers had long suspected the existence of such a sphere, but few thought it would be so large.

Historically, interpreting HVCs has been difficult because being stuck within the galaxy, we have no direct way to know their locations. We can see their two-dimensional positions on the sky but lack depth perception. Over the past four decades, this ambiguity has led to many alternative hypotheses, some placing HVCs close to our own stellar neighborhood, others locating them deep in intergalactic space. The recent breakthroughs have occurred mainly because ground-based and orbiting telescopes have finally managed to get a three-dimensional fix on the clouds—and thereby a better perspective on our celestial hometown.

Overview/*High-Velocity Clouds*

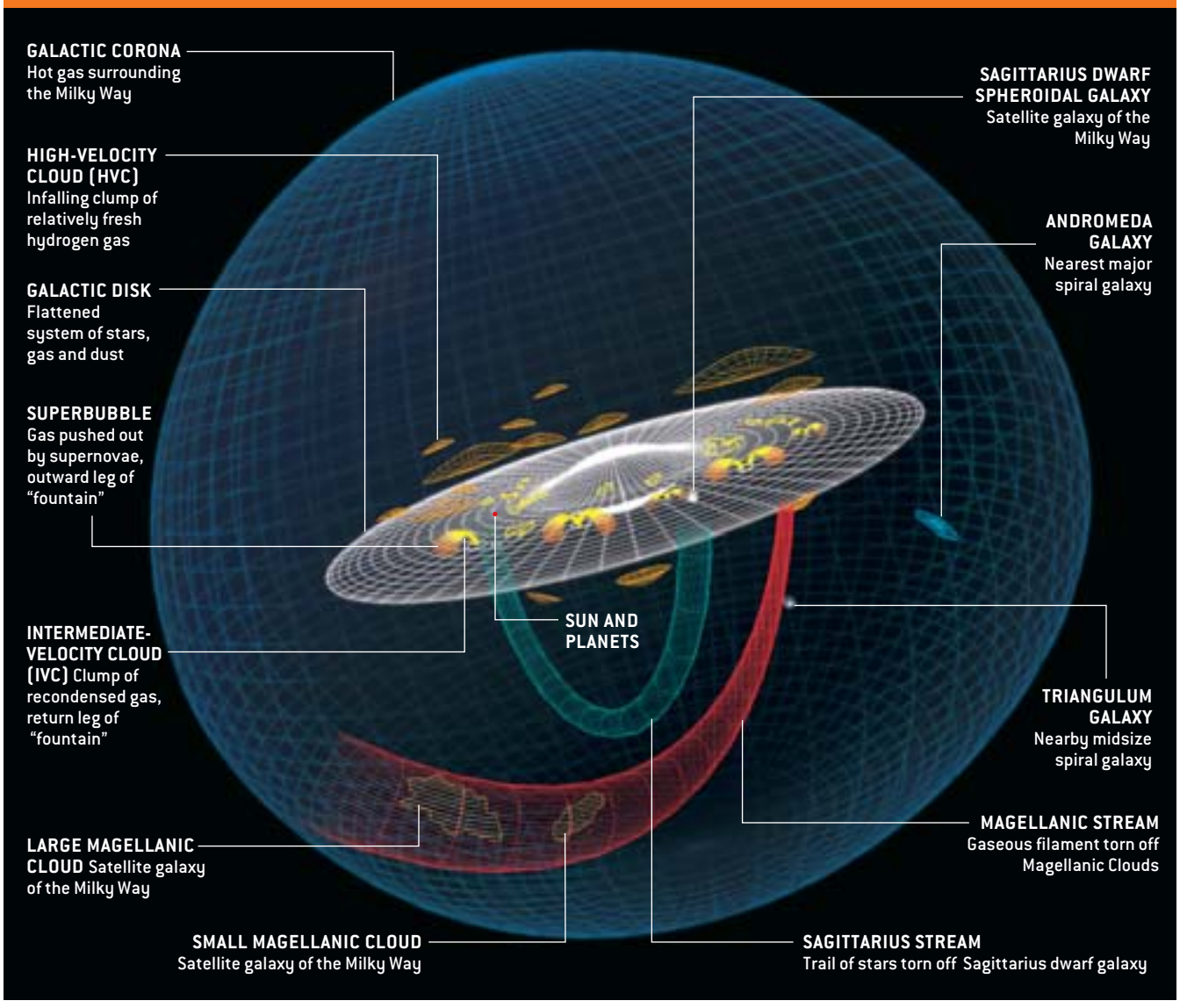
- Since the early 1960s astronomers have thought that the Milky Way and other galaxies were born early in cosmic history and then evolved slowly. Today, however, evidence indicates that galaxies are continuing to grow. They cannibalize their smaller brethren and gulp down fresh gas from intergalactic space.
- In our Milky Way we have a close-up view of the ongoing construction work. The incoming gas takes the form of high-velocity clouds discovered decades ago. Only recently were some of these clouds proved to be fresh material; observationally, they get entangled with circulating gas.
- These clouds come in several guises: clumps of neutral hydrogen reminiscent of intergalactic gas; a stream of gas torn out of nearby small galaxies; and highly ionized hot gas that may be dispersed throughout the intergalactic vicinity.

Virgin or Recycled?

OUR GALAXY CONTAINS about 100 billion stars, most of which are concentrated in a thin disk about 100,000 light-years across and 3,000 light-years thick. These stars revolve around the galactic center in nearly circular orbits. The sun, for example, trundles around at nearly 200 kilometers per second. Another 10 billion stars form the galactic "halo," a huge spheri-

RON MILLER (*preceding pages*)

OUR GALAXY AND ITS NEIGHBORHOOD



SLIM FILMS

cal envelope that surrounds the disk. Between the stars lie gas and dust, forming the interstellar medium, most of which also moves in nearly circular orbits around the galactic center and is even more narrowly concentrated in a disk than the stars are. Like a planet's atmosphere, the gas in the medium is densest at its "bottom" (the galactic plane) and thins out with height. But up to about 10 percent of the interstellar medium lies outside the plane and moves up to 400 kilometers per second faster than rotation would imply. This gas constitutes the HVCs.

The story of HVCs began in the mid-1950s, when Guido Münch of the California Institute of Technology discovered dense pockets of gas outside the plane—a clear exception to the rule that the density of gas diminishes with height. Left to themselves, those dense pockets should quickly dissipate, so in 1956 Lyman Spitzer, Jr., of Princeton University proposed that they were stabilized by a hot, gaseous corona that surrounded the Milky Way, a galactic-scale version of the corona around the sun [see "The Coronas of Galaxies," by Klaas S. de Boer and Blair D. Savage; Scientific American, August 1982].

Inspired by Spitzer's proposal, Jan Oort of Leiden University in the Netherlands conjectured that the galactic halo might also contain cold gas very far from the galactic plane. A search for radio emission from cold clouds resulted in their discovery in 1963. Unlike the gas found by Münch, these clouds did not follow the overall rotation of the galaxy; instead they seemed to be falling toward the galactic disk at high speed, so they became known as HVCs. A slower-moving but still anomalous type of cloud, an intermediate-velocity cloud, or IVC, was spotted the same year.

Oort later fleshed out his idea and suggested that after the initial formation of the galaxy, gas near the edge of its gravitational sphere of influence was left over. This gas reached the disk only after 10 billion years or more, becoming observable as HVCs. Oort's idea fit in well with models that try to explain

the observed chemical composition of the galaxy. Stars produce heavy elements and scatter them into interstellar space when they die. Newly born stars incorporate those elements and produce even more. Therefore, if the galaxy were evolving in isolation, each generation of stars should contain more heavy elements than its predecessors.

Yet most stars in the solar neighborhood, regardless of age, have about the same abundance of heavy elements. The favored explanation for this apparent discrepancy is that the galaxy is not isolated and that interstellar gas is constantly being diluted by more pristine material. Several researchers surmised that some or all of the HVCs represent this fresh gas, but the proposition lacked direct observational evidence.

An alternative hypothesis holds that HVCs have nothing to do with an influx of gas but are instead part of a "galactic fountain." This idea was proposed in the mid-1970s by Paul Shapiro, now at the University of Texas at Austin, and George B. Field of the Harvard-Smithsonian Center for Astrophysics. Gas heated and ionized by massive stars rises out of the disk into the corona, forming an atmosphere. Some regions then cool off, rain back down and become electrically neutral again, setting up a cycle of gas between the disk and the corona. In 1980 Joel Bregman, now at the University of Michigan at Ann Arbor, suggested that HVCs could be the returning gas, and for a while this idea was the leading explanation for their origin.

Going Out with the Tide

NEITHER OORT'S HYPOTHESIS nor the fountain model, however, could explain all characteristics of all HVCs. The problem was further complicated by the discovery in the early 1970s of the Magellanic Stream, a filament of gas that arcs around the galaxy. The stream follows the orbits of the Large and Small Magellanic Clouds, two small companion galaxies that revolve around the Milky Way like moons around a planet. Although astronomers usually reserve the term "cloud" for a clump of gas or dust, these full-fledged galaxies containing billions of stars are so named because they resemble clouds in the night sky. They are currently about 150,000 light-years from our galaxy, about as close as they ever get on their highly elongated paths.

The stream behaves in many ways like a string of HVCs. Much of it moves at velocities that are incompatible with normal galactic rotation. Yet it cannot be explained by the two hypotheses described above. According to the most detailed model of the stream, published in 1996 by Lance T. Gardiner of Sun Moon University in South Korea and Masafumi Noguchi of Tohoku University in Japan, the filament is our galaxy's version of the tidal streams that astronomers see around many other galaxies. When the Magellanic Clouds made their previous close approach to the Milky Way, 2.2 billion years ago, the combined force of our galaxy and the Large Magellanic Cloud ripped off some of the gas in the outer parts of the Small Magellanic Cloud. About half the gas was decelerated and lagged behind the Magellanic Clouds in their orbits. The other half was accelerated and pulled ahead of the galaxies, forming what is called a leading arm. A similar process may also be ripping apart some of the Milky Way's other satellite galaxies [*see box on page 45*].

Our galaxy is tearing apart its satellite galaxies, and gas clouds are arriving from intergalactic space.

An alternative model ascribes the stream to frictional forces. If the Milky Way has a very extended corona (much bigger than the one proposed by Spitzer), this corona could strip off gas from the Magellanic Clouds. In either model, however, the Magellanic Clouds have lost large amounts of gas, producing many of the HVCs.

Yet another twist in the saga of HVCs came in 1999, when Leo Blitz of the University of California at Berkeley and his collaborators suggested that they are much farther away than most of their colleagues thought possible. Instead of buzzing through the outskirts of the Milky Way, HVCs could be floating around in the Local Group of galaxies—a conglomeration of the Milky Way, Andromeda and some 40 smaller galaxies that occupies a volume of space roughly four million light-years across. In this case, HVCs would be remnants of the group's, rather than only our galaxy's, formation.

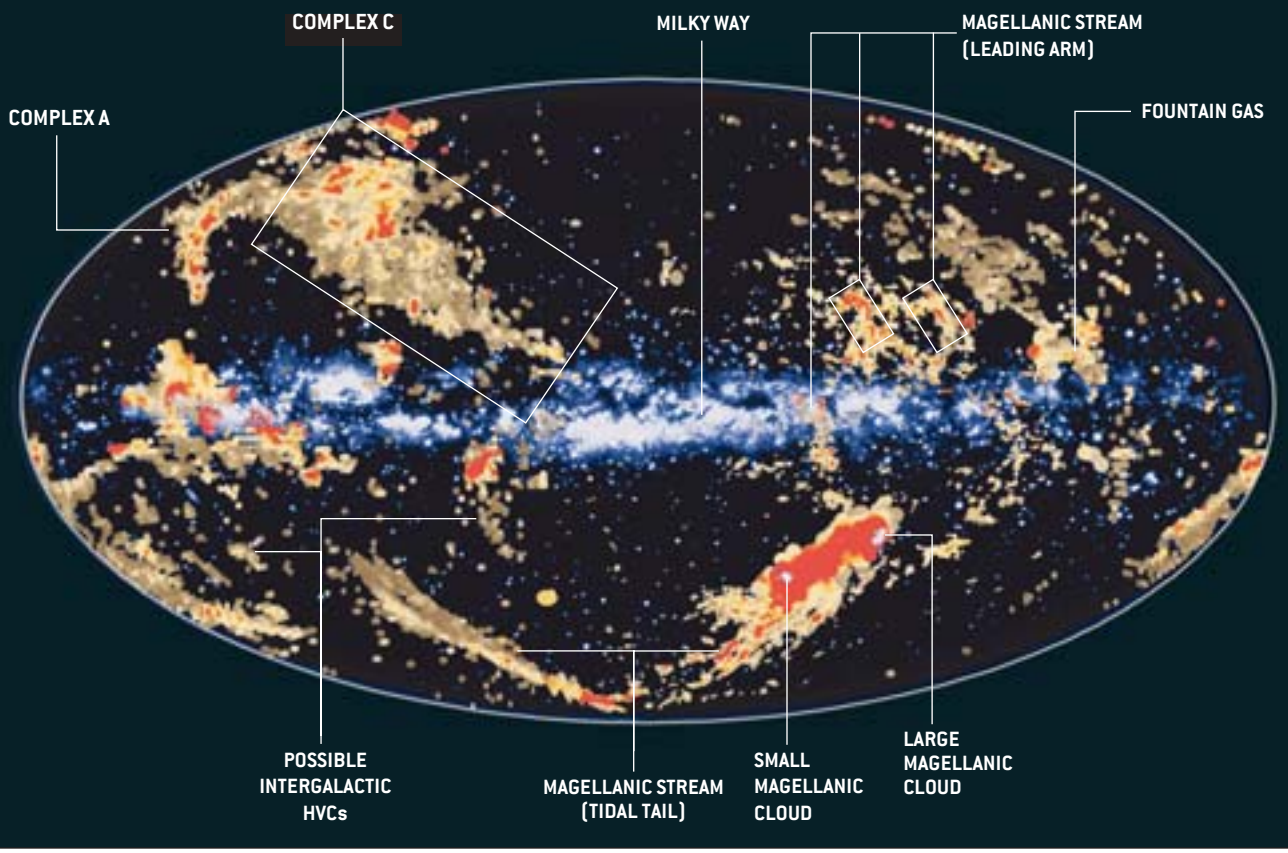
Similar ideas had been put forward more than 30 years ago and excluded because gas clouds should not be stable at the proposed distances. Blitz conjectured that HVCs are not, in fact, clouds of gas but clumps of dark matter with a small amount of gas mixed in. If so, HVCs are 10 times as massive as astronomers had assumed and therefore able to hold themselves together. An attractive feature of this hypothesis is that it alleviates what has become a major embarrassment for as-

THE AUTHORS

BART P. WAKKER and *PHILIPP RICHTER* are observers, primarily in the ultraviolet and radio bands of the electromagnetic spectrum. They joined forces to investigate high-velocity clouds in late 1999, when Richter took up a postdoctoral position at the University of Wisconsin–Madison, where Wakker was doing research. Wakker traces his interest in astronomy to the *Apollo 8* moonflight. He did his doctoral thesis on HVCs at the University of Groningen in the Netherlands, then spent five years at the University of Illinois before moving to Wisconsin in 1995. Richter received his Ph.D. from the University of Bonn in Germany, where he studied diffuse molecular gas in the Magellanic Clouds and the halo of the Milky Way. After leaving Wisconsin in 2002, he worked at the Arcetri Astrophysical Observatory in Florence, Italy, and recently returned to Bonn.

CLOUDY SKY

MAP OF GALACTIC GAS combines radio observations of neutral hydrogen (*colored splotches*) with a visible-light image of the Milky Way (*white*). The map depicts our sky, reprojected so that the galactic disk runs across the middle; the core of the galaxy lies at the center. High-velocity clouds of hydrogen, such as complexes A and C, are located above and below the disk.



INGRID KALLICK; SOURCE: BART WAKKER (*radio*), LUND OBSERVATORY (*visible*)

tronomers—namely, that models of galaxy formation predict more leftover dark matter halos than have been found [see "The Life Cycle of Galaxies," by Guinevere Kauffmann and Frank van den Bosch; SCIENTIFIC AMERICAN, June 2002]. HVCs could be the missing leftovers.

Getting Warmer

THUS, ASTRONOMERS ENTERED the third millennium with four hypotheses for HVCs: fresh gas left over from galaxy formation, gas cycling through a galactic fountain, shreds of the Magellanic Clouds, or intergalactic amalgams of gas and dark matter. Each hypothesis had bits and pieces of supporting evidence, but researchers needed new data to break the deadlock, and since the mid-1990s they have made major progress.

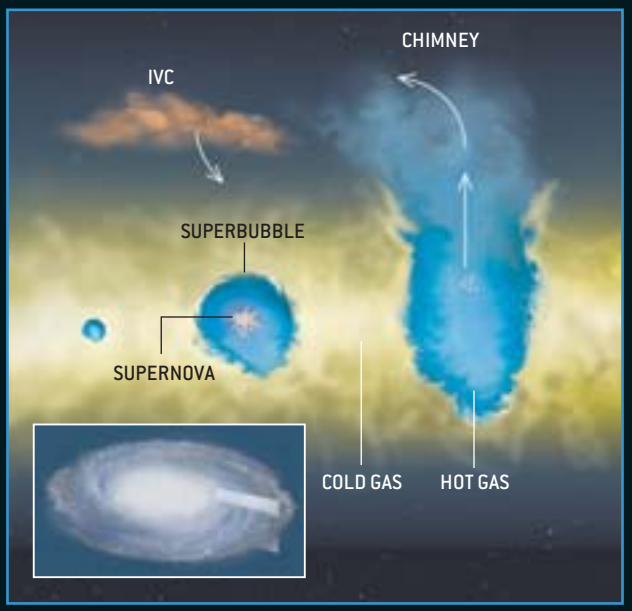
First, they have completed an all-sky survey for radio emission from neutral hydrogen, which traces gas at temperatures of about 100 kelvins. Aad Hulsbosch of the University of Nijmegen and one of us (Wakker), using the Dwingeloo radio telescope in the Netherlands, finished the northern half of this survey in 1988. Ricardo Morras and his collaborators, using the Villa Elisa radio telescope in Argentina, covered the southern sky in 2000 [*see illustration above*]. A third survey, by Dap Hartmann and Butler Burton of Leiden Observatory, became available in 1997 and mapped all of the Milky Way's neutral hydrogen, including both HVCs and IVCs.

A further contribution came from observations in visible light, made by instruments such as the Wisconsin Hydrogen-Alpha Mapper [see "The Gas between the Stars," by Ronald J. Reynolds; SCIENTIFIC AMERICAN, January 2002]. Although neutral hydrogen does not shine at visible wavelengths, ionized gas does, and the outer parts of HVCs are ionized by far-ultraviolet light from the Milky Way and other objects. The radiation also heats the clouds' exteriors to 8,000 kelvins. The amount of visible light is a measure of the intensity of the radiation field surrounding the HVC, which in turn depends on its distance from the galactic disk. Thus, these observations offer a rough way to estimate the location of HVCs.

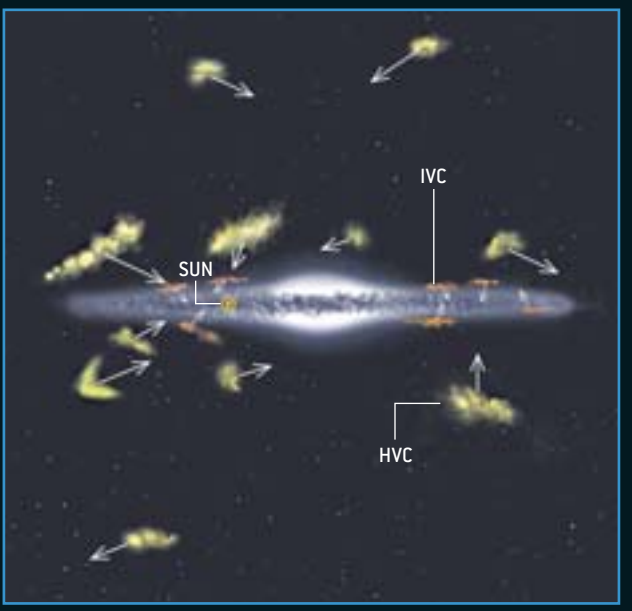
The most important progress has come from observations of spectral absorption lines in HVCs. Instead of looking for light given off by the gas, this work analyzes light blocked by the gas—specific atoms filter out specific wavelengths of light. Three observatories have made the largest contributions: the

FOUR PROCESSES THAT SHAPE THE GALAXY

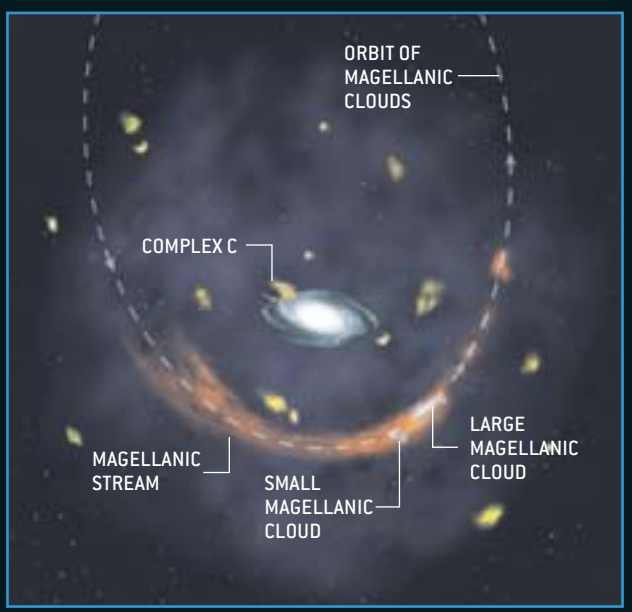
GALACTIC FOUNTAIN: Intermediate-velocity clouds are probably the return leg of a vast cycle of gas. Clusters of supernova explosions generate bubbles of hot gas (*blue*) that break through the surrounding cold gas (*yellow*) and feed a hot corona. Chunks of the gas cool and fall back to the disk.



GAS INFALL: Many of the high-velocity clouds (*yellow*) are gas raining onto the Milky Way, continuing its formation nearly 12 billion years after it started. Such gas could provide fresh fuel for star formation. Observationally, they are easily confused with the intermediate-velocity clouds (*orange*).



GALACTIC CANNIBALIZATION: The Milky Way is ripping gas from two of its satellite galaxies, the Large and Small Magellanic Clouds. Along their orbits astronomers see the Magellanic Stream (*orange*). Other, unrelated high-velocity clouds (*yellow*), possibly condensing out of a hot corona, float in the same space.



INTERGALACTIC REPLENISHMENT: The Milky Way and Andromeda galaxies may be embedded in a massive sea of hot intergalactic gas (*blue*). Out of this gas, cold clumps may condense and get captured by the galaxies—forming new high-velocity clouds that eventually fall in. This model is still uncertain.

INGRID KALLICK

La Palma Observatory in the Canary Islands, the Hubble Space Telescope and the Far Ultraviolet Spectroscopic Explorer (FUSE), launched in 1999.

Using such data, Laura Danly, now at the University of Denver, and her collaborators put limits on the distance to an IVC 11 years ago. More recently, Hugo van Woerden of the University of Groningen in the Netherlands and his collaborators gauged the distance to an HVC for the first time [*see box on next page*]. Meanwhile we and our colleagues measured the chemical composition of the clouds, rounding out the information needed to distinguish among the various hypotheses.

A very warm component of HVCs emerged in data from FUSE. This satellite detected absorption by highly ionized oxygen (specifically, oxygen atoms that have lost five of their eight electrons), which implies a temperature of about 300,000 kelvins. Such temperatures can occur where cool (100 kelvins) neutral hydrogen comes into contact with extremely hot (one million kelvins) gas. Alternatively, the presence of gas at 300,000 kelvins shows that the extremely hot gas is cooling down. Together with Blair D. Savage of the University of Wisconsin–Madison and Kenneth Sembach of the Space Telescope Science Institute in Baltimore, we have traced this component of HVCs.

Complex Behavior

HAVING EXPLORED ALL these new data, we can now present a coherent picture of HVCs. We begin with two of the largest, known as complexes A and C, which were the first HVCs discovered back in 1963. Complex A is 25,000 to 30,000 light-years away, which clearly puts it in the galactic halo. The distance to complex C remains uncertain: at least 14,000 light-years but probably no more than 45,000 light-years above the galactic plane.

The two clouds are deficient in heavy elements, having about a tenth of the concentration found in the sun. The nitrogen content of complex C is especially low, about $1/50$ of the sun's. The paucity of nitrogen suggests that the heavy elements came mostly from high-mass stars, which produce less nitrogen relative to other heavy elements than low-mass stars do. In fact, recent models of the young universe predict that the earliest stars are uncommonly heavy. Complex C thus appears to be a fossil from the ancient universe.

Brad Gibson of Swinburne University in Melbourne, Australia, has looked at a different part of complex C and measured a heavy-element concentration that was twice as high as our earlier results. This variation in composition indicates that complex C has begun to mix with other gas clouds in the galactic halo, which have higher concentrations of heavy elements. In addition, Andrew Fox and his collaborators at Wisconsin used the data for highly ionized oxygen and other ions to show that the gas at 300,000 kelvins in complex C represents an interface between hot and cool gas. We seem to be catching complex C in the process of assimilating into the galaxy.

Clouds such as complexes A and C thus provide the first direct evidence for the infall of fresh gas. Complex C brings between 0.1 and 0.2 solar mass of new material every year, and complex A represents about half of that. This is 10 to 20 percent of the total needed to dilute galactic gas and account for the chemical composition of stars. Other HVCs may make up the remainder. It is somewhat unclear, though, whether the ultimate source of this gas is a remnant halo (as proposed by Oort), deep intergalactic space, or even a small dwarf galaxy that the Milky Way swallowed.

A Multiplicity of Origins

THE RESULTS ELIMINATE three of the hypotheses for the origin of complexes A and C. The fountain hypothesis implies that they originate in the disk and have a composition similar to that of the sun, which is not the case. The Magellanic Stream hypothesis also gets the heavy-element content wrong. Finally, the dark matter hypothesis fails because these two HVCs do not lie in intergalactic space. It turns out, however, that these three explanations are not completely incorrect. We simply have to look elsewhere to find where they apply.

For a long time, IVCs stood in the shadow of the more

Conscious of Streams

MOST OF THE MILKY WAY is as thoroughly mixed as a well-stirred gravy. Two stars that originated in the same region may be located in completely different parts of the sky today. But during the past few years, astronomers have found groups of stars that move in unison, forming what they call stellar streams. They are like lumps that a cook has just thrown into a pot but that have not had time to mix in.

The streams are believed to be the remnants of satellite galaxies of the Milky Way that were torn apart by tides, the same process that formed some of the high-velocity clouds. The streams thus trace a flow of stars from dwarf galaxies to the Milky Way. They differ from the Magellanic Stream, which consists of gas rather than stars. They represent independent evidence for the ongoing growth of our galaxy.

One spectacular example is a stream of stars being pulled off the Sagittarius dwarf spheroidal galaxy, which was discovered in 1994 by Rodrigo Ibata of the Strasbourg Observatory in France and his colleagues [*see artist's conception above*]. More recently, several other stellar streams were found in the data gathered by the Sloan Digital Sky Survey, a program to map a large portion of the sky systematically. One may be related to the Canis Major dwarf galaxy, which Ibata, Nicolas Martin of Strasbourg and their collaborators discovered two months ago. Over the past two billion years, this galaxy has been stretched into a spiraling ring of stars along the galactic plane. *—B.W. and P.R.*

RON MILLER

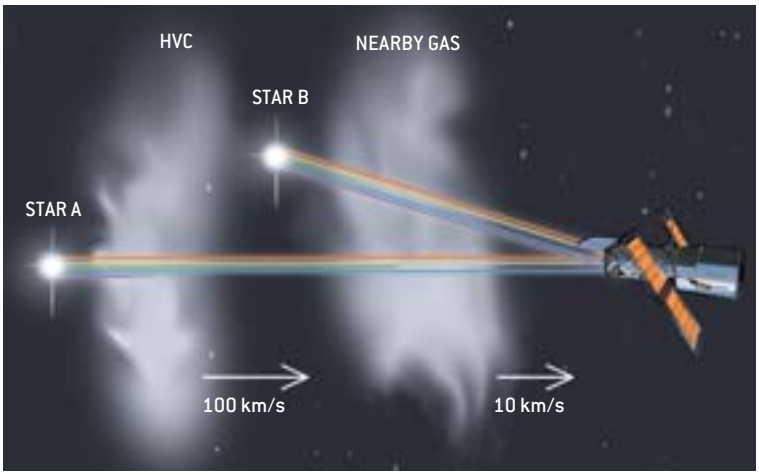
PEEKING BEHIND THE CLOUDS

HIGH-VELOCITY CLOUDS stymied astronomers for decades because their distances and compositions were uncertain. The only known technique to measure these properties is the absorption-line method. Stars and galaxies located behind HVCs act as bulbs that shine through the clouds from behind. Most of the light passes through the clouds, but a few wavelengths are absorbed, allowing properties of the clouds to be measured.

If the spectrum of a star contains absorption lines, it means a cloud must be sitting between us and the star. The distance to the star sets an upper limit on the distance to the cloud. Conversely, the lack of an absorption line implies a lower limit on the distance to the cloud. These limits assume that other factors can be ruled out: uncertainties in the stellar distance, lack of enough heavy elements to produce a detectable absorption line, and absorption lines created by material within the star itself.

To determine HVC distances, the most useful lightbulbs are so-called RR Lyrae variables and blue horizontal branch (BHB) stars. They are numerous, their distances can be measured accurately, and few of their spectral lines overlap with those of the clouds. In principle, the absorption lines of any element could be used. To determine the heavy element content, however, the best measurements rely on the spectral lines of neutral oxygen and ionized sulfur. These lines lie in the ultraviolet part of the spectrum, requiring properly equipped satellites such as the Hubble Space Telescope or Far Ultraviolet Spectroscopic Explorer (FUSE). In this case, the best lightbulbs are distant active galaxies such as quasars, because they often have featureless spectra and are brighter ultraviolet emitters than stars.

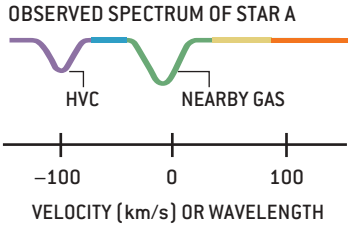
A single star or galaxy can illuminate more than one gas cloud. Each cloud moves at a different velocity, so each absorbs at a slightly different wavelength because of the Doppler effect. To distinguish the clouds requires a spectrometer with high spectral resolution, which in turn requires a large telescope. *—B.W. and P.R.*



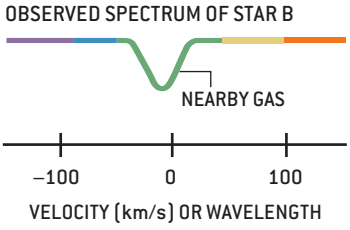
CROSS-SECTIONAL VIEW (not to scale)

BIGGEST HASSLE in studying high-velocity clouds is to measure their distances. The best available technique is indirect and approximate. Consider an HVC that lies between two stars, labeled A and B. Another, slower-moving cloud of gas lies between us and Star B.

VIEW FROM EARTH



WHEN ASTRONOMERS take a spectrum of Star A, they notice two absorption lines and infer the presence of two clouds. The lines occur at different wavelengths because of the different cloud velocities.



STAR B has only a single absorption line, so it must lie in front of the HVC. Thus, the two stars, whose distances can be estimated by independent means, place an upper and lower limit on the cloud distance.

flashy and mysterious HVCs. Several teams have now measured their composition, and it matches that of gas in the disk. Moreover, IVCs lie some 4,000 light-years above the plane, the place where fountains would operate. Both facts indicate that they, rather than HVCs, represent the return flow of a fountain.

A piece of corroborating evidence has been the detection of hydrogen molecules in IVCs. Forming these molecules in space requires interstellar dust grains, which will be sufficiently abundant only if the ambient gas is chemically enriched. In line with this idea, molecular hydrogen was not found in complex C. Thus, IVCs are recycled gas from within the galaxy, whereas HVCs are primarily gas from outside.

As for the Magellanic Stream hypothesis, at least one HVC does seem to be a castoff from the stream. Its composition is

INGRID KALLICK

similar to that of the Small Magellanic Cloud, as Limin Lu and his co-workers at Wisconsin found in 1998. The HVC is located in the leading arm of the stream, meaning that whatever pulled it off the Small Magellanic Cloud also accelerated it. Frictional forces cannot do that; only tidal forces can. Lu's discovery finally settles the question of the origin of the stream.

Frictional forces may still be important, however. FUSE found highly ionized oxygen associated with the Magellanic Stream, suggesting that it, too, is embedded in hot gas. The galactic corona must therefore extend much farther out than was originally proposed by Spitzer—out to a few hundred thousand light-years, rather than a few thousand. This corona is not dense enough to strip gas from the Magellanic Clouds, but once the gas has been drawn out by tidal forces, friction with the corona causes it to decelerate, slowly rain down on the galaxy and contribute to the growth of the Milky Way.

clouds would weigh at most 10 million solar masses, and rather than roaming throughout the Local Group, most would stay within half a million light-years of the main galaxies.

Although neutral HVCs do not appear to be dispersed throughout the Local Group, other types of high-velocity gas may be. The highly ionized gas in one HVC lies far outside the Milky Way. FUSE has also discovered high-velocity, highly ionized oxygen on its own, without any neutral gas. Similar clouds of hot gas have been found elsewhere in the universe by Todd M. Tripp of Princeton and his co-workers. This hot gas may constitute a filament running through intergalactic space. Such filaments show up in simulations of the broad-scale evolution of the cosmos [see "The Emptiest Places," by Evan Scannapieco, Patrick Petitjean and Tom Broadhurst; SCIENTIFIC AMERICAN, October 2002], and the total amount of matter in these filaments may be larger than that in all galaxies combined,

Filaments of hot intergalactic gas form a reservoir that the Milky Way can draw on to make new stars.

Similarly, the dark matter hypothesis, although it does not explain complexes A and C, may fit into the broader scheme of things. Blitz originally proposed that the intergalactic HVCs weigh 10 million to 100 million solar masses. Yet such clouds have not been detected in nearby galaxy groups similar to the Local Group, even though observations are now sensitive enough to do so. Furthermore, the hypothesis predicts that visible-light emission from HVCs should be too faint to detect, but in almost all cases that this emission has been looked for, it has been detected. Finally, theoretical arguments show that if the HVCs are distant, they must be either fully ionized or extremely massive, and both options are inconsistent with observations. It thus appears that HVCs are not the predicted population of dark matter clouds.

Robert Braun of Dwingeloo Observatory and Butler Burton and Vincent de Heij of Leiden instead propose that the Milky Way and Andromeda galaxies are surrounded by several hundred small clouds made mostly of dark matter and ionized gas, with a small fraction of neutral hydrogen. These forming a reservoir that the Milky Way can draw on to make new stars.

The HVCs surrounding the Milky Way remind us that we are living in a galaxy that is still forming and evolving. Originally our galaxy was surrounded by many smaller satellite galaxies and a lot of leftover gas. Over the past several billion years, it has incorporated most of those satellites. It may also have accreted much of the pristine gas from its intergalactic environs, and plenty of gas may still lie out there. Gas is still trickling in, taking the form of HVCs. At the same time, the galaxy expels gas loaded with heavy elements into its halo and maybe even into intergalactic space.

Within the next 10 or so billion years, more satellite galaxies will merge with the Milky Way, forming more of the stellar streams now being discovered in the halo. Our galaxy is on a collision course with the Andromeda galaxy. We cannot tell exactly how the Milky Way, or what is left of it, will look in the distant future, but we know that its formation has not come to an end yet. SA

MORE TO EXPLORE

High-Velocity Clouds. Bart P. Wakker and Hugo van Woerden in *Annual Review of Astronomy and Astrophysics,* Vol. 35, pages 217–266; September 1997.

A Confirmed Location in the Galactic Halo for the High-Velocity Cloud "Chain A." Hugo van Woerden, Ulrich J. Schwarz, Reynier F. Peletier, Bart P. Wakker and Peter M. W. Kalberla in *Nature,* Vol. 400, pages 138–141; July 8, 1999. Available online at **arXiv.org/abs/astro-ph/9907107**

Accretion of Low-Metallicity Gas by the Milky Way. Bart P. Wakker, J. Chris Howk, Blair D. Savage, Hugo van Woerden, Steve L. Tufte, Ulrich J. Schwarz, Robert Benjamin, Ronald J. Reynolds, Reynier F. Peletier and Peter M. W. Kalberla in *Nature,* Vol. 402, No. 6760; pages 388–390; November 25, 1999.

The Formation and Evolution of the Milky Way. Cristina Chiappini in *American Scientist,* Vol. 89, No. 6, pages 506–515; November–December 2001.

A Far Ultraviolet Spectroscopic Explorer Survey of Molecular Hydrogen in Intermediate-Velocity Clouds in the Milky Way Halo. P. Richter, B. P. Wakker, B. D. Savage and K. R. Sembach in *Astrophysical Journal,* Vol. 586, No. 1, pages 230–248; March 20, 2003. **arXiv.org/abs/astro-ph/0211356**

Highly Ionized High-Velocity Gas in the Vicinity of the Galaxy. K. R. Sembach, B. P. Wakker, B. D. Savage, P. Richter, M. Meade, J. M. Shull, E. B. Jenkins, G. Sonneborn and H. W. Moos in *Astrophysical Journal,* Supplement Series, Vol. 146, No. 1, pages 165–208; May 2003. **arXiv.org/abs/astro-ph/0207562**

INNER WORLD of people with schizophrenia is often confused, punctuated by alien voices, paranoia and illogical thoughts.

Decoding Schizophrenia

A fuller understanding of signaling in the brain of people with this disorder offers new hope for improved therapy

By Daniel C. Javitt and Joseph T. Coyle

COOP PHOTO ART COLLECTIVE

Today the word "schizophrenia" brings to mind

such names as John Nash and Andrea Yates. Nash, the subject of the Oscar-winning film *A Beautiful Mind,* emerged as a mathematical prodigy and eventually won a Nobel Prize for his early work, but he became so profoundly disturbed by the brain disorder in young adulthood that he lost his academic career and floundered for years before recovering. Yates, a mother of five who suffers from both depression and schizophrenia, infamously drowned her young children in a bathtub to "save them from the devil" and is now in prison.

The experiences of Nash and Yates are typical in some ways but atypical in others. Of the roughly 1 percent of the world's population stricken with schizophrenia, most remain largely disabled throughout adulthood. Rather than being geniuses like Nash, many show below-average intelligence even before they become symptomatic and then undergo a further decline in IQ when the illness sets in, typically during young adulthood. Unfortunately, only a minority ever achieve gainful employment. In contrast to Yates, fewer than half marry or raise families. Some 15 percent reside for long periods in state or county mental health facilities, and another 15 percent end up incarcerated for petty crimes and vagrancy. Roughly 60 percent live in poverty, with one in 20 ending up homeless. Because of poor social support, more individuals with schizophrenia become victims than perpetrators of violent crime.

Medications exist but are problematic. The major options today, called antipsychotics, stop all symptoms in only about 20 percent of patients. (Those lucky enough to respond in this way tend to function well as long as they continue treatment; too many, however, abandon their medicines over time, usually because of side effects, a desire to be "normal" or a loss of access to mental health care). Two thirds gain some relief from antipsychotics yet remain symptomatic throughout life, and the remainder show no significant response.

An inadequate arsenal of medications is only one of the obstacles to treating this tragic disorder effectively. Another is the theories guiding drug therapy. Brain cells (neurons) communicate by releasing chemicals called neurotransmitters that either excite or inhibit other neurons. For decades, theories of schizophrenia have focused on a single neurotransmitter: dopamine. In the past few years, though, it has become clear that a disturbance in dopamine levels is just a part of the story and that, for many, the main abnormalities lie elsewhere. In particular, suspicion has fallen on deficiencies in the neurotransmitter glutamate. Scientists now realize that schizophrenia affects virtually all parts of the brain and that, unlike dopamine, which plays an important role only in isolated regions, glutamate is critical virtually everywhere. As a result, investigators are searching for treatments that can reverse the underlying glutamate deficit.

Overview/*Schizophrenia*

- Scientists have long viewed schizophrenia as arising out of a disturbance in a particular brain system—one in which brain cells communicate using a signaling chemical, or neurotransmitter, called dopamine.
- Yet new research is shifting emphasis from dopamine to another neurotransmitter, glutamate. Impaired glutamate signaling appears to be a major contributor to the disorder.
- Drugs are now in development to treat the illness based on this revised understanding of schizophrenia's underlying causes.

Multiple Symptoms

TO DEVELOP better treatments, investigators need to understand how schizophrenia arises—which means they need to account for all its myriad symptoms. Most of these fall into categories termed "positive," "negative" and "cognitive." Positive symptoms generally imply occurrences beyond normal experience; negative symptoms generally connote diminished experience. Cognitive, or "disorganized," symptoms refer to difficulty maintaining a logical, coherent flow of conversation, maintaining attention, and thinking on an abstract level.

The public is most familiar with the positive symptoms, particularly agitation, paranoid delusions (in which people feel conspired against) and hallucinations, commonly in the form of spoken voices. Command hallucinations, where voices tell people to hurt themselves or others, are an especially ominous sign: they can be difficult to resist and may precipitate violent actions.

The negative and cognitive symptoms are less dramatic but more pernicious. These can include a cluster called the 4 A's: *a*utism (loss of interest in other people or the surroundings), *a*mbivalence (emotional withdrawal), blunted *a*ffect (manifested by a bland and unchanging facial expression), and the cognitive problem of loose *a*ssociation (in which people join thoughts without clear logic, frequently jumbling words together into a meaningless word salad). Other common symptoms include a lack of spontaneity, impoverished speech, difficulty establishing rapport and a slowing of movement. Apathy and disinterest especially can

cause friction between patients and their families, who may view these attributes as signs of laziness rather than manifestations of the illness.

When individuals with schizophrenia are evaluated with pencil-and-paper tests designed to detect brain injury, they show a pattern suggestive of widespread dysfunction. Virtually all aspects of brain operation, from the most basic sensory processes to the most complex aspects of thought are affected to some extent. Certain functions, such as the ability to form new memories either temporarily or permanently or to solve complex problems, may be particularly impaired. Patients also display difficulty solving the types of problems encountered in daily living, such as describing what friends are for or what to do if all the lights in the house go out at once. The inability to handle these common problems, more than anything else, accounts for the difficulty such individuals have in living independently. Overall, then, schizophrenia conspires to rob people of the very qualities they need to thrive in society: personality, social skills and wit.

PERCEIVING FRAGMENTS as parts of a whole can be difficult for people with schizophrenia. When normal subjects view fractured images like those above in sequence, they identify the object quickly, but schizophrenic patients often cannot make that leap swiftly.

Beyond Dopamine

THE EMPHASIS on dopamine-related abnormalities as a cause of schizophrenia emerged in the 1950s, as a result of the fortuitous discovery that a class of medication called the phenothiazines was able to control the positive symptoms of the disorder. Subsequent studies demonstrated that these substances work by blocking the functioning of a specific group of chemical-sensing molecules called dopamine D2 receptors, which sit on the surface of certain nerve cells and convey dopamine's signals to the cells' interior. At the same time, research led by the recent Nobel laureate Arvid Carlsson revealed that amphetamine, which was known to induce hallucinations and delusions in habitual abusers, stimulated dopamine release in the brain. Together these two findings led to the "dopamine theory," which proposes that most symptoms of schizophrenia stem from excess dopamine release in important brain regions, such as the limbic system (thought to regulate emotion) and the frontal lobes (thought to regulate abstract reasoning).

Over the past 40 years, both the strengths and limitations of the theory have become apparent. For some patients, especially those with prominent positive symptoms, the theory has proved robust, fitting symptoms and guiding treatment well. The minority of those who display only positive manifestations frequently function quite well—holding jobs, having families and suffering relatively little cognitive decline over time—if they stick with their medicines.

Yet for many, the hypothesis fits poorly. These are the people whose symptoms come on gradually, not dramatically, and in whom negative symptoms overshadow the positive. The sufferers grow withdrawn, often isolating themselves for years. Cognitive functioning is poor, and patients improve slowly, if at all, when treated with even the best existing medications on the market.

Such observations have prompted some researchers to modify the dopamine hypothesis. One revision suggests, for example, that the negative and cognitive symptoms may stem from *reduced* dopamine levels in certain parts of the brain, such as the frontal lobes, and increased dopamine in other parts of the brain, such as the limbic system. Because dopamine receptors in the frontal lobe are primarily of the D1 (rather than D2) variety, in-

THE AUTHORS

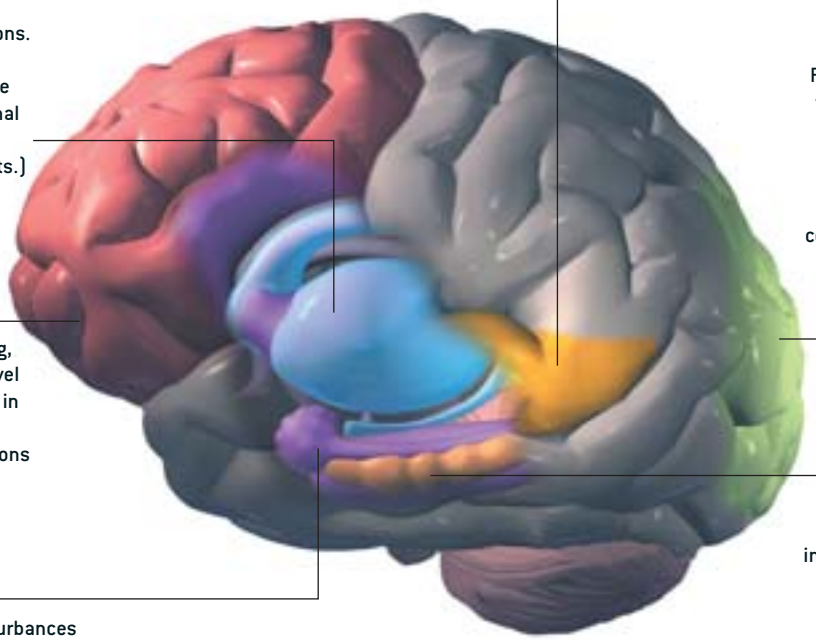
DANIEL C. JAVITT and *JOSEPH T. COYLE* have studied schizophrenia for many years. Javitt is director of the Program in Cognitive Neuroscience and Schizophrenia at the Nathan Kline Institute for Psychiatric Research in Orangeburg, N.Y., and professor of psychiatry at the New York University School of Medicine. His paper demonstrating that the glutamate-blocking drug PCP reproduces the symptoms of schizophrenia was the second-most cited schizophrenia publication of the 1990s. Coyle is Eben S. Draper Professor of Psychiatry and Neuroscience at Harvard Medical School and also editor in chief of the *Archives of General Psychiatry*. Both authors have won numerous awards for their research. Javitt and Coyle hold independent patents for use of NMDA modulators in the treatment of schizophrenia, and Javitt has significant financial interests in Medifoods and Glytech, companies attempting to develop glycine and D-serine as treatments for schizophrenia.

THE BRAIN IN SCHIZOPHRENIA

MANY BRAIN REGIONS and systems operate abnormally in schizophrenia, including those highlighted below. Imbalances in the neurotransmitter dopamine were once thought to be the prime cause of schizophrenia. But new findings suggest that impoverished signaling by the more pervasive neurotransmitter glutamate—or, more specifically, by one of glutamate's key targets on neurons (the NMDA receptor)—better explains the wide range of symptoms in this disorder.

BASAL GANGLIA
Involved in movement and emotions and in integrating sensory information. Abnormal functioning in schizophrenia is thought to contribute to paranoia and hallucinations. (Excessive blockade of dopamine receptors in the basal ganglia by traditional antipsychotic medicines leads to motor side effects.)

AUDITORY SYSTEM
Enables humans to hear and understand speech. In schizophrenia, overactivity of the speech area (called Wernicke's area) can create auditory hallucinations—the illusion that internally generated thoughts are real voices coming from the outside.



OCCIPITAL LOBE
Processes information about the visual world. People with schizophrenia rarely have full-blown visual hallucinations, but disturbances in this area contribute to such difficulties as interpreting complex images, recognizing motion, and reading emotions on others' faces.

FRONTAL LOBE
Critical to problem solving, insight and other high-level reasoning. Perturbations in schizophrenia lead to difficulty in planning actions and organizing thoughts.

HIPPOCAMPUS
Mediates learning and memory formation, intertwined functions that are impaired in schizophrenia.

LIMBIC SYSTEM
Involved in emotion. Disturbances are thought to contribute to the agitation frequently seen in schizophrenia.

DIFFERENT NEUROTRANSMITTERS, SAME RESULTS

SOME SCIENTISTS have proposed that too much dopamine leads to symptoms emanating from the basal ganglia and that too little dopamine leads to symptoms associated with the frontal cortex. Insufficient glutamate signaling could produce those same symptoms, however.

IN THE REST OF THE CORTEX, glutamate is prevalent, but dopamine is largely absent.

IN THE FRONTAL CORTEX, where dopamine promotes cell firing (by acting on D1 receptors), glutamate's stimulatory signals amplify those of dopamine; hence, a shortage of glutamate would decrease neural activity, just as if too little dopamine were present.

IN THE BASAL GANGLIA, where dopamine normally inhibits cell firing (by acting on D2 receptors on nerve cells), glutamate's stimulatory signals oppose those of dopamine; hence, a shortage of glutamate would increase inhibition, just as if too much dopamine were present.

ALFRED T. KAMAJIAN

vestigators have begun to search, so far unsuccessfully, for medications that stimulate D1 receptors while inhibiting D2s.

In the late 1980s researchers began to recognize that some pharmaceuticals, such as clozapine (Clozaril), were less likely to cause stiffness and other neurologic side effects than older treatments, such as chlorpromazine (Thorazine) or haloperidol (Haldol), and were more effective in treating persistent positive and negative symptoms. Clozapine, known as an atypical antipsychotic, inhibits dopamine receptors less than the older medications and affects the activity of various other neurotransmitters more strongly. Such discoveries led to the development and wide adoption of several newer atypical antipsychotics based on the actions of clozapine (certain of which, unfortunately, now turn out to be capable of causing diabetes and other unexpected side effects). The discoveries also led to the proposal that dopamine was not the only neurotransmitter disturbed in schizophrenia; others were involved as well.

Theories focusing largely on dopamine are problematic on additional grounds. Improper dopamine balance cannot account for why one individual with schizophrenia responds almost completely to treatment, whereas someone else shows no apparent response. Nor can it explain why positive symptoms respond so much better than negative or cognitive ones do. Finally, despite decades of research, investigations of dopamine have yet to uncover a smoking gun. Neither the enzymes that produce this neurotransmitter nor the receptors to which it binds appear sufficiently altered to account for the panoply of observed symptoms.

The Angel Dust Connection

IF DOPAMINE CANNOT account well for schizophrenia, what is the missing link? A critical clue came from the effects of another abused drug: PCP (phencyclidine), also known as angel dust. In contrast to amphetamine, which mimics only the positive symptoms of the disease, PCP induces symptoms that resemble the full range of schizophrenia's manifestations: negative and cognitive and, at times, positive. These effects are seen not just in abusers of PCP but also in individuals given brief, low doses of PCP or ketamine (an anesthetic with similar effects) in controlled drug-challenge trials.

Such studies first drew parallels between the effects of PCP and the symptoms of schizophrenia in the 1960s. They showed, for example, that individuals receiving PCP exhibited the same type of disturbances in interpreting proverbs as those with schizophrenia. More recent studies with ketamine have produced even more compelling similarities. Notably, during ketamine challenge, normal individuals develop difficulty thinking abstractly, learning new information, shifting strategies or placing information in temporary storage. They show a general motor slowing and reduction in speech output just like that seen in schizophrenia. Individuals given PCP or ketamine also grow withdrawn, sometimes even mute; when they talk, they speak tangentially and concretely. PCP and ketamine rarely induce schizophrenialike hallucinations in normal volunteers, but they exacerbate these disturbances in those who already have schizophrenia.

The ability of PCP and ketamine to induce a broad spectrum of schizophrenia-like symptoms suggests that these drugs replicate some key molecular disturbance in the brain of schizophrenic patients. At the molecular level the drugs impair the functioning of the brain signaling systems that rely on glutamate, the main excitatory neurotransmitter in the brain. More precisely, they block the action of a form of glutamate receptor known as the NMDA receptor, which plays a critical role in brain development, learning, memory and neural processing in general. This receptor also participates in regulating dopamine release, and blockade of NMDA receptors produces the same disturbances of dopamine function typically seen in schizophrenia. Thus, NMDA receptor dysfunction, by itself, can explain both negative and cognitive symptoms of schizophrenia as well as the dopamine abnormalities at the root of the positive symptoms.

One example of the research implicating NMDA receptors in schizophrenia relates to the way the brain normally pro-

Drug Classes in Development

Unless otherwise noted, the compounds mentioned below are in the early stages of human testing. Their developers or producers are listed in parentheses.

Stimulators of NMDA-type glutamate receptors aim to overcome the signaling deficits that apparently contribute to many schizophrenic symptoms.
Examples: Glycine (Medifoods), D-serine (Glytech). As natural substances, both of them are sold, but they remain under evaluation specifically for their value in treating schizophrenia.

Stimulators of AMPA-type glutamate receptors—also called ampakines—may improve some aspects of memory and cognition in people with schizophrenia.
Example: CX516 (Cortex Pharmaceuticals)

Modulators of another class of glutamate receptors—metabotropic receptors—can regulate glutamate release and potentially restore the balance between the activity of NMDA and AMPA receptors.
Example: LY354740 (Eli Lilly)

Inhibitors of glycine transport reduce glycine removal from synapses, which should increase signaling by NMDA receptors.
Example: GlyT-1 (NPS Pharmaceuticals and Janssen Pharmaceutica)

Stimulators of alpha 7 nicotinic receptors, the same receptors activated by the nicotine in cigarettes, indirectly stimulate the brain's NMDA receptors. Schizophrenics often smoke heavily, probably because the nicotine, acting on alpha 7 receptors, helps them to focus.
Example: DMXB-A (University of Colorado Health Sciences Center)

Stimulators of D1 dopamine receptors are being developed mainly for Parkinson's disease and have passed initial safety trials. They might also correct dopamine deficiencies in schizophrenia, but clinical trials for that purpose have not yet been performed.
Example: ABT-431 (Abbott Laboratories)

cesses information. Beyond strengthening connections between neurons, NMDA receptors amplify neural signals, much as transistors in old-style radios boosted weak radio signals into strong sounds. By selectively amplifying key neural signals, these receptors help the brain respond to some messages and ignore others, thereby facilitating mental focus and attention. Ordinarily, people respond more intensely to sounds presented infrequently than to those presented frequently and to sounds heard while listening than to sounds they make themselves while speaking. But people with schizophrenia do not respond this way, which implies that their brain circuits reliant on NMDA receptors are out of kilter.

If reduced NMDA receptor activity prompts schizophrenia's symptoms, what then causes this reduction? The answer remains unclear. Some reports show that people with schizophrenia have fewer NMDA receptors, although the genes that give rise to the receptors appear unaffected. If NMDA receptors are intact and present in proper amounts, perhaps the problem lies with a flaw in glutamate release or with a buildup of compounds that disrupt NMDA activity.

Some evidence supports each of these ideas. For instance, postmortem studies of schizophrenic patients reveal not only lower levels of glutamate but also higher levels of two compounds (NAAG and kynurenic acid) that impair the activity of NMDA receptors. Moreover, blood levels of the amino acid homocysteine are elevated; homocysteine, like kynurenic acid, blocks NMDA receptors in the brain. Overall, schizophrenia's pattern of onset and symptoms suggests that chemicals disrupting NMDA receptors may accumulate in sufferers' brains, although the research verdict is not yet in. Entirely different mechanisms may end up explaining why NMDA receptor transmission becomes attenuated.

New Treatment Possibilities

REGARDLESS OF what causes NMDA signaling to go awry in schizophrenia, the new understanding—and preliminary studies in patients—offers hope that drug therapy can correct the problem. Support for this idea comes from studies showing that clozapine, one of the most effective medications for schizophrenia identified to date, can reverse the behavioral effects of PCP in animals, something that older antipsychotics cannot do. Further, short-term trials with agents known to stimulate NMDA receptors have produced encouraging results. Beyond adding support to the glutamate hypothesis, these results have enabled long-term clinical trials to begin. If proved effective in large-scale tests, agents that activate NMDA receptors will become the first entirely new class of medicines developed specifically to target the negative and cognitive symptoms of the disorder.

The two of us have conducted some of those studies. When we and our colleagues administered the amino acids glycine and D-serine to patients with their standard medications, the subjects showed a 30 to 40 percent decline in cognitive and negative symptoms and some improvement in positive symptoms. Delivery of a medication, D-cycloserine, that is primarily used for treating tuberculosis but happens to cross-react with the NMDA receptor, produced similar results. Based on such findings, the National Institute of Mental Health has organized multicenter clinical trials at four hospitals to determine the effectiveness of D-cycloserine and glycine as therapies for schizophrenia; results should be available this year. Trials of D-serine, which is not yet approved for use in the U.S., are ongoing elsewhere with encouraging preliminary results as well. These agents have also been helpful when taken with the newest generation of atypical antipsychotics, which raises the hope that therapy can be developed to control all three major classes of symptoms at once.

None of the agents tested to date may have the properties needed for commercialization; for instance, the doses required may be too high. We and others are therefore exploring alternative avenues. Molecules that slow glycine's removal from brain synapses—known as glycine transport inhibitors—might enable glycine to stick around longer than usual, thereby increasing stimulation of

Steep Social Costs

SCHIZOPHRENIA, which affects about two million Americans, takes an enormous toll on society. Because it tends to arise in young adulthood and persist, it rings up a huge tally in health care bills and lost wages and ranks among the costliest illnesses in the U.S.

Treatment and strong social support enable some individuals to lead relatively productive and satisfying lives, but most are not so lucky. Fewer than a third can hold a job, and half of those do so only because they have intensive assistance. Men (who tend to become symptomatic earlier than women) usually do not marry, and women who tie the knot frequently enter into marriages that do not last. Because individuals with schizophrenia often isolate themselves and lack jobs, they constitute a disproportionate share of the chronically homeless population.

People with this disorder also have a high likelihood of becoming substance abusers. About 60 percent of symptomatic individuals smoke cigarettes, and half abuse alcohol, marijuana or cocaine. Such activities can lead to poor compliance with treatment and can exacerbate psychotic symptoms, increasing propensities toward violence. (Abstainers, however, behave no more violently than the general population.) Homelessness and substance abuse combine to land many with schizophrenia in prisons and county jails, where they often fail to get the treatment they require.

The grim figures do not end there: roughly 10 percent of people with schizophrenia commit suicide (usually during the illness's early stages), a higher rate than results from major depression. But there is one bright note: clozapine, the atypical antipsychotic introduced in 1989, has recently been shown to reduce the risk of suicide and substance abuse. Whether newer atypical agents exert a similar effect remains to be determined, however. *—D.C.J. and J.T.C.*

NMDA receptors. Agents that directly activate "AMPA-type" glutamate receptors, which work in concert with NMDA receptors, are also under active investigation. And agents that prevent the breakdown of glycine or D-serine in the brain have been proposed.

Many Avenues of Attack

SCIENTISTS INTERESTED in easing schizophrenia are also looking beyond signaling systems in the brain to other factors that might contribute to, or protect against, the disorder. For example, investigators have applied so-called gene chips to study brain tissue from people who have died, simultaneously comparing the activity of tens of thousands of genes in individuals with and without schizophrenia. So far they have determined that many genes important to signal transmission across synapses are less active in those with schizophrenia—but exactly what this information says about how the disorder develops or how to treat it is unclear.

Genetic studies in schizophrenia have nonetheless yielded intriguing findings recently. The contribution of heredity to schizophrenia has long been controversial. If the illness were dictated solely by genetic inheritance, the identical twin of a schizophrenic person would always be schizophrenic as well, because the two have the same genetic makeup. In reality, however, when one twin has schizophrenia, the identical twin has about a 50 percent chance of also being afflicted. Moreover, only about 10 percent of first-degree family members (parents, children or siblings) share the illness even though they have on average 50 percent of genes in common with the affected individual. This disparity suggests that genetic inheritance can strongly predispose people to schizophrenia but that environmental factors can nudge susceptible individuals into illness or perhaps shield them from it. Prenatal infections, malnutrition, birth complications and brain injuries are all among the influences suspected of promoting the disorder in genetically predisposed individuals.

Over the past few years, several genes have been identified that appear to increase susceptibility to schizophrenia. Interestingly, one of these genes codes for an enzyme (catechol-O-methyltransferase) involved in the metabolism of dopamine, particularly in the prefrontal cortex. Genes coding for proteins called dysbindin and neuregulin seem to affect the number of NMDA receptors in brain. The gene for an enzyme involved in the breakdown of D-serine (D-amino acid oxidase) may exist in multiple forms, with the most active form producing an approximately fivefold increase in risk for schizophrenia. Other genes may give rise to traits associated with schizophrenia but not the disease itself. Because each gene involved in schizophrenia produces only a small increase in risk, genetic studies must include large numbers of subjects to detect an effect and often generate conflicting results. On the other hand, the existence of multiple genes predisposing for schizophrenia may help explain the variability of symptoms across individuals, with some people perhaps showing the greatest effect in dopamine pathways and others evincing significant involvement of other neurotransmitter pathways.

OBJECTS often have hidden meanings to people with schizophrenia, who may hoard news items, pictures or other things that would seem useless to others. This wall is a re-creation.

Finally, scientists are looking for clues by imaging living brains and by comparing brains of people who have died. In general, individuals with schizophrenia have smaller brains than unaffected individuals of similar age and sex. Whereas the deficits were once thought to be restricted to areas such as the brain's frontal lobe, more recent studies have revealed similar abnormalities in many brain regions: those with schizophrenia have abnormal levels of brain response while performing tasks that activate not only the frontal lobes but also other areas of the brain, such as those that control auditory and visual processing. Perhaps the most important finding to come out of recent research is that no one area of the brain is "responsible" for schizophrenia. Just as normal behavior requires the concerted action of the entire brain, the disruption of function in schizophrenia must be seen as a breakdown in the sometimes subtle interactions both within and between different brain regions.

Because schizophrenia's symptoms vary so greatly, many investigators believe that multiple factors probably cause the syndrome. What physicians diagnose as schizophrenia today may prove to be a cluster of different illnesses, with similar and overlapping symptoms. Nevertheless, as researchers more accurately discern the syndrome's neurological bases, they should become increasingly skilled at developing treatments that adjust brain signaling in the specific ways needed by each individual. SA

MORE TO EXPLORE

Recent Advances in the Phencyclidine Model of Schizophrenia. D. C. Javitt and S. R. Zukin in *American Journal of Psychiatry,* Vol. 148, No. 10, pages 1301–1308; October 1991.

Efficacy of High-Dose Glycine in the Treatment of Enduring Negative Symptoms of Schizophrenia. U. Heresco-Levy, D. C. Javitt, M. Ermilov, C. Mordel, G. Silipo and M. Lichtenstein in *Archives of General Psychiatry,* Vol. 56, No. 1, pages 29–36; January 1999.

A Beautiful Mind: The Life of Mathematical Genius and Nobel Laureate John Nash. Sylvia Nasar. Touchstone Books, 2001.

The Emerging Role of Glutamate in the Pathophysiology and Treatment of Schizophrenia. D. C. Goff and J. T. Coyle in *American Journal of Psychiatry,* Vol. 158, No. 9, pages 1367–1377; September 2001.

Revolution #9. Directed by Tim McCann. Wellspring Media, 2001. VHS and DVD release, 2003.

COOP PHOTO ART COLLECTIVE

RFID

A Key to Automating Everything

Already common in security systems and tollbooths, radio-frequency identification tags and readers stand poised to take over many processes now accomplished by human toil

By Roy Want

TALK TO ME: When interrogated by computer-controlled readers, widely dispersed RFID tags (*highlighted*) could indicate, for instance, that a bread bag has been thrown into the garbage and you need a new loaf.

"The most profound technologies are those that disappear. They weave themselves into the fabric of everyday life until they are indistinguishable from it."

Thirteen years ago, in an article for SCIENTIFIC AMERICAN, the late Mark Weiser, then my colleague at Xerox PARC, outlined his bold vision of "ubiquitous computing": small computers would be embedded in everyday objects all around us and, using wireless connections, would respond to our presence, desires and needs without being actively manipulated. This network of mobile and fixed devices would do things for us automatically and so invisibly that we would notice only their effects. Weiser called such systems "calm technology," because they would make it easier for us to focus on our work and other activities, instead of demanding that we interact with and control them, as the typical PC does today.

In a home equipped with this kind of technology, readers strategically placed in the bedroom, the bathroom door frame, the stairwell and the refrigerator would detect the identifying data in microchip tags sewn into your clothes and embedded in the packaging of foods and send the data to a home computer, which would take action based on that information.

The computer would notice as you got out of bed in the morning and would switch on the coffeemaker. As you entered the bathroom, the shower would come on, adjusted to your favorite temperature. When you started down the stairs, the preloaded toaster would heat up so that your breakfast would be done just the way you like it. When you opened the refrigerator, the appliance would remind you that you were out of milk and that the tub of coleslaw inside had passed its expiration date and should be thrown out.

Today systems based on radio-frequency identification (RFID) technology are helping to move Weiser's vision closer to reality. These systems consist of tags (small silicon chips that contain identifying data and sometimes other information) and of readers that automatically receive and decode that data.

The responsive RFID home—and conference room, office building and car—are still far away, but RFID technology is already in limited use. The tags, often as small as a grain of rice, now hide in ID cards and wristbands, windshield-mounted toll tags, gasoline quick-purchase tokens, and electronic ear tags for livestock, and they have begun to appear in auto key-chain antitheft devices, toys (Hasbro *Star Wars* figures) and other products. They have also timed runners in road races, and last year a company in Mexico began a service to implant tags under the skin of children as an antikidnapping measure.

In the near term, RFID tags will probably be found in airline luggage labels (British Airways has conducted extensive trials), and they may eventually be embedded in paper currency to inhibit counterfeiters and enable governments to track the movement of cash. (Hitachi in Japan recently announced that it has developed tags minute enough for this application.) Meanwhile the retail, security, transportation, manufacturing and shipping industries are all testing or starting to implement sophisticated RFID applications.

But the RFID revolution is not without a downside: the technology's growth raises important social issues, and as RFID systems proliferate, we will be forced to address new problems related to privacy, law and ethics. Controversy has already erupted: in mid-2003 two major retailers—Wal-Mart in the U.S. and the international clothing maker Benetton—canceled large-scale tests of in-store RFID-centered inventory control systems apparently partly as a response to public reactions that raised the specter of wholesale monitoring of citizens through tags embedded in consumer products [*see box on page 61*].

The Inside Story

RFID TECHNOLOGY is based on the simple idea that an electronic circuit in an unpowered, or "passive," tag—which requires no batteries or maintenance—can be intermittently powered from a distance by a reader device that broadcasts energy to it. So powered, the tag exchanges information with the reader. Tags essentially consist of a plain antenna bonded to a silicon chip and encapsulated inside a glass or plastic module.

Tags operate differently depending on several factors, especially the frequency at which they function. Initially RFID tags worked only at frequency bands of 13.56 megahertz or lower. Such tags, which are still the most widely used, typically need to be less than a meter away from a reader and offer poor discrimination (a reader cannot quickly interpret a multitude of individual tags grouped closely together).

More sophisticated, higher-frequency tags now enable a reader to quickly identify many individual tags grouped together, even haphazardly—although they are not yet able to distinguish perfectly among all the items in a loaded grocery cart. (The ability to swiftly and reliably scan a shopping cart full of jumbled, closely spaced RFID-tagged items is a major aim of this technology. Once perfect-

Overview/*RFID Technology*

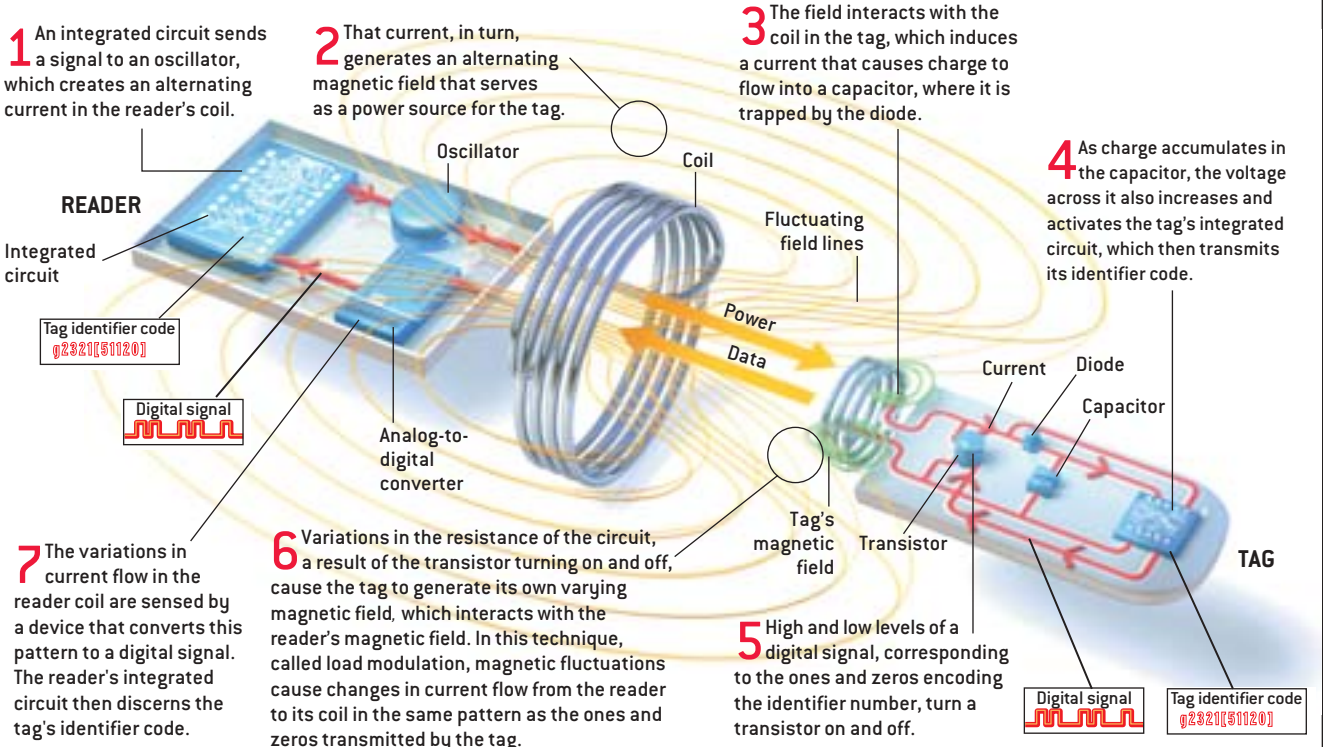
- RFID systems consist of tags—chips that contain identifying and often other data—and reading devices that convey information from the tags to computers.
- Such systems are already in limited use and are being tested widely for applications involving the tracking of inventory from manufacturers to stores.
- As the technology improves and its costs fall, it could form the core of networks that will handle many activities, from monitoring the structural integrity of bridges to reminding you that the tub of coleslaw in the fridge is past its due date.
- Privacy advocates worry that RFID systems will give marketers and others access to more personal information than individuals would want them to have.

JOHNER *Photonica* [*photograph*] [*preceding pages*]

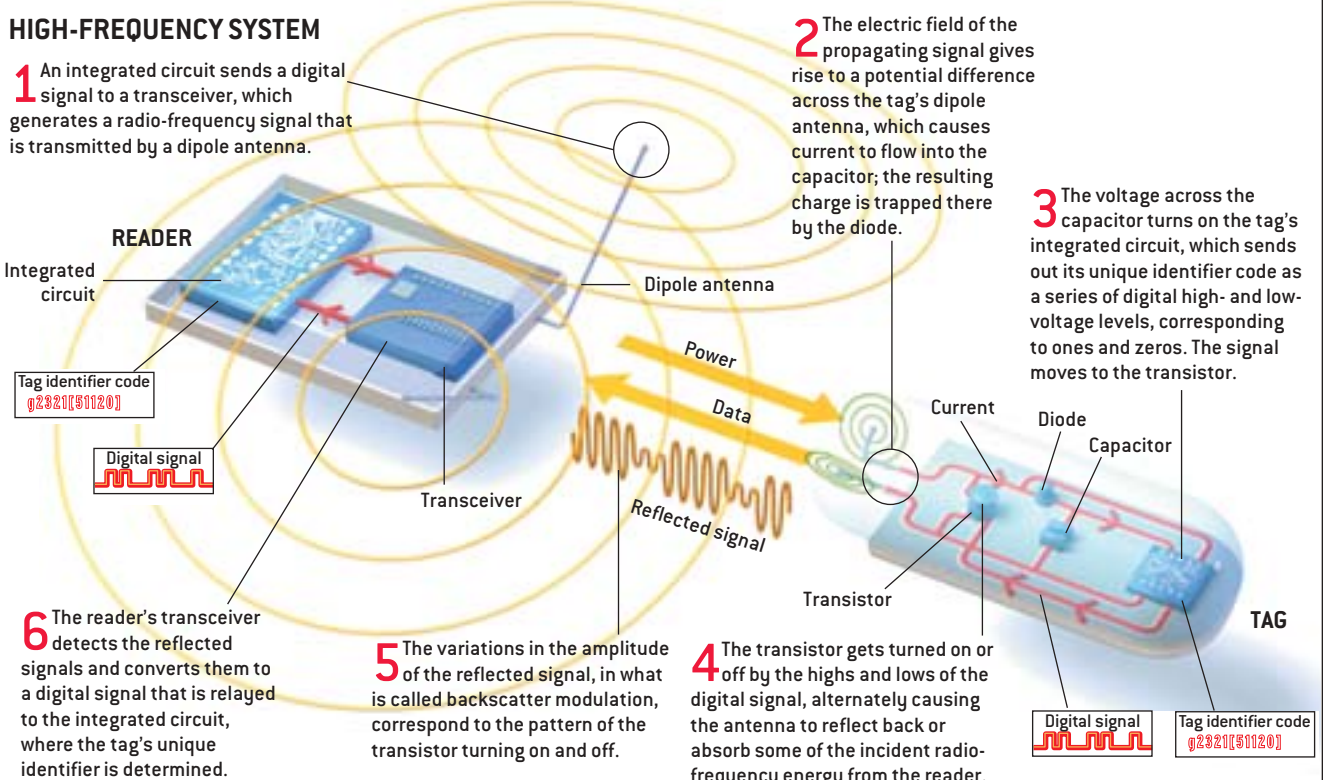
HOW RFID WORKS

RFID systems operate in both low frequency (less than 100 megahertz) and high frequency (greater than 100 megahertz) modes. Unlike their low-frequency counterparts, high-frequency tags can have their data read at distances of greater than one meter, even while closely spaced together. New data can also be transmitted to the tags, a process not shown here.

LOW-FREQUENCY SYSTEM



HIGH-FREQUENCY SYSTEM



BRYAN CHRISTIE DESIGN

ed, such RFID scanning should streamline inventory and checkout procedures and save millions of dollars for retailers.)

The higher-frequency tags can potentially be read from much greater distances than their lower-frequency counterparts, although so far their range has been extended only to a few meters (largely because of tag electronics that operate at very low power derived from the reader's signal, improved antennas and inexpensive high-sensitivity receivers). The updated tags can also hold significantly more information than earlier models, which allows manufacturers to incorporate useful data beyond a mere ID code. The tags can, for instance, use the energy they capture to power an onboard sensor. Tags with sensors that assess tire pressure and temperature while a vehicle is in motion are already in some cars, and Michelin, Philips Semiconductor and BMW are developing prototypes for the mass market.

A major aim of advanced RFID technology is to SWIFTLY AND RELIABLY SCAN a shopping cart full of jumbled, closely spaced tagged items.

RFID Now

RFID DEVICES are beginning to replace magnetic-stripe security cards for unlocking doors and granting access to secured areas—especially at facilities with special security needs, such as military installations. The most visible use of RFID, though, is probably the automatic toll-payment systems that rely on readers at toll plazas to scan tags attached to the windshields of passing cars. The reader records the tag's ID and then deducts money from a prepaid account. These systems are designed to allow cars to zip through toll plazas ideally without stopping or even slowing down very much.

Known as E-ZPass in New York, New Jersey, Delaware and other states, as FasTrak in California, and by different names elsewhere, RFID-based automatic toll systems have been operating for several years. FasTrak, in place on the San Francisco Bay Bridge and on Interstate Highway 15 near San Diego, has been quite successful, but the East Coast E-ZPass system had some early teething problems related to administrative and political issues, not to the technology itself. The San Francisco Bay Bridge system requires drivers to slow to 25 miles per hour while passing the reader, but only for safety reasons, because the tollbooths are narrow. The FasTrak system on I-15, however, operates at freeway speeds and, further, is being used to monitor traffic.

RFID systems are also in the early stages of replacing those familiar Universal Product Code (UPC) bar codes, which are read optically at very short distances to identify products, track inventory and semiautomatize the checkout process at stores. RFID tags, unlike bar codes, can be molded into a product's casing and can use encryption and other strategies to make them difficult to forge. In addition, some RFID tags permit readers to write new data to their onboard memories for later retrieval. For example, each transaction between reader and tag can record the time, date and identity of whoever accessed the tag. This capability should be useful for creating an audit trail in a tag attached to, say, a car, to indicate where it was manufactured and to record each time it was sold, its previous owners, its service history and its accidents.

Based on the growing number of business sectors that are beginning to test tag-and-reader systems, some experts in the field believe RFID will be widely used, especially in retail, by 2010. Others say such broad application will not happen until around 2015 or later, when the cost of RFID tags falls enough to make them economically viable for labeling inexpensive consumer products.

The Near Future

RFID TRACKING technology is starting to be used to follow merchandise as it travels from factory to stores. It will probably be fully established for such applications before it makes deep inroads into stores proper, because warehouse systems are easier to develop and are less likely to fuel public concern that RFID tags in consumer goods could be used to monitor customers once they leave a store. Recently Wal-Mart announced that it will require its top 100 suppliers to place high-frequency tags on cartons and pallets shipped to its stores. And the U.S. Department of Defense has similarly called on its suppliers to adopt high-frequency RFID inventory labeling by 2005.

But the potential—and inevitable—uses for RFID in stores themselves remain tantalizing for retailers. The canceled Wal-Mart in-store test, planned in partnership with Gillette, would have evaluated the ability of RFID-based "smart shelves," equipped with built-in readers, to monitor the movement of millions of shavers and other Gillette products embedded with RFID tags. (In principle, the 96-bit code allotted for identification of each RFID tag would allow every person on earth to have about 50 quadrillion tags apiece.) The ability to keep tabs on individual products on store shelves is generally accepted as the most difficult task for RFID technology—but one that could pay off royally for retailers.

Notably, RFID smart-shelf systems could save money on labor and help to increase sales by ensuring that shelves are always stocked. If the systems monitored stock levels, employees would not have to do it: when the computers sensed that stock was running low, they could automatically alert someone to order more or

THE AUTHOR

ROY WANT is a principal engineer at Intel Research/CTG in Santa Clara, Calif., where he leads a project that is setting a long-range research agenda for ubiquitous computing. Early in his career, at Olivetti Research, Want pursued an automated system for locating people inside buildings. At Xerox PARC's Ubiquitous Computing Program, he headed the development of one of the first context-aware computer systems, managed the Embedded Systems group, and worked on applications of electronic tagging and on the design of PDAs with manipulative user interfaces. Want holds more than 50 patents related to mobile and distributed computer systems and is associate editor in chief of *IEEE Pervasive Computing*.

Dealing with the Darker Side

WHAT WILL BE the social consequences of a world full of embedded RFID tags and readers? Will our privacy be further eroded as RFID technology makes it possible for our movements to be tracked and allows our personal information to be available in unprecedented detail? These and many other questions must be answered before RFID systems become commonplace.

One of the major worries for privacy advocates is that RFID tags identifying individual items purchased with credit or debit cards would link buyers to the specific items in the card's or the store's databases. Marketers could then use these data to keep track of exactly what particular people bought, down to the color, size, style and price—more information than UPC bar codes reveal. In an amplification of the way that phone and direct-mail solicitors use similar, less accurate data to target people for sales pitches, those equipped with RFID-derived data might home in on consumers with very specific sales pitches.

Another concern is that RFID equipment will produce automatic audit trails of commercial transactions: in a totally tagged world, it will be easier to detect when we lie about how we spent our time or what we did and where. This capability could have great consequences for the workplace, and the legal system might look to using logs kept by tag readers as courtroom evidence. We may need laws to specify who can access data logs and for what purpose. In Europe, the Data Protection Act already limits access to computer records of this kind, and the U.S. will probably enact similar legislation.

We will also have to grapple with the inevitable displacement of workers by RFID systems. Opposition to tagging could well come from the industrial labor force, which stands to lose significant numbers of jobs as industry adopts RFID tools able to perform tasks that now depend on human effort. A bitter strike by longshoremen on the West Coast in 2002, partly over new technology that threatened future jobs, may have been a preview of conflicts to come over RFID systems.

PROTEST ENSUED when the British supermarket chain Tesco began testing RFID technology.

Privacy Advocates Protest

The backlash against perceived invasions of consumer privacy by RFID applications began in March 2003, when Philips Semiconductor announced that it was shipping 15 million RFID tags to the clothing manufacturer and retailer Benetton to be incorporated into labels during production. The tags were to interact with a network of RFID readers in Benetton's store shelves and warehouses to track inventory throughout the company's 5,000 retail outlets worldwide.

Despite Philips's reassurances that tagged clothing could not be tracked outside Benetton stores, some industry experts said that criminals could increase the Benetton tags' tracking distance by creating more sensitive RFID readers. Privacy advocates worried that the tags could be scanned by RFID readers other than those in Benetton stores, which would allow people wearing the clothes to be monitored without their knowledge by, say, criminals or the government. Consumers Against Supermarket Privacy Invasion and Numbering, a U.S.-based privacy group, called for a worldwide boycott of Benetton until the company abandoned RFID tracking technology. Benetton quickly issued statements saying that although it had already tested RFID systems, it was not using RFID inventory tracking and had no firm plans to insert the millions of Philips tags into its products.

Similar concerns—that corporations might keep consumers' products under surveillance in purchasers' homes and on the streets—surfaced about a test of an in-store RFID inventory system that was planned by Wal-Mart and Gillette. To answer consumer concern, Gillette announced that it was embedding its RFID tags in packaging, not products, so purchasers would discard the tags with the packaging. But Declan McCullagh, a commentator who writes for computing publications and who favors RFID for its practical value, has written: "Future burglars could canvass alleys with RFID detectors, looking for RFID tags on discarded packaging that indicates expensive electronic gear is nearby.... [T]he ability to remain anonymous is eroded."

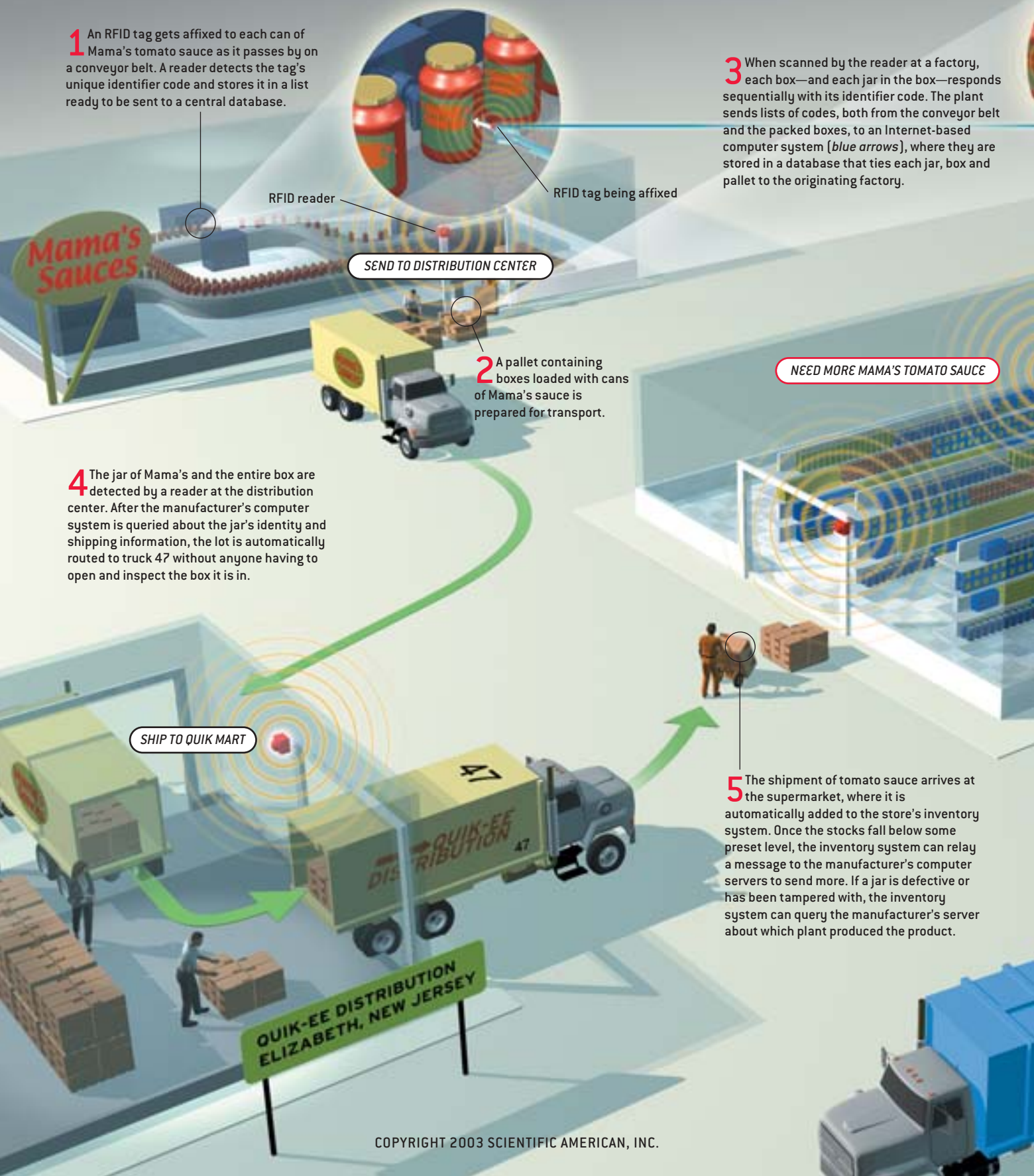
One way to avoid such possibilities is to put a kill switch into each RFID tag on a consumer item, which would allow the tag to be turned off after purchase. Indeed, the Auto-ID Center—a research consortium funded by information technology companies and headquartered at the Massachusetts Institute of Technology—has released guidelines saying that retailers must be able to disable RFID tags at checkout counters, and manufacturers, including Alien Technology, Matrics and Philips, are now producing tags with kill switches.

McCullagh has suggested four requirements for the use of RFID tags on consumer products: Consumers should be notified when RFID tags are present in what they are buying (this could be done with a printed notice on a checkout receipt). All tags should be readily visible and easily removable. The tags should be disabled by default at the checkout counter. And, when possible, RFID tags should be placed only on the product's packaging, not embedded in the product. *—R.W.*

CHRIS McDERMOTT/NOTAGS.CO.UK

TRACKING THINGS HERE, THERE AND EVERYWHERE

RFID SYSTEMS will let products, such as the hypothetical Mama's tomato sauce, be monitored and reordered automatically, once depleted. Even an individual item will be traced. Here is one scenario.



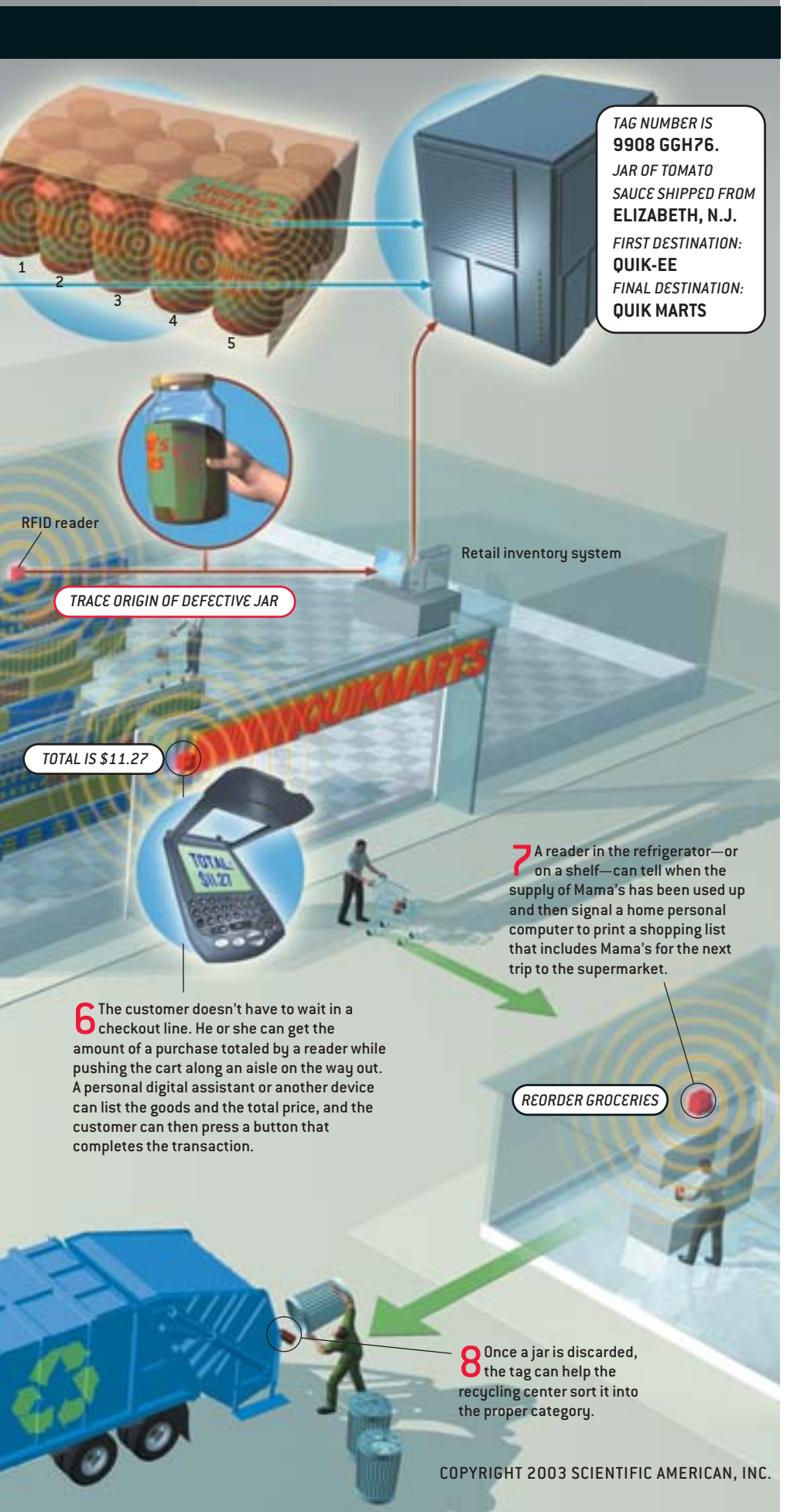
1 An RFID tag gets affixed to each can of Mama's tomato sauce as it passes by on a conveyor belt. A reader detects the tag's unique identifier code and stores it in a list ready to be sent to a central database.

2 A pallet containing boxes loaded with cans of Mama's sauce is prepared for transport.

3 When scanned by the reader at a factory, each box—and each jar in the box—responds sequentially with its identifier code. The plant sends lists of codes, both from the conveyor belt and the packed boxes, to an Internet-based computer system (*blue arrows*), where they are stored in a database that ties each jar, box and pallet to the originating factory.

4 The jar of Mama's and the entire box are detected by a reader at the distribution center. After the manufacturer's computer system is queried about the jar's identity and shipping information, the lot is automatically routed to truck 47 without anyone having to open and inspect the box it is in.

5 The shipment of tomato sauce arrives at the supermarket, where it is automatically added to the store's inventory system. Once the stocks fall below some preset level, the inventory system can relay a message to the manufacturer's computer servers to send more. If a jar is defective or has been tampered with, the inventory system can query the manufacturer's server about which plant produced the product.



could place orders directly with the manufacturer. The systems could offer other benefits as well. Because inventory tags are programmable, their data can include information about where the item was manufactured and sold. And like pinned-on magnetic antishoplifting tags, the RFID inventory tags could be detected leaving the store to prevent theft (estimated to cost $50 billion a year).

Wal-Mart said it canceled its in-store test to free up resources for developing behind-the-scenes RFID capabilities in its warehouses, which will require fewer tags and less powerful computing. This is probably true; industry insiders, however, have suggested that consumer concerns over RFID systems invading individual privacy also played a significant role in the decision. That the backlash had an influence would not be surprising, given that it was at about the same time that Benetton aborted its own large-scale in-store test of an inventory system after its plans were criticized by consumers and the media. The Benetton trial would have examined RFID technology's ability to scan entire cases of tag-bearing clothes in many different colors, sizes and styles and to capture and upload the inventory data to its tracking system, obviating the need for workers to hand-check each garment.

Other tests of warehouse and in-store inventory systems continue, by Procter & Gamble, Canon, and International Paper. And last spring, Metro, a German retail chain, opened a "future store" equipped with an RFID inventory management system involving both smart shelves and scales equipped with RFID readers that can identify types of produce. In addition, tagged shopping carts are scanned to measure in-store customer traffic and to signal automatically for the opening or closing of checkout stations. The Metro pilot is the work of Intel, where I work today, and the German software developer SAP, along with more than 30 other companies, including Hewlett-Packard, Cisco Systems and Philips.

Over the Horizon

RFID INVENTORY systems still fall far short of Weiser's vision: they do not help us perform everyday tasks. Indeed, com-

puters and chips scattered throughout our homes—in toasters, games, entertainment systems and other devicess—demand more, not less, of our attention. We must configure and control dozens of devices, transfer data between them, and try to figure out what went wrong when a failure occurs. Simple tasks, such as setting a wristwatch or operating a television, require an instruction manual. It is clear that for computing to become invisible, we need not only ubiquitous computing but what David L. Tennenhouse of Intel calls "proactive computing"—systems that anticipate what we need and provide it without forcing us to do a lot of work first.

For proactive computing to function on a major scale, networks of RFID readers must be placed throughout the environment. Forward thinkers envision two main types of proactive RFID networks, both of which include a web of interacting readers that monitor many RFID tags and convey the information they collect to remote computers.

One type is made up of readers set permanently in place and connected together by cables. These devices would power and read tags—some with sensors—that are also permanently fixed in place. (If necessary, the tags could also be read by mobile readers passing by.)

This kind of network might be installed on a bridge: tags would be buried deep inside concrete structural members, welded into joints between steel beams and put in other places where their sensors could measure stress and change in various parts of the structure. They would collect and store such information as the discovery that a structural member had been flexed beyond its safe limits during a seismic tremor. The readers would be powered from ordinary AC electric lines or through the interreader network cables and would be hardwired to an Internet connection, so they could send their data to computers that would analyze the input and take action in response.

The second type of system—called an ad hoc wireless network—does not have all its readers and sensor tags permanently in place. Instead it is made up of RFID readers put wherever they are needed, in the same way you would choose a spot to plug in a lamp. They read tags that surround them: some of the tags are fixed and stationary; some have sensors and some do not; and some are mobile, attached to devices and people that pass through the network. Readers may be AC-powered if they are near power outlets or may be battery-powered. These readers, also known as network nodes, can form short-range wireless connections to one another on the fly: information moves across the network by hopping wirelessly from node to node (which is why these are sometimes called multihop networks) and flows toward a gateway node with an Internet connection.

You might create an ad hoc network with many readers monitoring hundreds of tag sensors spread out across tens of

Technical Challenges Ahead

- RFID tags and readers are orientation-dependent. Tags must be positioned properly relative to readers so that the antenna coils can exchange signals. The solution to this problem will come with the development of multiple-reader systems that use an array of readers positioned to cover all the possible orientations for tagged items that might, for example, be found in a display bin in a store. Part of this solution will involve protocols to coordinate the operation of these reader arrays.

- RFID signals are easily blocked. Over short ranges, these signals can be attenuated by certain materials (the most common is packing made from metallic substances). Over longer ranges, the signals—which are much weaker than commercial radio broadcast signals—can be blocked by common objects, including the human body. Researchers are working to solve this problem by using novel designs for tag antennas and more sensitive reader arrays.

- At an average cost of around 20 to 30 cents apiece, RFID tags are still too costly, especially for retail applications and certainly for use on inexpensive and low-margin products, such as a 50-cent candy bar or a $1 bar of soap. This is a key reason why mass-market consumer retail businesses —which operate on very thin profit margins—have been slow to adopt RFID-based smart shelf and smart checkout technology. RFID tag developers are working to lower the cost of tags to 10 cents, or even five cents, over the next few years. Some experts in the tag-manufacturing field believe that at these costs, RFID may not be widely adopted until at least 2010—if ever. Others say the cost must shrink to a fraction of a cent before we will see a tag on every item in the grocery store, which may take until 2015 or later.

DIVERSITY of RFID tags reflects the lack of standards for the technology.

- Competing technical standards for RFID readers and tags prevent their universal adoption. Different manufacturers are developing tag protocols operating at different frequencies with a variety of packet formats—the way in which a tag's digital data are packaged and transmitted to a reader. Ideally, a single standard should be adopted to make all tags compatible with all readers.

Both the cost and standardization challenges are being addressed by individual companies and by the Auto-ID Center and the International Organization for Standardization (ISO), industry consortia working to set standards for RFID tags.

COURTESY OF ROY WANT

square miles. Such a network could provide the data to make improved weather forecasts. If the sensors could simultaneously detect wind speeds at many locations across the whole area, the computer could even sense the formation of a tornado at an early stage and generate an earlier alert than is possible today.

An ad hoc RFID network in an office building could perform many tasks. Readers could monitor sensor tags that indicate the temperature in different rooms so that the central computer could maintain constant conditions throughout the building or on a single floor. Other readers would scan employees' security badges and recognize the tags in their laptops so that workers could access centrally stored data or link up with colleagues elsewhere in the building. The design of all kinds of sensor networks is being researched by Deborah Estrin's team at the Center for Embedded Networked Sensing at the University of California at Los Angeles, by David E. Culler's team at the University of California at Berkeley, by Gaetano Borriello's team at the University of Washington, at Intel Research's Network of Labs, and at several small companies, including Crossbow in Santa Clara, Calif., Dust Inc. in Berkeley, Calif., and Sensoria in San Diego.

The Responsive Environment

WHEN RFID NETWORKS are finally in place everywhere and we are surrounded by tags and readers feeding responsive computer systems, we will have reached the point at which Weiser believed computing could be blended invisibly into everyday tasks. At this level of integration, RFID technology will support even our simplest activities. For example, RFID-enhanced computer products could "talk" to one another and independently configure their connections. My Intel colleague Trevor Pering has been exploring a way to automatically configure wireless network links between mobile computers and peripherals. If you purchased a printer with a wireless networking capability (such as Bluetooth) and an RFID tag, you might simply unpack the device and bring it near your computer: the computer would read the printer's RFID tag and connect to the printer automatically, eliminating messy configuration dialogues.

The scope of possible RFID applications is vast and could even include assisting people with Alzheimer's disease. Eric Dishman, also at Intel, is working on a system aimed at helping those with memory impairment maintain their independence. In one prototype system, all the objects needed for making a cup of tea are tagged. If the patient picks up at least two objects—say, a sugar jar and a tea bag—the system infers, by knowing the ID and location of the objects in relation to one another, that the patient needs help. The system also tracks the sequence in which the objects are used in order to infer whether the person is "stuck" and then delivers recorded voice assistance.

In a totally different realm, PSA Corporation, Hutchinson-Whampoa and P&O Ports—the three largest seaport operators in the world—have taken what could be the early steps toward developing an RFID-based antiterrorism security system that would outfit cargo containers with hidden sensor tags designed to detect radiation or chemical or biological agents in smuggled weapons. Right now the system can detect only whether a container has been opened by an unauthorized person during transit. It could be expanded so that at each stage of a container's journey, from its initial site to ground transportation, dockside storage and transport ships, readers would interrogate the tag to determine if it had detected dangerous materials. The tag's sensor would permanently register even very brief exposures to these substances and flag the incident at the next reading station.

Eventually PDAs (personal digital assistants) could be designed to operate as RFID tag readers so that we could receive proactive assistance from tags placed virtually everywhere in our environment. From a tagged sign on a train station, your PDA could retrieve a Web address linking you to an Internet-based timetable. Similarly, realtors could tag the signs on homes for sale: driving past, you could simply beam your PDA at the realtor's sign and then download photographs and information about the property from the Internet.

Benetton aborted a large-scale in-store test of an RFID inventory system after its plans were CRITICIZED BY CONSUMERS and the media.

Important technical challenges remain, and so it will be years, perhaps decades, before we can reap the benefits of such fully realized RFID applications. As RFID reader-and-tag networks begin appearing in our environment, however, we will increasingly see how this technology can extend the ability of computers—in combination with the Internet—to sense and respond to the physical world.

In his 1991 article in this magazine, Weiser wrote: "There is more information available at our fingertips during a walk in the woods than in any computer system, yet people find a walk among trees relaxing and computers frustrating. Machines that fit the human environment, instead of forcing humans to enter theirs, will make using a computer as refreshing as taking a walk in the woods." Wielded sensibly, RFID has the power to make computing an unobtrusive, intuitive part of everyday life—indeed, as refreshing as a walk through nature. SA

MORE TO EXPLORE

The Computer for the 21st Century. Mark Weiser in *Scientific American,* Vol. 265, No. 3, pages 94–104; September 1991.

Ubiquitous Electronic Tagging. Roy Want and Dan M. Russell in *IEEE Distributed Systems Online,* Vol. 1, No. 2; September 2000.

Connecting the Physical World with Pervasive Networks. Deborah Estrin, David Culler, Kris Pister and Gaurav Sukhatme in *IEEE Pervasive Computing,* Vol. 1, No. 1, pages 59–69; January–March 2002.

Comparing Autonomic and Proactive Computing. Roy Want, Trevor Pering and David Tennenhouse in *IBM Systems Journal,* Vol. 42, No. 1, pages 129–135; January 2003.

The RFID Handbook. Klaus Finkenzeller. John Wiley & Sons, 2003.

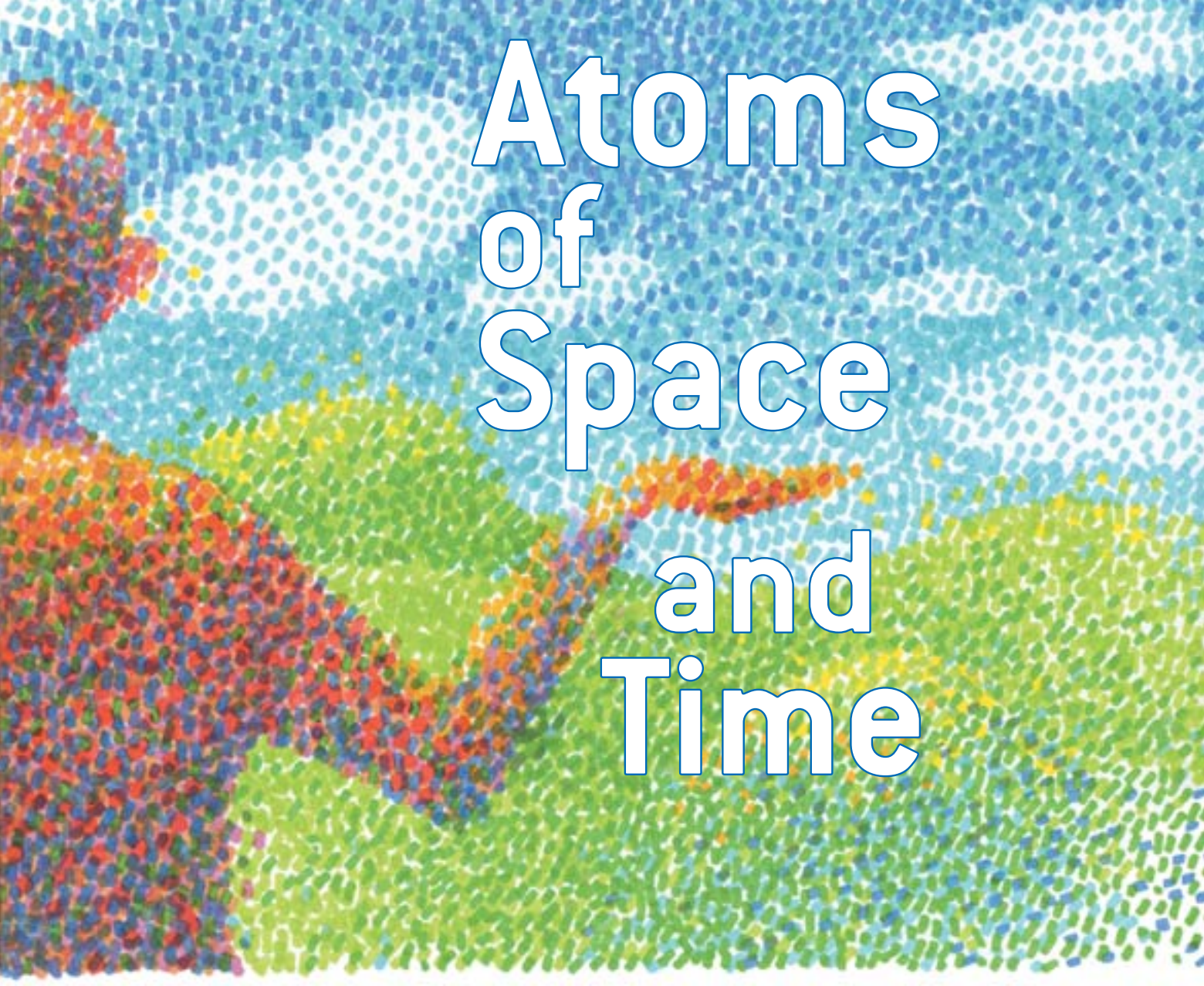
We perceive space and time to be continuous, but if the amazing theory of loop quantum gravity is correct, they actually come in discrete pieces

By Lee Smolin

A little more than 100 years ago most people—and most scientists—thought of matter as continuous. Although since ancient times some philosophers and scientists had speculated that if matter were broken up into small enough bits, it might turn out to be made up of very tiny atoms, few thought the existence of atoms could ever be proved. Today we have imaged individual atoms and have studied the particles that compose them. The granularity of matter is old news.

In recent decades, physicists and mathematicians have asked if space is also made of discrete pieces. Is it continuous, as we learn in school, or is it more like a piece of cloth, woven out of individual fibers? If we could probe to size scales that were small enough, would we see "atoms" of space, irreducible pieces of volume that cannot be broken into anything smaller? And what about time: Does nature change continuously, or does the world

DUSAN PETRICIC



Atoms of Space and Time

evolve in series of very tiny steps, acting more like a digital computer?

The past 16 years have seen great progress on these questions. A theory with the strange name of "loop quantum gravity" predicts that space and time are indeed made of discrete pieces. The picture revealed by calculations carried out within the framework of this theory is both simple and beautiful. The theory has deepened our understanding of puzzling phenomena having to do with black holes and the big bang. Best of all, it is testable; it makes predictions for experiments that can be done in the near future that will enable us to detect the atoms of space, if they are really there.

Quanta

MY COLLEAGUES AND I developed the theory of loop quantum gravity while struggling with a long-standing problem in physics: Is it possible to develop a quantum theory of gravity? To explain why this is an important question—and what it has to do with the granularity of space and time—I must first say a bit about quantum theory and the theory of gravity.

The theory of quantum mechanics was formulated in the first quarter of the 20th century, a development that was closely connected with the confirmation that matter is made of atoms. The equations of quantum mechanics require that certain quantities, such as the energy of an atom, can come only in specific, discrete units. Quantum theory successfully predicts the properties and behavior of atoms and the elementary particles and forces that compose them. No theory in the history of science has been more successful than quantum theory. It underlies our understanding of chemistry, atomic and subatomic physics, electronics and even biology.

In the same decades that quantum mechanics was being formulated, Albert Einstein constructed his general theory of relativity, which is a theory of gravity. In his theory, the gravitational

force arises as a consequence of space and time (which together form "spacetime") being curved by the presence of matter. A loose analogy is that of a bowling ball placed on a rubber sheet along with a marble that is rolling around nearby. The balls could represent the sun and the earth, and the sheet is space. The bowling ball creates a deep indentation in the rubber sheet, and the slope of this indentation causes the marble to be deflected toward the larger ball, as if some force—gravity—were pulling it in that direction. Similarly, any piece of matter or concentration of energy distorts the geometry of spacetime, causing other particles and light rays to be deflected toward it, a phenomenon we call gravity.

Quantum theory and Einstein's theory of general relativity separately have each been fantastically well confirmed by experiment—but no experiment has explored the regime where both theories predict significant effects. The problem is that quantum effects are most prominent at small size scales, whereas general relativistic effects require large masses, so it takes extraordinary circumstances to combine both conditions.

Allied with this hole in the experimental data is a huge conceptual problem: Einstein's theory of general relativity is thoroughly classical, or nonquantum. For physics as a whole to be logically consistent, there has to be a theory that somehow unites quantum mechanics and general relativity. This long-sought-after theory is called quantum gravity. Because general relativity deals in the geometry of spacetime, a quantum theory of gravity will in addition be a quantum theory of spacetime.

SPACE IS WOVEN out of distinct threads.

Physicists have developed a considerable collection of mathematical procedures for turning a classical theory into a quantum one. Many theoretical physicists and mathematicians have worked on applying those standard techniques to general relativity. Early results were discouraging. Calculations carried out in the 1960s and 1970s seemed to show that quantum theory and general relativity could not be successfully combined. Consequently, something fundamentally new seemed to be required, such as additional postulates or principles not included in quantum theory and general relativity, or new particles or fields, or new entities of some kind. Perhaps with the right additions or a new mathematical structure, a quantumlike theory could be developed that would successfully approximate general relativity in the nonquantum regime. To avoid spoiling the successful predictions of quantum theory and general relativity, the exotica contained in the full theory would remain hidden from experiment except in the extraordinary circumstances where both quantum theory and general relativity are expected to have large effects. Many different approaches along these lines have been tried, with names such as twistor theory, noncommutative geometry and supergravity.

An approach that is very popular with physicists is string theory, which postulates that space has six or seven dimensions—all so far completely unobserved—in addition to the three that we are familiar with. String theory also predicts the existence of a great many new elementary particles and forces, for which there is so far no observable evidence. Some researchers believe that string theory is subsumed in a theory called M-theory [see "The Theory Formerly Known as Strings," by Michael J. Duff; Scientific American, February 1998], but unfortunately no precise definition of this conjectured theory has ever been given. Thus, many physicists and mathematicians are convinced that alternatives must be studied. Our loop quantum gravity theory is the best-developed alternative.

Overview/*Quantum Spacetime*

- To understand the structure of space on the very smallest size scale, we must turn to a quantum theory of gravity. Gravity is involved because Einstein's general theory of relativity reveals that gravity is caused by the warping of space and time.
- By carefully combining the fundamental principles of quantum mechanics and general relativity, physicists are led to the theory of "loop quantum gravity." In this theory, the allowed quantum states of space turn out to be related to diagrams of lines and nodes called spin networks. Quantum spacetime corresponds to similar diagrams called spin foams.
- Loop quantum gravity predicts that space comes in discrete lumps, the smallest of which is about a cubic Planck length, or 10^{-99} cubic centimeter. Time proceeds in discrete ticks of about a Planck time, or 10^{-43} second. The effects of this discrete structure might be seen in experiments in the near future.

A Big Loophole

In the mid-1980s a few of us—including Abhay Ashtekar, now at Pennsylvania State University, Ted Jacobson of the University of Maryland and Carlo Rovelli, now at the University of the Mediterranean in Marseille—decided to reexamine the question of whether quantum mechanics could be combined consistently with general relativity using the standard techniques. We knew that the negative results from the 1970s had an important loophole. Those calculations assumed that the geometry of space is continuous and smooth, no matter how minutely we examine it, just as people

DUSAN PETRICIC

had expected matter to be before the discovery of atoms. Some of our teachers and mentors had pointed out that if this assumption was wrong, the old calculations would not be reliable.

So we began searching for a way to do calculations without assuming that space is smooth and continuous. We insisted on not making any assumptions beyond the experimentally well tested principles of general relativity and quantum theory. In particular, we kept two key principles of general relativity at the heart of our calculations.

The first is known as background independence. This principle says that the geometry of spacetime is not fixed. Instead the geometry is an evolving, dynamical quantity. To find the geometry, one has to solve certain equations that include all the effects of matter and energy. Incidentally, string theory, as currently formulated, is not background independent; the equations describing the strings are set up in a predetermined classical (that is, nonquantum) spacetime.

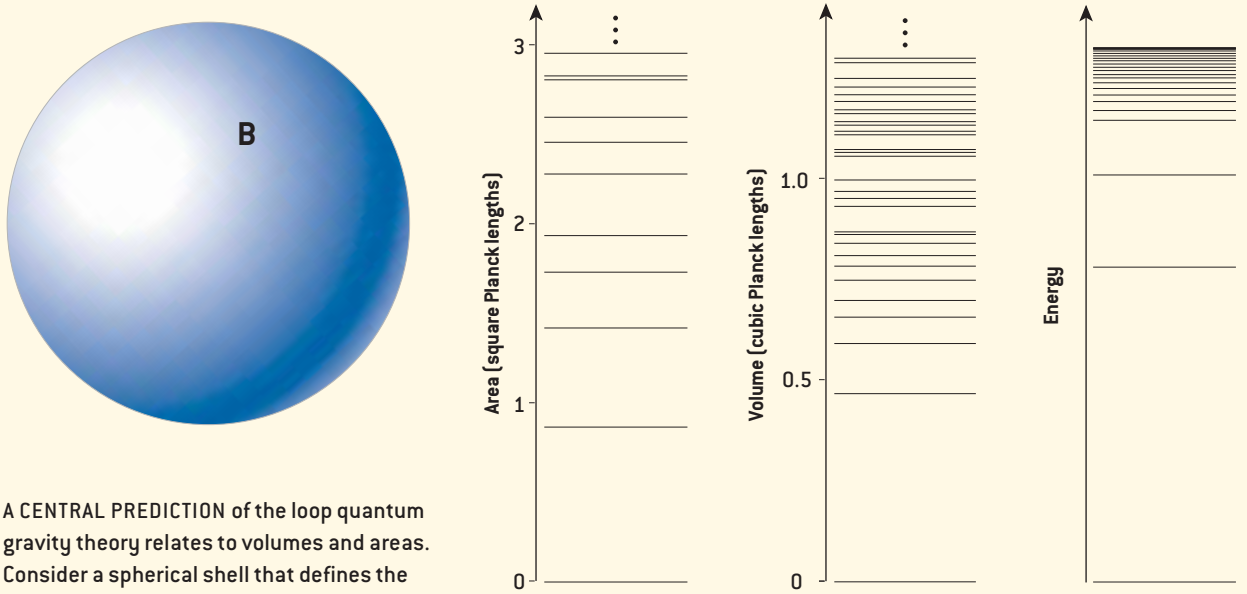
The second principle, known by the imposing name diffeomorphism invariance, is closely related to background independence. This principle implies that, unlike theories prior to general relativity, one is free to choose any set of coordinates to map spacetime and express the equations. A point in spacetime is defined only by what physically happens at it, not by its location according to some special set of coordinates (no coordinates are special). Diffeomorphism invariance is very powerful and is of fundamental importance in general relativity.

By carefully combining these two principles with the standard techniques of quantum mechanics, we developed a mathematical language that allowed us to do a computation to determine whether space is continuous or discrete. That calculation revealed, to our delight, that space is quantized. We had laid the foundations of our theory of loop quantum gravity. The term "loop," by the way, arises from how some computations in the theory involve small loops marked out in spacetime.

The calculations have been redone by a number of physicists and mathematicians using a range of methods. Over the years since, the study of loop quantum gravity has grown into a healthy field of research, with many contributors around the world; our combined efforts give us confidence in the picture of spacetime I will describe.

Ours is a quantum theory of the structure of spacetime at the smallest size scales, so to explain how the theory works we need to consider what it predicts for a small region or volume. In dealing with quantum physics, it is essential to specify precisely what physical quantities are to be measured. To do so, we consider a region somewhere that is marked out by a boundary, B [*see illustration below*].

QUANTUM STATES OF VOLUME AND AREA



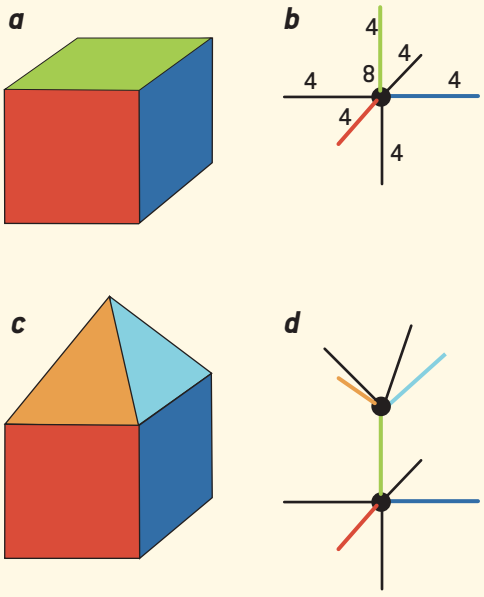
A CENTRAL PREDICTION of the loop quantum gravity theory relates to volumes and areas. Consider a spherical shell that defines the boundary, B, of a region of space having some volume (*above*). According to classical (nonquantum) physics, the volume could be any positive real number. The loop quantum gravity theory says, however, that there is a nonzero absolute minimum volume (about one cubic Planck length, or 10^{-99} cubic centimeter), and it restricts the set of larger volumes to a discrete series of numbers. Similarly, there is a nonzero minimum area (about one square Planck length, or 10^{-66} square centimeter) and a discrete series of larger allowed areas. The discrete spectrum of allowed quantum areas (*left*) and volumes (*center*) is broadly similar to the discrete quantum energy levels of a hydrogen atom (*right*).

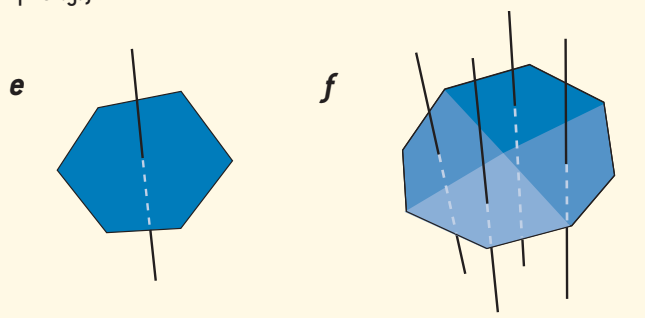
SOURCE: AREA AND VOLUME DATA FROM ROBERTO DE PIETRI AND CARLO ROVELLI (http://arxiv.org/abs/gr-qc/9602023); NADIA STRASSER

VISUALIZING QUANTUM STATES OF VOLUME

DIAGRAMS CALLED SPIN NETWORKS are used by physicists who study loop quantum gravity to represent quantum states of space at a minuscule scale. Some such diagrams correspond to polyhedra-shaped volumes. For example, a cube (*a*) consists of a volume enclosed within six square faces. The corresponding spin network (*b*) has a dot, or node, representing the volume and six lines that represent the six faces. The complete spin network has a number at the node to indicate the cube's volume and a number on each line to indicate the area of the corresponding face. Here the volume is eight cubic Planck lengths, and the faces are each four square Planck lengths. (The rules of loop quantum gravity restrict the allowed volumes and areas to specific quantities: only certain combinations of numbers are allowed on the lines and nodes.)

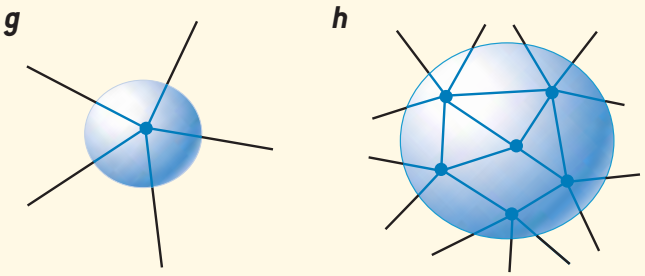
If a pyramid sat on the cube's top face (*c*), the line representing that face in the spin network would connect the cube's node to the pyramid's node (*d*). The lines corresponding to the four exposed faces of the pyramid and the five exposed faces of the cube would stick out from their respective nodes. (The numbers have been omitted for simplicity.)





One quantum of area

Larger area



One quantum of volume

Larger volume

In general, in a spin network, one quantum of area is represented by a single line (*e*), whereas an area composed of many quanta is represented by many lines (*f*). Similarly, a quantum of volume is represented by one node (*g*), whereas a larger volume takes many nodes (*h*). If we have a region of space defined by a spherical shell, the volume inside the shell is given by the sum of all the enclosed nodes and its surface area is given by the sum of all the lines that pierce it.

The spin networks are more fundamental than the polyhedra: any arrangement of polyhedra can be represented by a spin network in this fashion, but some valid spin networks represent combinations of volumes and areas that cannot be drawn as polyhedra. Such spin networks would occur when space is curved by a strong gravitational field or in the course of quantum fluctuations of the geometry of space at the Planck scale.

The boundary may be defined by some matter, such as a cast-iron shell, or it may be defined by the geometry of spacetime itself, as in the event horizon of a black hole (a surface from within which even light cannot escape the black hole's gravitational clutches).

What happens if we measure the volume of the region? What are the possible outcomes allowed by both quantum theory and diffeomorphism invariance? If the geometry of space is continuous, the region could be of any size and the measurement result could be any positive real number; in particular, it could be as close as one wants to zero volume. But if the geometry is granular, then the measurement result can come from just a discrete set of numbers and it cannot be smaller than a certain minimum possible volume. The question is similar to asking how much energy electrons orbiting an atomic nucleus have. Classical mechanics predicts that that an electron can possess any amount of energy, but quantum mechanics allows only specific energies (amounts in between those values do not occur). The difference is like that between the measure of something that flows continuously, like the 19th-century conception of water, and something that can be counted, like the atoms in that water.

The theory of loop quantum gravity

NADIA STRASSER

predicts that space is like atoms: there is a discrete set of numbers that the volume-measuring experiment can return. Volume comes in distinct pieces. Another quantity we can measure is the area of the boundary B. Again, calculations using the theory return an unambiguous result: the area of the surface is discrete as well. In other words, space is not continuous. It comes only in specific quantum units of area and volume.

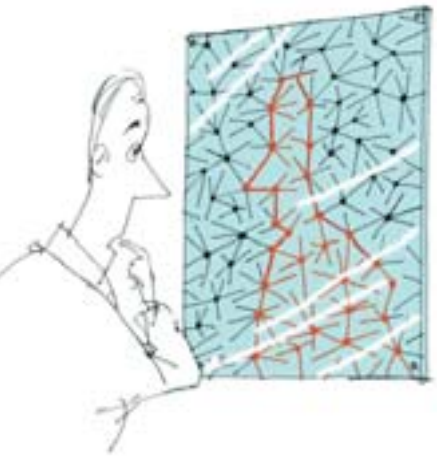
The possible values of volume and area are measured in units of a quantity called the Planck length. This length is related to the strength of gravity, the size of quanta and the speed of light. It measures the scale at which the geometry of space is no longer continuous. The Planck length is very small: 10^{-33} centimeter. The smallest possible nonzero area is about a square Planck length, or 10^{-66} cm^2. The smallest nonzero volume is approximately a cubic Planck length, 10^{-99} cm^3. Thus, the theory predicts that there are about 10^{99} atoms of volume in every cubic centimeter of space. The quantum of volume is so tiny that there are more such quanta in a cubic centimeter than there are cubic centimeters in the visible universe (10^{85}).

Spin Networks

WHAT ELSE DOES our theory tell us about spacetime? To start with, what do these quantum states of volume and area look like? Is space made up of a lot of little cubes or spheres? The answer is no—it's not that simple. Nevertheless, we can draw diagrams that represent the quantum states of volume and area. To those of us working in this field, these diagrams are beautiful because of their connection to an elegant branch of mathematics.

To see how these diagrams work, imagine that we have a lump of space shaped like a cube, as shown in the illustration on the opposite page. In our diagrams, we would depict this cube as a dot, which represents the volume, with six lines sticking out, each of which represents one of the cube's faces. We have to write a number next to the dot to specify the quantity of volume, and on each line we write a number to specify the area of the face that the line represents.

Next, suppose we put a pyramid on top of the cube. These two polyhedra, which share a common face, would be depicted as two dots (two volumes) connected by one of the lines (the face that joins the two volumes). The cube has five other faces (five lines sticking out), and the pyramid has four (four lines sticking out). It is clear how more complicated arrangements involving polyhedra other than cubes and pyramids could be depicted with these dot-and-line diagrams: each polyhedron of volume becomes a dot, or node, and each flat face of a polyhedron becomes a line, and the lines join the nodes in the way that the faces join the polyhedra together. Mathematicians call these line diagrams graphs.



MATTER EXISTS at the nodes of the spin network.

Now in our theory, we throw away the drawings of polyhedra and just keep the graphs. The mathematics that describes the quantum states of volume and area gives us a set of rules for how the nodes and lines can be connected and what numbers can go where in a diagram. Every quantum state corresponds to one of these graphs, and every graph that obeys the rules corresponds to a quantum state. The graphs are a convenient shorthand for all the possible quantum states of space. (The mathematics and other details of the quantum states are too complicated to discuss here; the best we can do is show some of the related diagrams.)

The graphs are a better representation of the quantum states than the polyhedra are. In particular, some graphs connect in strange ways that cannot be converted into a tidy picture of polyhedra. For example, whenever space is curved, the polyhedra will not fit together properly in any drawing we could do, yet we can still easily draw a graph. Indeed, we can take a graph and from it calculate how much space is distorted. Because the distortion of space is what produces gravity, this is how the diagrams form a quantum theory of gravity.

For simplicity, we often draw the graphs in two dimensions, but it is better to imagine them filling three-dimensional space, because that is what they represent. Yet there is a conceptual trap here: the lines and nodes of a graph do not live at specific locations in space. Each graph is defined only by the way its pieces connect together and how they relate to well-defined boundaries such as boundary B. The continuous, three-dimensional space that you are imagining the graphs occupy *does not exist* as a separate entity. All that exist are the lines and nodes; they *are* space, and the way they connect defines the geometry of space.

These graphs are called spin networks because the numbers on them are related to quantities called spins. Roger Penrose of the University of Oxford first proposed in the early 1970s that spin networks might play a role in theories of quantum gravity. We were very pleased when we found, in 1994, that precise calculations confirmed his intuition. Readers familiar with Feynman diagrams should note that our spin networks are *not* Feynman diagrams, despite the superficial resemblance. Feynman diagrams represent quantum interactions between particles, which proceed from one quantum state to another. Our diagrams represent fixed quan-

THE AUTHOR

LEE SMOLIN is a researcher at the Perimeter Institute for Theoretical Physics in Waterloo, Ontario, and an adjunct professor of physics at the University of Waterloo. He has a B.A. from Hampshire College and a Ph.D. from Harvard University and has been on the faculty of Yale, Syracuse and Pennsylvania State universities. In addition to his work on quantum gravity, he is interested in elementary particle physics, cosmology and the foundations of quantum theory. His 1997 book, *The Life of the Cosmos* (Oxford University Press), explored the philosophical implications of developments in contemporary physics and cosmology.

DUSAN PETRICIC

tum states of spatial volumes and areas.

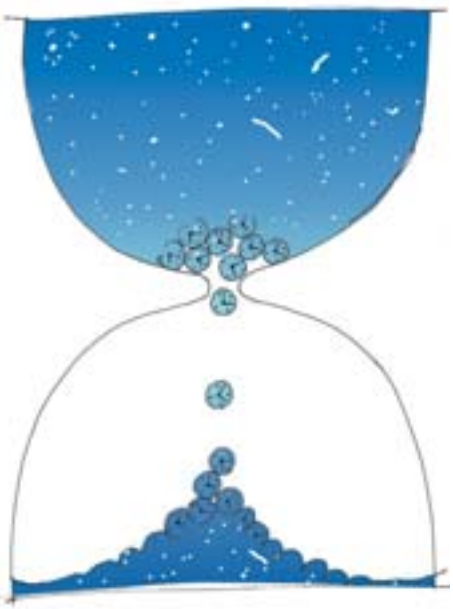
The individual nodes and edges of the diagrams represent extremely small regions of space: a node is typically a volume of about one cubic Planck length, and a line is typically an area of about one square Planck length. But in principle, nothing limits how big and complicated a spin network can be. If we could draw a detailed picture of the quantum state of our universe—the geometry of its space, as curved and warped by the gravitation of galaxies and black holes and everything else—it would be a gargantuan spin network of unimaginable complexity, with approximately 10^{184} nodes.

These spin networks describe the geometry of space. But what about all the matter and energy contained in that space? How do we represent particles and fields occupying positions and regions of space? Particles, such as electrons, correspond to certain types of nodes, which are represented by adding more labels on nodes. Fields, such as the electromagnetic field, are represented by additional labels on the lines of the graph. We represent particles and fields moving through space by these labels moving in discrete steps on the graphs.

Moves and Foams

PARTICLES AND FIELDS are not the only things that move around. According to general relativity, the geometry of space changes in time. The bends and curves of space change as matter and energy move, and waves can pass through it like ripples on a lake [see "Ripples in Space and Time," by W. Wayt Gibbs; SCIENTIFIC AMERICAN, April 2002]. In loop quantum gravity, these processes are represented by changes in the graphs. They evolve in time by a succession of certain "moves" in which the connectivity of the graphs changes [*see illustration on opposite page*].

When physicists describe phenomena quantum-mechanically, they compute probabilities for different processes. We do the same when we apply loop quantum gravity theory to describe phenomena, whether it be particles and fields moving on the spin networks or the geometry of space itself evolving in time. In particular, Thomas Thiemann of the Perimeter Institute for Theoretical Physics in Waterloo, Ontario, has derived precise quantum probabilities for the spin network moves. With these the theory is completely specified: we have a well-defined procedure for computing the probability of any process that can occur in a world that obeys the rules of our theory. It remains only to do the computations and work out predictions for what could be observed in experiments of one kind or another.



TIME ADVANCES by the discrete ticks of innumerable clocks.

Einstein's theories of special and general relativity join space and time together into the single, merged entity known as spacetime. The spin networks that represent space in loop quantum gravity theory accommodate the concept of spacetime by becoming what we call spin "foams." With the addition of another dimension—time—the lines of the spin networks grow to become two-dimensional surfaces, and the nodes grow to become lines. Transitions where the spin networks change (the moves discussed earlier) are now represented by nodes where the lines meet in the foam. The spin foam picture of spacetime was proposed by several people, including Carlo Rovelli, Mike Reisenberger (now of the University of Montevideo), John Barrett of the University of Nottingham, Louis Crane of Kansas State University, John Baez of the University of California at Riverside and Fotini Markopoulou of the Perimeter Institute for Theoretical Physics.

In the spacetime way of looking at things, a snapshot at a specific time is like a slice cutting across the spacetime. Taking such a slice through a spin foam produces a spin network. But it would be wrong to think of such a slice as moving continuously, like a smooth flow of time. Instead, just as space is defined by a spin network's discrete geometry, time is defined by the sequence of distinct moves that rearrange the network, as shown in the illustration on the opposite page. In this way time also becomes discrete. Time flows not like a river but like the ticking of a clock, with "ticks" that are about as long as the Planck time: 10^{-43} second. Or, more precisely, time in our universe flows by the ticking of innumerable clocks—in a sense, at every location in the spin foam where a quantum "move" takes place, a clock at that location has ticked once.

Predictions and Tests

I HAVE OUTLINED what loop quantum gravity has to say about space and time at the Planck scale, but we cannot verify the theory directly by examining spacetime on that scale. It is too small. So how can we test the theory? An important test is whether one can derive classical general relativity as an approximation to loop quantum gravity. In other words, if the spin networks are like the threads woven into a piece of cloth, this is analogous to asking whether we can compute the right elastic properties for a sheet of the material by averaging over thousands of threads. Similarly, when averaged over many Planck lengths, do spin networks describe the geometry of space and its evolution in a way that agrees roughly with the "smooth cloth" of Einstein's classical theory? This is a difficult problem, but recently researchers have made progress for some cases, for certain configurations of the material, so to speak. For example, long-wavelength gravitational waves propagating on otherwise flat (uncurved) space can be described as excitations of specific quantum states described by the loop quantum gravity theory.

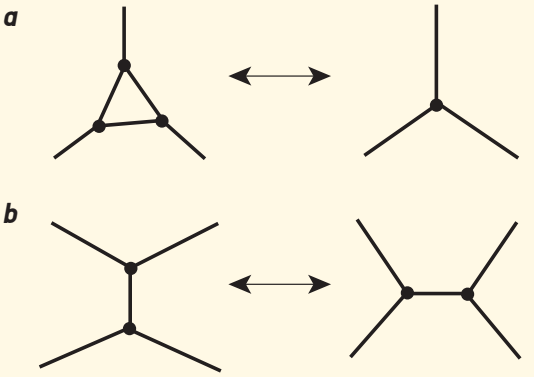
Another fruitful test is to see what loop quantum gravity has to say about one of the long-standing mysteries of

DUSAN PETRICIC

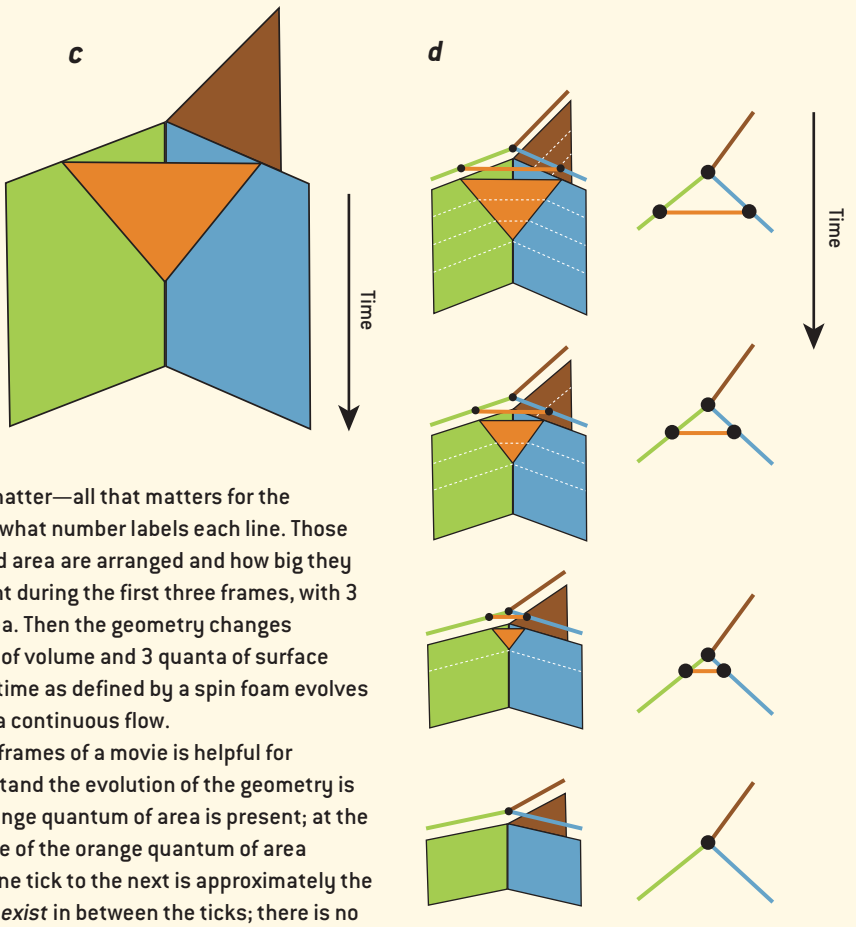
gravitational physics and quantum theory: the thermodynamics of black holes, in particular their entropy, which is related to disorder. Physicists have computed predictions regarding black hole thermodynamics using a hybrid, approximate theory in which matter is treated quantum-mechanically but spacetime is not. A full quantum theory of gravity, such as loop quantum gravity, should be able to reproduce these predictions. Specifically, in the 1970s Jacob D. Bekenstein, now at the Hebrew University of Jerusalem, inferred that black holes must be ascribed an entropy proportional to their surface area [see "Information in a Holographic Universe," by Jacob D. Bekenstein; Scientific American, August 2003]. Shortly after, Stephen Hawking deduced that black holes, particularly small ones, must emit radiation. These predictions are

EVOLUTION OF GEOMETRY IN TIME

CHANGES IN THE SHAPE of space—such as those occurring when matter and energy move around within it and when gravitational waves flow by—are represented by discrete rearrangements, or moves, of the spin network. In *a,* a connected group of three volume quanta merge to become a single volume quantum; the reverse process can also occur. In *b,* two volumes divide up space and connect to adjoining volumes in a different way. Represented as polyhedra, the two polyhedra would merge on their common face and then split like a crystal cleaving on a different plane. These spin-network moves take place not only when large-scale changes in the geometry of space occur but also incessantly as quantum fluctuations at the Planck scale.



ANOTHER WAY to represent moves is to add the time dimension to a spin network—the result is called a spin foam (*c*). The lines of the spin network become planes, and the nodes become lines. Taking a slice through a spin foam at a particular time yields a spin network; taking a series of slices at different times produces frames of a movie showing the spin network evolving in time (*d*). But notice that the evolution, which at first glance appears to be smooth and continuous, is in fact discontinuous. All the spin networks that include the orange line (*first three frames shown*) represent exactly the same geometry of space. The length of the orange line doesn't matter—all that matters for the geometry is how the lines are connected and what number labels each line. Those are what define how the quanta of volume and area are arranged and how big they are. Thus, in *d,* the geometry remains constant during the first three frames, with 3 quanta of volume and 6 quanta of surface area. Then the geometry changes discontinuously, becoming a single quantum of volume and 3 quanta of surface area, as shown in the last frame. In this way, time as defined by a spin foam evolves by a series of abrupt, discrete moves, not by a continuous flow.

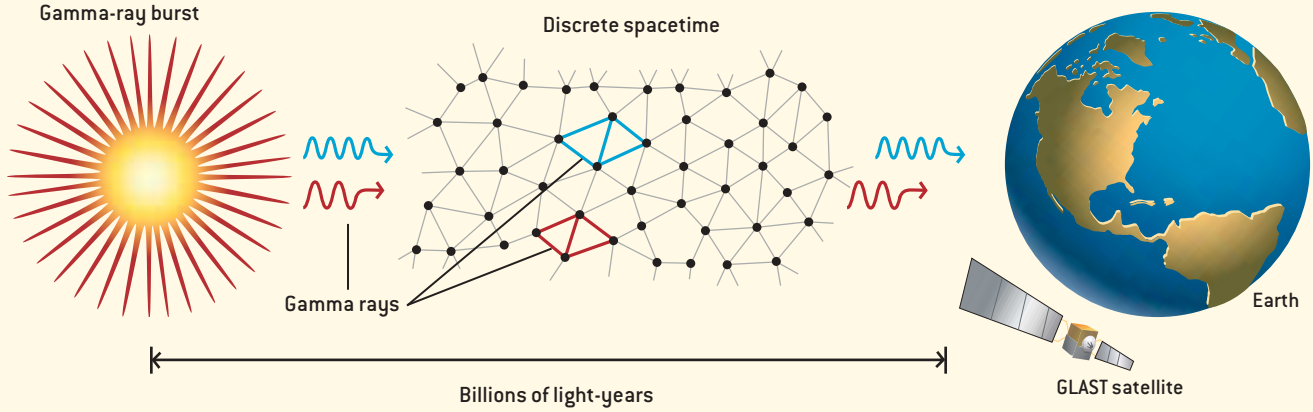


Although speaking of such sequences as frames of a movie is helpful for visualization, the more correct way to understand the evolution of the geometry is as discrete ticks of a clock. At one tick the orange quantum of area is present; at the next tick it is gone—in fact, the disappearance of the orange quantum of area *defines* the tick. The difference in time from one tick to the next is approximately the Planck time, 10^{-43} second. But time *does not exist* in between the ticks; there is no "in between," in the same way that there is no water in between two adjacent molecules of water.

SOURCE: *A* AND *B* ADAPTED FROM FOTINI MARKOPOULOU (http://arxiv.org/abs/gr-qc/9704013/); *C* ADAPTED FROM CARLO ROVELLI (http://arxiv.org/abs/gr-qc/9806121/); NADIA STRASSER

AN EXPERIMENTAL TEST

RADIATION from distant cosmic explosions called gamma-ray bursts might provide a way to test whether the theory of loop quantum gravity is correct. Gamma-ray bursts occur billions of light-years away and emit a huge amount of gamma rays within a short span. According to loop quantum gravity, each photon occupies a region of lines at each instant as it moves through the spin network that is space (in reality a very large number of lines, not just the five depicted here). The discrete nature of space causes higher-energy gamma rays to travel slightly faster than lower-energy ones. The difference is tiny, but its effect steadily accumulates during the rays' billion-year voyage. If a burst's gamma rays arrive at Earth at slightly different times according to their energy, that would be evidence for loop quantum gravity. The GLAST satellite, which is scheduled to be launched in 2006, will have the required sensitivity for this experiment.



among the greatest results of theoretical physics in the past 30 years.

To do the calculation in loop quantum gravity, we pick the boundary B to be the event horizon of a black hole. When we analyze the entropy of the relevant quantum states, we get *precisely* the prediction of Bekenstein. Similarly, the theory reproduces Hawking's prediction of black hole radiation. In fact, it makes further predictions for the fine structure of Hawking radiation. If a microscopic black hole is ever observed, this prediction could be tested by studying the spectrum of radiation it emits. That may be far off in time, however, because we have no technology to make black holes, small or otherwise.

Indeed, any experimental test of loop quantum gravity would appear at first to be an immense technological challenge. The problem is that the characteristic effects described by the theory become significant only at the Planck scale, the very tiny size of the quanta of area and volume. The Planck scale is 16 orders of magnitude below the scale probed in the highest-energy particle accelerators currently planned (higher energy is needed to probe shorter distance scales). Because we cannot reach the Planck scale with an accelerator, many people have held out little hope for the confirmation of quantum gravity theories.

In the past several years, however, a few imaginative young researchers have thought up new ways to test the predictions of loop quantum gravity that can be done now. These methods depend on the propagation of light across the universe. When light moves through a medium, its wavelength suffers some distortions, leading to effects such as bending in water and the separation of different wavelengths, or colors. These effects also occur for light and particles moving through the discrete space described by a spin network.

Unfortunately, the magnitude of the effects is proportional to the ratio of the Planck length to the wavelength. For visible light, this ratio is smaller than 10^{-28}; even for the most powerful cosmic rays ever observed, it is about one billionth. For any radiation we can observe, the effects of the granular structure of space are very small. What the young researchers spotted is that these effects accumulate when light travels a long distance. And we detect light and particles that come from billions of light years away, from events such as gamma-ray bursts [see "The Brightest Explosions in the Universe," by Neil Gehrels, Luigi Piro and Peter J. T. Leonard; SCIENTIFIC AMERICAN, December 2002].

A gamma-ray burst spews out photons in a range of energies in a very brief explosion. Calculations in loop quantum gravity, by Rodolfo Gambini of the University of the Republic in Uruguay, Jorge Pullin of Louisiana State University and others, predict that photons of different energies should travel at slightly different speeds and therefore arrive at slightly different times [*see illustration above*]. We can look for this effect in data from satellite observations of gamma-ray bursts. So far the precision is about a factor of 1,000 below what is needed, but a new satellite observatory called GLAST, planned for 2006, will have the precision required.

The reader may ask if this result would mean that Einstein's theory of special relativity is wrong when it predicts a universal speed of light. Several people, including Giovanni Amelino-Camelia of the University of Rome "La Sapienza"

NADIA STRASSER

and João Magueijo of Imperial College London, as well as myself, have developed modified versions of Einstein's theory that will accommodate high-energy photons traveling at different speeds. Our theories propose that the universal speed is the speed of very low energy photons or, equivalently, long-wavelength light.

Another possible effect of discrete spacetime involves very high energy cosmic rays. More than 30 years ago researchers predicted that cosmic-ray protons with an energy greater than 3×10^{19} electron volts would scatter off the cosmic microwave background that fills space and should therefore never reach the earth. Puzzlingly, a Japanese experiment called AGASA has detected more than 10 cosmic rays with an energy over this limit. But it turns out that the discrete structure of space can raise the energy required for the scattering reaction, allowing higher-energy cosmic-ray protons to reach the earth. If the AGASA observations hold up, and if no other explanation is found, then it may turn out that we have already detected the discreteness of space.

The Cosmos

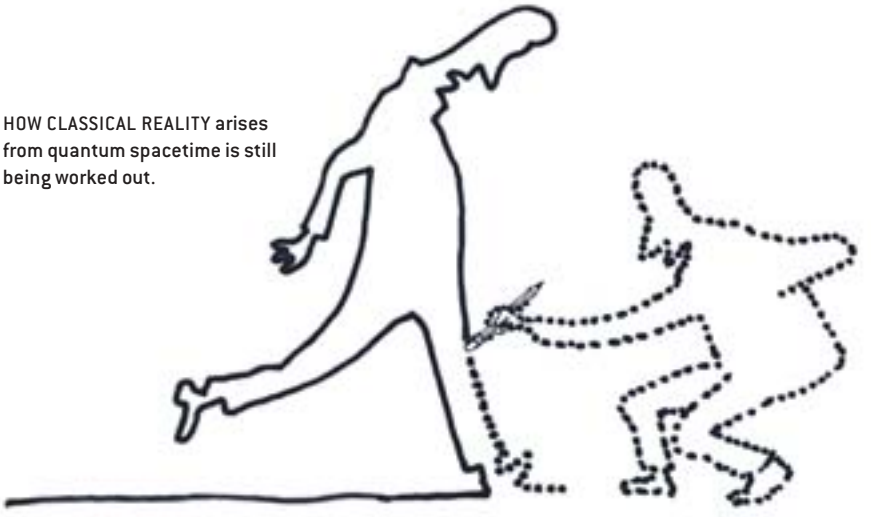
IN ADDITION to making predictions about specific phenomena such as high-energy cosmic rays, loop quantum gravity has opened up a new window through which we can study deep cosmological questions such as those relating to the origins of our universe. We can use the theory to study the earliest moments of time just after the big bang. General relativity predicts that there was a first moment of time, but this conclusion ignores quantum physics (because general relativity is not a quantum theory). Recent loop quantum gravity calculations by Martin Bojowald of the Max Planck Institute for Gravitational Physics in Golm, Germany, indicate that the big bang is actually a big bounce; before the bounce the universe was rapidly contracting. Theorists are now hard at work developing predictions for the early universe that may be testable in future cosmological observations. It is not impossible that in our lifetime we could see evidence of the time before the big bang.

A question of similar profundity concerns the cosmological constant—a positive or negative energy density that could permeate "empty" space. Recent observations of distant supernovae and the cosmic microwave background strongly indicate that this energy does exist and is positive, which accelerates the universe's expansion [see "The Quintessential Universe," by Jeremiah P. Ostriker and Paul J. Steinhardt; SCIENTIFIC AMERICAN, January 2001]. Loop quantum gravity has no trouble incorporating the positive energy density. This fact was demonstrated in 1990, when Hideo Kodama of Kyoto University wrote down equations describing an exact quantum state of a universe having a positive cosmological constant.

Many open questions remain to be answered in loop quantum gravity. Some are technical matters that need to be clarified. We would also like to understand how, if at all, special relativity must be modified at extremely high energies. So far our speculations on this topic are not solidly linked to loop quantum gravity calculations. In addition, we would like to know that classical general relativity is a good approximate description of the theory for distances much larger than the Planck length, in all circumstances. (At present we know only that the approximation is good for certain states that describe rather weak gravitational waves propagating on an otherwise flat spacetime.) Finally, we would like to understand whether or not loop quantum gravity has anything to say about unification: Are the different forces, including gravity, all aspects of a single, fundamental force? String theory is based on a particular idea about unification, but we also have ideas for achieving unification with loop quantum gravity.

Loop quantum gravity occupies a very important place in the development of physics. It is arguably *the* quantum theory of general relativity, because it makes no extra assumptions beyond the basic principles of quantum theory and relativity theory. The remarkable departure that it makes—proposing a discontinuous spacetime described by spin networks and spin foams—emerges from the mathematics of the theory itself, rather than being inserted as an ad hoc postulate.

Still, everything I have discussed is theoretical. It could be that in spite of all I have described here, space really is continuous, no matter how small the scale we probe. Then physicists would have to turn to more radical postulates, such as those of string theory. Because this is science, in the end experiment will decide. The good news is that the decision may come soon. SA



HOW CLASSICAL REALITY arises from quantum spacetime is still being worked out.

DUSAN PETRICIC

MORE TO EXPLORE

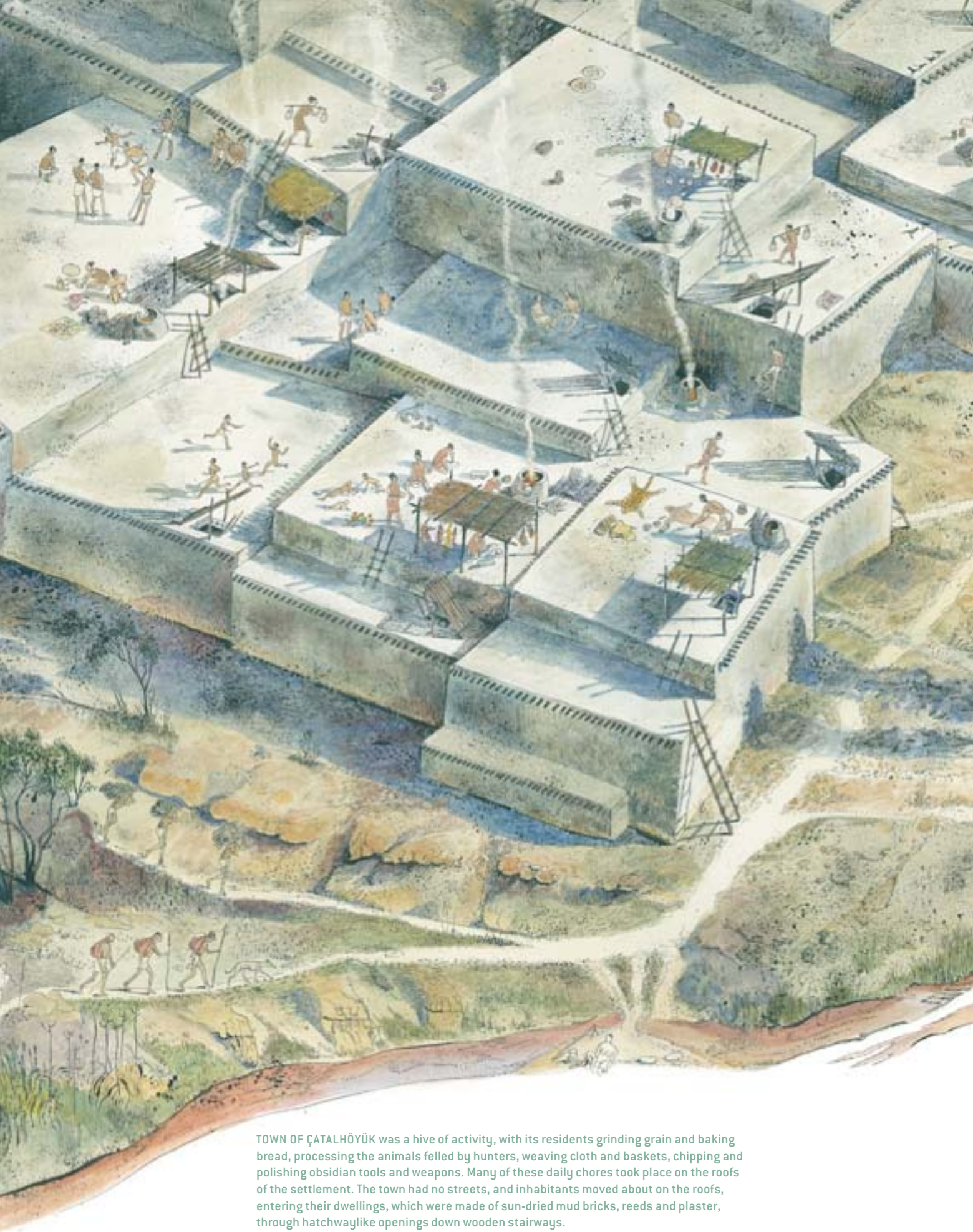
Three Roads to Quantum Gravity. Lee Smolin. Basic Books, 2001.

The Quantum of Area? John Baez in *Nature,* Vol. 421, pages 702–703; February 2003.

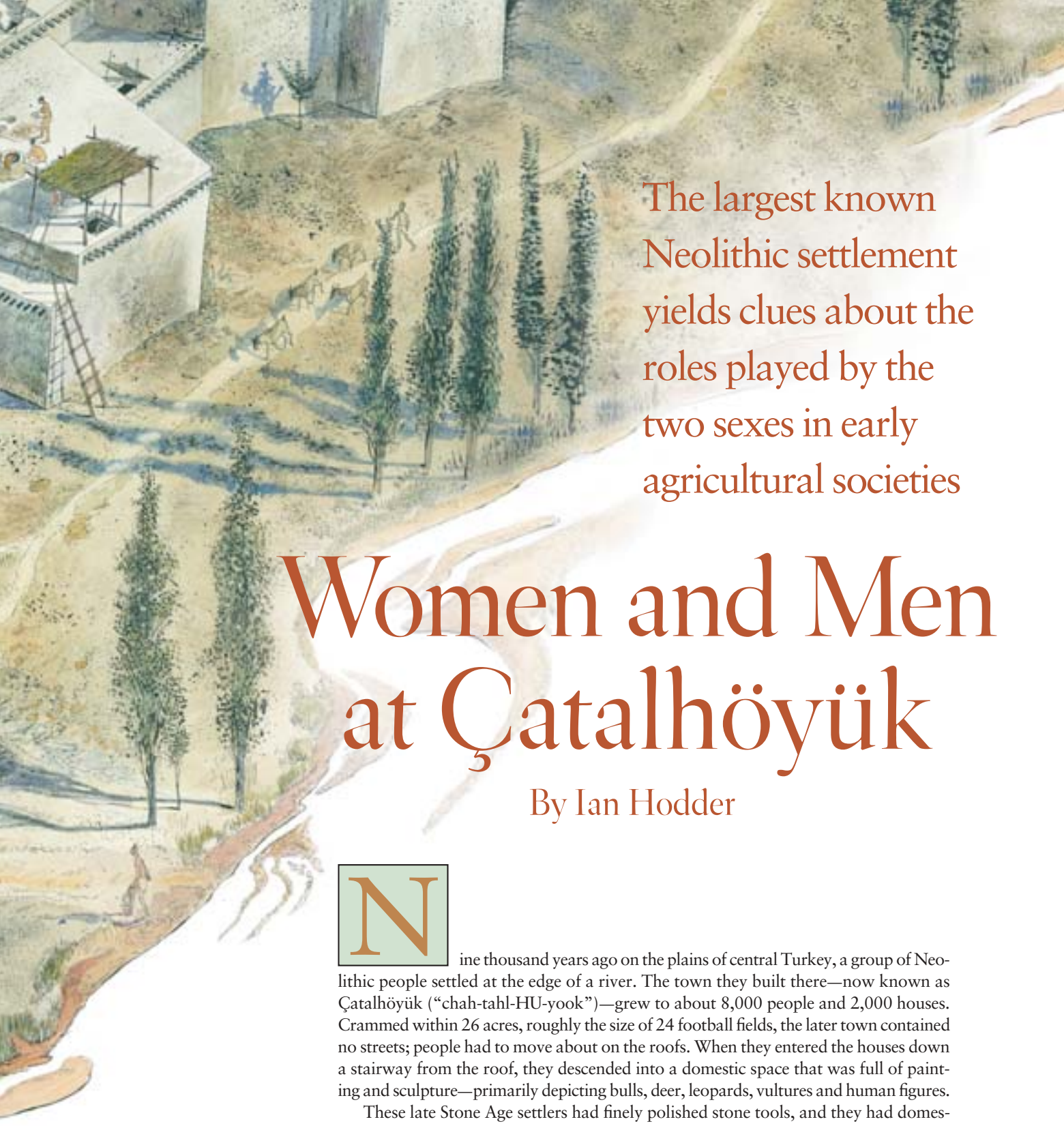
How Far Are We from the Quantum Theory of Gravity? Lee Smolin. March 2003. Preprint available at **http://arxiv.org/hep-th/0303185**

Welcome to Quantum Gravity. Special section. *Physics World,* Vol. 16, No. 11, pages 27–50; November 2003.

Loop Quantum Gravity. Lee Smolin. Online at **www.edge.org/3rd_culture/smolin03/smolin03_index.html**



TOWN OF ÇATALHÖYÜK was a hive of activity, with its residents grinding grain and baking bread, processing the animals felled by hunters, weaving cloth and baskets, chipping and polishing obsidian tools and weapons. Many of these daily chores took place on the roofs of the settlement. The town had no streets, and inhabitants moved about on the roofs, entering their dwellings, which were made of sun-dried mud bricks, reeds and plaster, through hatchwaylike openings down wooden stairways.



The largest known Neolithic settlement yields clues about the roles played by the two sexes in early agricultural societies

Women and Men at Çatalhöyük

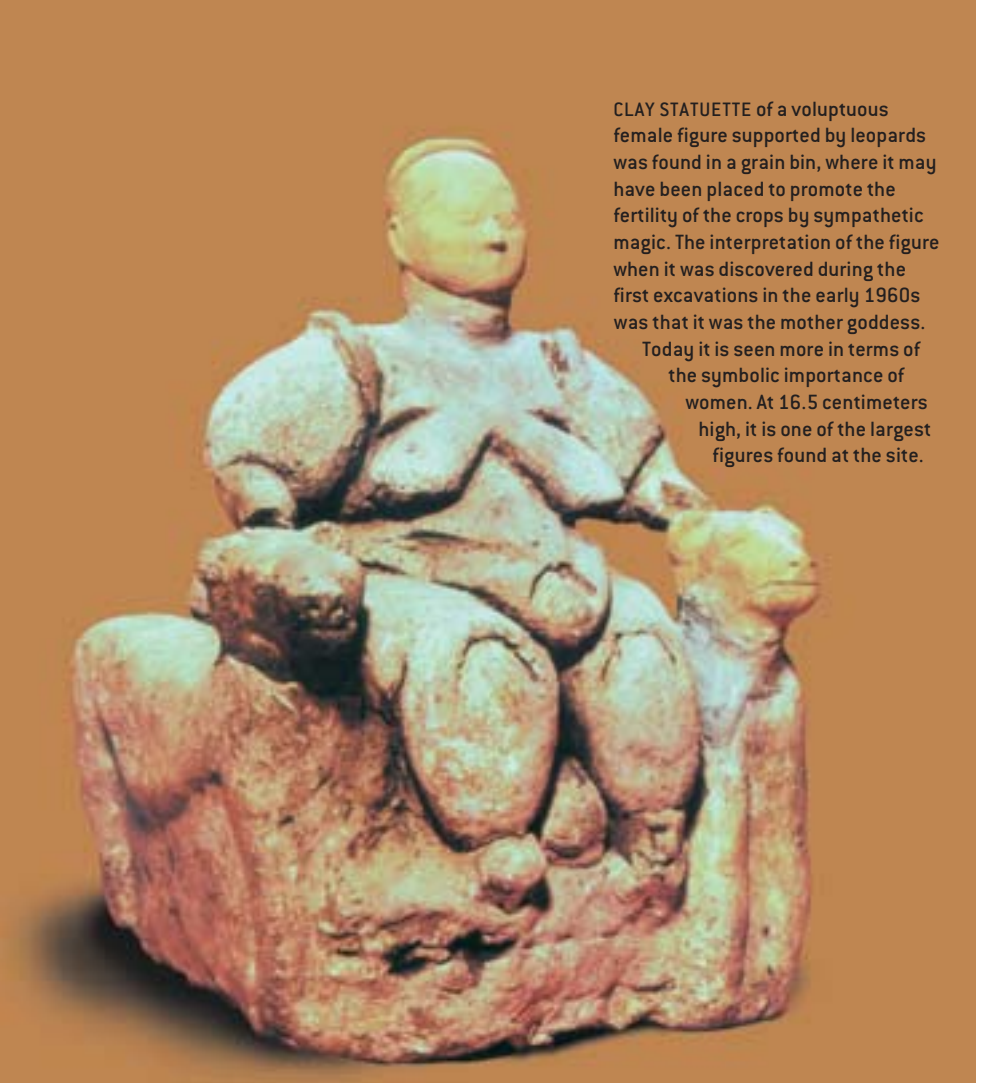
By Ian Hodder

Nine thousand years ago on the plains of central Turkey, a group of Neolithic people settled at the edge of a river. The town they built there—now known as Çatalhöyük ("chah-tahl-HU-yook")—grew to about 8,000 people and 2,000 houses. Crammed within 26 acres, roughly the size of 24 football fields, the later town contained no streets; people had to move about on the roofs. When they entered the houses down a stairway from the roof, they descended into a domestic space that was full of painting and sculpture—primarily depicting bulls, deer, leopards, vultures and human figures.

These late Stone Age settlers had finely polished stone tools, and they had domesticated cereals and sheep. In addition, they hunted wild cattle, pigs and horses and made use of many wild plants. The site is not the earliest agricultural settlement, but its large size at an early date and its elaborate art mean that it has always played a part in discussions about early farmers and their way of life.

One of the questions in which Çatalhöyük was immediately embroiled was the role of women in early agricultural societies. A long tradition in European thought holds that most of these societies were matriarchies (women were the leaders, descent was through the female line, and inheritance passed from mother to daughters) and that they worshipped a powerful mother goddess. The idea of an agricultural phase

JACK UNRUH



CLAY STATUETTE of a voluptuous female figure supported by leopards was found in a grain bin, where it may have been placed to promote the fertility of the crops by sympathetic magic. The interpretation of the figure when it was discovered during the first excavations in the early 1960s was that it was the mother goddess. Today it is seen more in terms of the symbolic importance of women. At 16.5 centimeters high, it is one of the largest figures found at the site.

in which the goddess was a potent symbol became a central tenet of the New Age goddess movements in the last decades of the 20th century, and many goddess tours have visited Çatalhöyük to pray, to hold circle dances and to feel the sway of the goddess.

Was Çatalhöyük the bastion of female power it has been thought to be? The resumption of excavations at the site in the 1990s, after a gap of a quarter of a century, has turned up fresh evidence of the relative power of the sexes at this place in central Turkey 9,000 years ago, and we can begin to answer this question—and to paint a picture of what it was like to be a woman or a man at Çatalhöyük.

Overview/*Life circa 7000 B.C.*

- A 9,000-year-old site at Çatalhöyük in Turkey reveals a curious town of thousands of houses crammed together with no streets between them.
- Inhabitants climbed down stairways from the roof to enter dwellings filled with wall paintings and sculptures.
- The lives of men and women in the town do not appear to have differed greatly. One or the other sex may have exercised more power in certain spheres—men in hunting, women in growing plants, for example—over various periods in the town's history, but evidence to date indicates that both sexes played key roles in social and religious life.

The Mother Goddess

RESEARCH ON EARLIER and later agricultural sites provided some context for thinking about this question—and warned against expecting clear-cut answers. Before the 18th century, scholars in Europe had believed, based on Aristotle and interpretations of the Bible, that the political development of society began with patriarchy. During the 18th century, however, reports from North America told of societies that traced heritage through the female line, and in the early 19th century a Swiss jurist named Johann Bachofen argued that a phase of women's social power had preceded the patriarchal family. These ideas influenced many scholars in the second half of the 19th century and throughout the 20th century, including Sigmund Freud and archaeologists such as V. Gordon Childe and Jacques Cauvin.

The first excavator at Çatalhöyük was James Mellaart of the University of London, who, with his wife, Arlette, worked at the site from 1961 to 1965. He was steeped in the scholarship of the European tradition, so it is not surprising that when he discovered opulent female imagery, such as the figurine at the left, he presumed that it represented the mother goddess. The powerful naked woman sitting on a seat of felines (probably leopards), with her hands resting on their heads, seems to conjure up precisely the tamer of nature.

Mellaart's publications about the site, complete with images of potent women, reached a wide audience, but it was another archaeologist who most effectively took up the mother goddess view of Çatalhöyük. Marija Gimbutas of the University of California at Los Angeles in a number of publications, including her 1974 book *Gods and Goddesses of Old Europe,* argued forcefully for an early phase of matriarchal society, evident at Çatalhöyük but also found across Europe with the advance of agriculture. Patriarchal societies came later, she contended, in conjunction with metallurgy, horse riding and warring.

More recently, cultural anthropologists—those scholars who compare and analyze societies—have withdrawn from making such sweeping generalizations, because human groups living today or in the recent past offer a diverse picture when it comes to the roles of the two sexes. Furthermore, cultural anthropology provides no substantiated claims for true matriarchies. The record does show, how-

JAMES MELLAART

ever, that in most recent and contemporary societies women have some form of authority or that women at certain stages in their lives, or in certain contexts, have power. Rather than talking simplistically about matriarchies and patriarchies, we should expect, according to the ethnographic evidence, a more complicated picture, which is just what we find at Çatalhöyük.

You Are What You Eat

SO FAR EXCAVATIONS at Çatalhöyük have extended over only 4 percent of the site. We have discovered 18 levels of habitation (each built on top of the previous level), covering a total of about 1,200 years. Most of our understanding to date comes from the middle and earlier levels, which have been examined most closely.

Some of our strongest scientific evidence about the relative status of men and women in the early and middle levels of Çatalhöyük concerns diet. If women and men lived notably different lives, and if one or the other was dominant, then we might expect to uncover disparities in diet, with the dominant group having more access to certain foods, such as meat or better joints of meat. So we have searched hard for such evidence, but we have not uncovered clear differences.

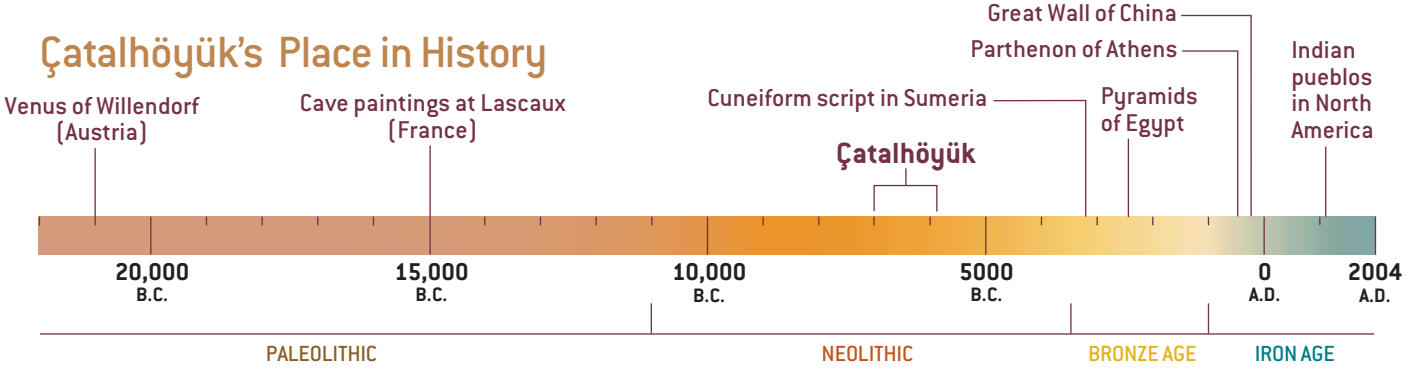
Two of my colleagues, Michael P. Richards of the University of Bradford in England and Jessica Pearson of the University of Oxford, have analyzed the stable isotopes in ancient bones at Çatalhöyük, to discover what people ate. The inhabitants of the settlement buried their dead underneath the floors of the houses, and in one building we found 62 bodies. The analysis of these skeletons detected no statistical variation between the isotopes in male and female bones. The same is true of the teeth, which were studied by Basak Boz, a graduate student at Hacettepe University in Ankara, in collaboration with Peter Andrews and Theya Molleson of the Natural History Museum in London. Women tend to have more cavities than men, but in terms of wear on the teeth the researchers found no difference.

By analyzing the patterns of wear and tear on the bones, Molleson has also been able to demonstrate that the people seem to have carried out very similar tasks during their lives [see "The Eloquent Bones of Abu Hureyra," by Theya Molleson; SCIENTIFIC AMERICAN, August 1994]. An intriguing piece of evidence supports this finding. Andrews and Molleson had noticed a black deposit often lining the inside of the ribs, which when analyzed proved to include carbon. The inhabitants of Çatalhöyük lived in small houses with little draft and with much smoke from the fire. Indeed, the wall plasters were covered with soot. The same soot got into people's lungs.

The hole in the roof through which inhabitants entered their houses was also where the smoke from the fire came out. Winters in the area are extremely cold, so families may have spent a great deal of time indoors, breathing smoky air. As a result, soot built up in their lungs. After burial and during the decay of the body, the soot was deposited on the inside of the ribs. But—and this is the crucial point for our purposes—both men and women had soot on their ribs. This finding implies that we cannot argue, for example, that men had more of an outdoor and women more of an indoor life. In fact, they appear to have lived quite similar lives in terms of the amount of time spent in the house.

ÇATALHÖYÜK originally lay beside a river and was surrounded by marshland. Today the river has dried up, and the area, 3,000 feet above sea level on the Anatolian plateau, is covered by fertile wheat fields.

Çatalhöyük's Place in History



NINA FINKEL

AT HOME IN THE STONE AGE

IN 1958 ARCHAEOLOGISTS DISCOVERED a late Stone Age settlement on a mound rising out of the plain in central Turkey. To their surprise, the site, which dated back to 7000 B.C., turned out to be far more than a village; it could rightfully be called a town. Çatalhöyük ("forked mound") had been a community of several thousand inhabitants who had an impressive social organization, a rich religious life, a high level of technology (weaving, pottery, obsidian tools), and a genius for painting and sculpture.

Researchers eventually identified 18 superimposed building levels on the 26-acre site. The habit of building one structure on top of another, using the old walls as foundations, led to a certain uniformity of plan. But by subdividing rooms or joining others together and by creating an open space in place of one or more rooms, the builders managed to vary the plans of individual levels considerably, and pleasant irregularities break the monotony of row after row of dwellings.

Nevertheless, orderliness and planning prevail: in the standard layout of houses, in the size of bricks, in the heights of panels, doorways, hearths and ovens, even in the size of rooms. The houses themselves range in size from about 11 to 48 square meters, but they are invariably rectangular. Constructed of timber frames, sun-dried mud bricks, reeds and plaster, all the buildings appear to have been only one story high. Bundles of reeds formed the flat roofs, with a thick mud cover on top and a mat below to prevent bits of reeds from falling on the inhabitants.

Entry through the roof is one of the most characteristic features of the buildings; other access was rare. Each house had a wooden stairway made of squared timber, one side of which rested along the south wall, where it has left an easily recognizable diagonal mark in the plaster. The stairs descended from a hole in the roof, and through this same opening the smoke from the hearth, oven and lamps escaped. Movable ladders provided access to the roofs from outside the town. The interiors were rich in sculptures and wall paintings—the first paintings ever found on constructed walls. The main room served as the area for cooking, eating, sleeping and many other activities. The kitchen was always along the south side, to allow smoke to escape through the entrance hole in the roof. Secondary rooms, used for storage, were entered from the main room through low doorways; one could move through them only by squatting or crawling.

Along the walls of the main room were built-in, raised platforms for sitting, working and sleeping. These were as carefully plastered as the rest of the dwelling, and they were often covered with reed or rush matting as a base for cushions, textiles and bedding. No single building provided sleeping space for more than eight people, and in most cases the family was probably smaller. Below these platforms the dead lay buried. The bodies were tightly wrapped in a crouched position and often placed in baskets.

So far no wells have been discovered. But some buildings had toilet areas; the material was taken out and put in disposal heaps used for both sanitation and rubbish. The thick ash deposits in these heaps made efficient sterilizers. Houses were kept scrupulously clean; remains of meals such as broken bones are a rarity in any building.

Nothing suggests that defense was the reason for the peculiar way in which the people of Çatalhöyük constructed dwellings with sole entry through the roof. Nor is there evidence of any sack or massacre during the 1,200 years of the town's existence so far explored. At present, the most likely explanation for the close huddling of houses is that people wanted to be buried on or near their ancestors. As nearby spaces for building new dwellings disappeared, the cramming in of houses meant that the only access possible was from the roof. What caused the final abandonment of the settlement is still far from clear.

ARCHAEOLOGISTS working at Çatalhöyük have reconstructed a typical house. They plastered the internal walls and the built-in furniture and then smoothed them with rubbing stones. Then they experimented with pigment mixes to replicate wall paintings. Using both dung and various kinds of wood, the researchers lit the oven and analyzed the residue to see how it compared with the archaeological remains of burned fuel. The house has provided insight into the use and feel of the interior space and raised many questions about light sources and problems of air circulation when the ovens were in use.

ÇATALHÖYÜK RESEARCH PROJECT

The study of the human remains showed that men were taller than women, but the variation in size was slight. The bones reveal that women were sometimes fatter in relation to their height than men. So perhaps there is some truth to the images of "fat ladies" seen in the figurines discovered by Mellaart. But overall, various lines of evidence suggest similar diets and lifestyles for women and men. We see little indication that the sexes had specialized tasks or that daily life was highly gendered.

This is not to argue that differences based on sex did not exist. An obvious one relates to childbirth. Study of the human bones has shown a high rate of infant and child mortality and several cases of burials of women with babies, perhaps indicating death during childbirth. But dietary and bone analyses give no clear sign that any divergence in lifestyle between women and men was translated into differences of status or power.

In Life and in Death

WE SOUGHT MORE information on status by looking into a custom at Çatalhöyük that seems bizarre from a 21st-century perspective. Archaeologists have excavated burials of headless bodies at the site. Most people were buried with their heads intact, and they were left like that. But in some cases, perhaps a year or so after burial, the grave was reopened and the head was cut off with a knife, leaving cut marks on the bones. These heads were then kept and used for ceremonial purposes. They were sometimes later left as part of abandonment rituals in houses. These practices are part of a wider tradition among the early farmers of Turkey and the Near East. At such places as Jericho, the skulls were plastered to re-create human features of the face.

It appears likely that the heads were removed from especially notable individuals—perhaps literally family or lineage "heads." So it was of great interest to find that the skulls of both men and women were circulated and curated, thus suggesting that lineage or family could be traced through both female and male lines.

We reach a similar conclusion when we consider another aspect of burial. The 62 burials in one building that I mentioned earlier largely occurred below platforms and spaces around the edges of the main room. A particular platform would serve for a time for burial and then go out of use. It is possible that the death of a particular person, specifically the last one buried in a spot, influenced when the shift in use took place—and these last-buried individuals are both male and female.

Archaeologists are accustomed to studying the layout of graves and of the artifacts in them to assess social distinctions. We have looked carefully to see whether men are always buried in one part of the room and women in another, whether men are buried on their left side and women on their right, whether men face one direction and women another, whether certain artifacts are found in the graves of men and others in the graves of women. Naomi Hamilton, while a postgraduate student at the University of Edinburgh, searched for such patterning. Look as she might, she could not tease out any clear distinctions. In one way, this is very frustrating, but in other ways, it is fascinating. It suggests a society in which sex is relatively unimportant in assigning social roles.

The burials imply equality, but what about the use of space within the houses during life? Archaeologists have often argued, on the basis of much contemporary study of small-scale, non-Western societies, that men would have made the stone tools, whereas women would have made the pots and done much of the cooking. The trouble with such assumptions is that one can always find ethnographic examples in which the roles are reversed. But let us for the moment allow that some sexual division of labor may have existed at Çatalhöyük when it came

BODIES of family members, such as this nine-month-old child, were usually laid on one side, often placed in a basket or on a reed mat and then buried under the floors and sleeping platforms of the dwellings. The lime plaster floors above the burials and the smoke from the ovens may have masked the odor of decay.

to activities inside the house. Each domicile contains a hearth or oven. Around the oven we find large accumulations of ashy rake-out material from the fire as well as the remains of cooking and processing cereals. So, we might conclude, the area around the oven was for food processing, and it was mainly the domain of women. One piece of evidence could be taken to support such a view: neonate burials frequently occur near the ovens.

But the ashy rake-outs also contain high densities of obsidian that had been flaked and knapped to make stone tools. The obsidian was traded from Cappadocia in central Turkey and then placed beneath the floor near the oven until pieces were taken out and made into tools. Such trading and tool production are often the province of men. If this was so at Çatalhöyük, then forming the obsidian into tools does not seem to have taken place in an area separate from that linked to do-

THE AUTHOR

IAN HODDER received his doctorate from the University of Cambridge and was professor of archaeology there from 1977 to 1999, when he moved to Stanford University. He is now chair of the department of cultural and social anthropology at Stanford and project director of archaeological research and excavation efforts at Çatalhöyük in Turkey. His books include *Symbols in Action* (1982), *The Present Past* (1982), *Reading the Past* (1986), *The Domestication of Europe* (1990), *Theory and Practice in Archaeology* (1992) and *The Archaeological Process* (1999).

ÇATALHÖYÜK RESEARCH PROJECT

mestic activity. Whether men or women made the stone tools, we find no indication of a clear separation of roles and tasks in any of the levels that have been excavated to date.

Life Is Short, Art Is Long

THE PICTURE LOOKS quite coherent thus far. When we examine how people lived their daily lives—what they ate, what they did, where they were buried, who was paramount in terms of lineage and family—we see little in the way of radical division between men and women, no evidence for either patriarchy or matriarchy.

But in the world of symbolic representation and art, we see something quite different. Here the realms of influence seem distinct. Consider first the males of the community. The abundant paintings appear to concentrate on men. By and large, the paintings do not portray women, whereas they include many figures of men, often clad in leopard skins, hunting or teasing wild animals. In some panels, these images are unmistakably men because they are bearded.

Indeed, much of the art is very masculine, and much is concerned with wild animals, a number of them male—bulls and stags with erect penises, for example. The numerous animal heads fixed to walls of the houses are mostly those of wild bulls and rams. This male focus of the art has a long tradition in Anatolia. Excavations at the earlier site of Göbekli in southeastern Turkey have found fantastic images of wild animals, often with erect penises, as well as stone phalluses around the site.

Nerissa Russell of Cornell University and Louise Martin of University College London have identified concentrations of the bones of large wild animals—predominantly bulls—at Çatalhöyük. These deposits, which contain higher proportions of bull bones than do those from daily meals, seem to be the residues from special feasts. The many paintings depicting groups of men and bulls could well commemorate such feasts or other rituals, as could the heads of bulls and other wild animals that were installed in the houses and plastered and painted.

As we have seen, however, the isotopic analysis of the human remains indicates no differences in the diets of the sexes, leading us to conclude that women as well as men participated in eating at these events. Only in the *art* connected with hunting and feasting do we see a distinction.

And what of the powerful female figure on the seat of leopards? Surely that indicates a strong image of women. Moreover, a recent find at Çatalhöyük reinforces this presumption: we discovered an intriguing female figurine that has a wild seed lodged in its back [*see illustration below*]. This connection between women and plants is also evident in the place that the famous "goddess" with leopards was found: a grain bin. And the few paintings that unmistakably depict women appear to show them gathering plants.

But aside from these few examples, the art and symbolism on the whole downplay or even deny the significance of agriculture. The houses are filled with symbolic representation: in many dwellings, one seems hardly able to move without facing some bull's head or painting. Yet in all this, the grain stores are never elaborated with any form of symbolism. The domestic pots are not painted or decorated; neither are the baskets used to store grains. The entire area of plants and agriculture is marginal in the art and symbolism. The artistic evidence, then, points to a divided world, one dominated by males and their activities involving hunting and wild animals and the other, less frequently portrayed world involving women and plants.

The situation is of course more complicated than this simple division implies. We must consider, too, the evolution of this society as it is revealed in the various levels of occupation. The figurines of "fat ladies," and especially the woman on leopards found in a grain bin, as well as the figurine of a woman with a grain lodged in her back, come from the upper

RECENT FINDING of a female figurine, only 2.8 centimeters high, with a seed embedded in her back suggests the important role women played in the nascent domestication of plants at Çatalhöyük.

ÇATALHÖYÜK RESEARCH PROJECT

PAINTING of an enormous red bull (the now extinct aurochs, *Bos primigenius*) occupied the wall of one house. The bull itself is more than six feet long, and the disproportionately small size of the male figures that surround it suggests the awe that the creature inspired. The painting may commemorate a feast or other ritual.

levels of the site—specifically, the most recent three or four in the total of 18. Although agriculture and domesticated plants had existed for centuries, key aspects of social life, as revealed in the art and in the remains of feasts, continued to focus on wild animals. In the upper levels of Çatalhöyük, however, we may be observing agricultural products becoming more central to the life of the community, with rituals taking place that involved farming. We also see in the art, particularly in the figurines, women linked to the growing of plants.

This prominence of agriculture and the role women play in it is part of a wider set of changes that occur in the upper levels of the site. In particular, we find large-scale ovens outside the houses, in courtyards, which may indicate some specialization in food production. Certainly the specialization in the manufacture of stone tools and pottery increases in these upper levels. And stamp seals appear, suggesting a greater sense of ownership. It is in this overall context that we see gender divisions becoming more marked and a specific female domain—growing plants for food—becoming more manifest.

So the picture of women and men at Çatalhöyük is complex—in a way that echoes some of the conclusions I mentioned earlier that anthropologists have reached about the allocation of power between the sexes. We are not witnessing a patriarchy or a matriarchy. What we are seeing is perhaps more interesting—a society in which, in many areas, the question of whether you were a man or a woman did not determine the life you could lead.

Both men and women could carry out a series of roles and enjoy a range of positions, from making tools to grinding grain and baking to heading a household. The depictions of feasting rituals imply that men dominated in this realm. But we can discern no sign that they had an overarching influence on other areas of life. And in any case, such male dominance came to be contested when, several millennia after the domestication of cereals, plant agriculture began to play a fuller part in the life of the community.

At this point, women and plants are linked in the art, but even here, whether the dominance of women in agriculture had much effect on other aspects of life must await further scientific study. In particular, we have much less information from the upper levels, where we found the "fat ladies" and the large-scale ovens, than we do from the earlier levels, where we have analyzed bones and teeth. Only when excavation of the upper levels is renewed over the next five years will we be able to see how this story of the emergence of images of powerful women unfolds. SA

MORE TO EXPLORE

A Neolithic City in Turkey. James Mellaart in *Scientific American,* Vol. 210, No. 4, pages 94–104; April 1964.

Çatal Hüyük: A Neolithic Town in Anatolia. James Mellaart. McGraw-Hill, 1967.

On the Surface: Çatalhöyük 1993–95. Edited by Ian Hodder. Cambridge: McDonald Institute for Archaeological Research and British Institute of Archaeology at Ankara, 1996. ISBN 0951942034.

Why Settle Down? The Mystery of Communities. Michael Balter in *Science,* Vol. 282, pages 1442–1445; November 20, 1998.

A Tale of Two Archaeologists. Robert Kunzig in *Discover,* Vol. 20, No. 5, pages 84–92; May 1999.

More information can be found online at **www.catalhoyuk.org, catal.arch.cam.ac.uk/catal/catal.html** and **www.smm.org/catal**

JAMES MELLAART

MANY PLANT SPECIES around Oxford, England, bloomed earlier in the 1990s than they did from the 1954-to-1990 mean. Among the most dramatic changes was in *Lamium album*, also known as the white dead nettle: first flowering was January 23, compared with March 18.

As temperatures rise sooner in spring, interdependent species in many ecosystems are shifting dangerously out of sync

Spring Forward

By Daniel Grossman

GUSTO/SPL/PHOTO RESEARCHERS, INC.

Growing up in England in the 1950s, Alastair Fitter spent a lot of time wandering with his father through the countryside near their home.

The elder Fitter, Richard, now 90, is a noted naturalist (who has written almost three dozen books on flowers, birds and related topics). As a hobby, Richard jotted down the first flowering date of hundreds of plant species, the spring arrival time of scores of birds, the late-summer departure dates of butterflies, and other signs of the passing seasons. Richard, who insists that he is simply "an inveterate list maker," never thought the records would serve any scientific purpose: "When I was 10, I read I should be keeping notes."

Alastair grew up to be a naturalist like his father as well as a professor of ecology at York University. His father's notes, he realized as an adult, were one of the few cases in which a single observer had systematically recorded the timing details of so many species in one place for so long. So in 2001, when the elder Fitter moved from his home, the locus of 47 years of systematic observations in one place, Alastair decided to take a closer look at the scribblings. By then, climate researchers had confirmed that the earth is getting warmer with stunning speed. The near-surface temperature of the planet has risen about 0.6 degree Celsius (about 1.1 degrees Fahrenheit) in the past 100 years. The 1990s was the warmest decade on record. He thought his father's data might confirm with plant life what researchers had shown with thermometers.

What Alastair Fitter discovered astonished him. An analysis of the records that he had done in the early 1990s had not shown a consistent pattern. But comparing flowering dates for the entire decade of the 1990s with those of the previous four decades, he found that 385 plants were flowering an average of 4.5 days earlier [*see illustration on opposite page*]. A smaller subset, 60 species in all, flowered on average two full weeks earlier, an astounding change for a single decade. It shows that, at least in the neighborhood of Oxford, England, "climate change is happening with extreme suddenness," Richard Fitter says.

The research, which the Fitters published jointly in *Science* in 2002, was only one of the more startling of a number of recent studies showing rapidly occurring changes in the life patterns of the world's plants and animals. Also in 2002 the Intergovernmental Panel on Climate Change (IPCC) published an overview of this topic based on a review of 2,500 published papers. A number of these articles reported on the relation between species and temperature for at least the past 20 years. Of the more than 500 birds, amphibians, plants and other organisms studied in these publications, 80 percent had changed the timing of reproduction or migration, length of growing season, population size or population distribution in ways that might be expected from warming temperatures. The overview's authors concluded that "there has been a discernible impact of regional climate change, particularly increases in temperature, on biological systems in the 20th century."

As with the Fitters' study of plant flowering, most of the studies reviewed by the IPCC did not investigate whether changes have been or will be harmful, but Alastair Fitter nonetheless believes adverse effects are inevitable: "When things flower may be relatively innocuous. Whether things go extinct is not, and that's going to be the next stage."

A small group of studies is taking up the challenge of looking into whether global warming is having an adverse effect on the relations among plants and animals within ecosystems. The research is proving that, in some cases, Fitter's gloomy prophecy is already becoming a reality: rising temperatures are degrading the links of food chains and the fitness of some creatures to continue to live in their habitats. In at least one instance, a researcher predicts that global warming will extirpate one species from an entire region

Overview/*Ecosystems under Stress*

- The recent surge in warming at the planet's surface has begun to alter the relationships among species in some ecosystems by weakening links in the food chain—between, for example, birds and the caterpillars they eat.
- Although the examples are not definitive, the signs are troubling, and many species could be at risk.
- Adding to the concern is research showing that the changing climate at the end of the last ice age tore apart existing ecosystems and created new ones, leaving no haven for species that no longer fit in.

ECOLOGIST Alastair Fitter (*below*) and his father, Richard, documented that first flowering dates for hundreds of plants in England have shifted in recent years. Although many plants now bloom significantly earlier, a few, such as this butterfly bush (*Buddleja davidii, left*), are delayed.

within 15 years. Although the data are insufficient so far to prove that many ecosystems are coming apart, the findings already point in a disturbing direction.

Earliest Birds and the Worm

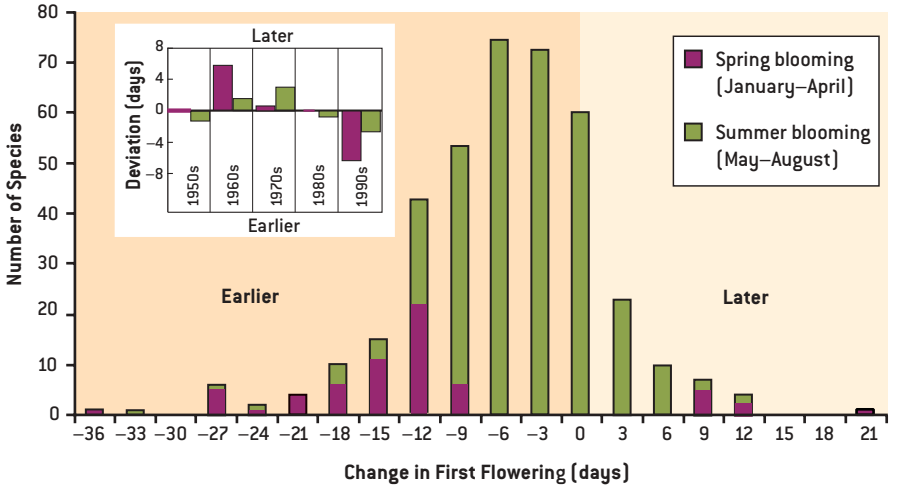
IN A SMALL OFFICE at the Netherlands Institute of Ecology, not far from Arnhem, Marcel E. Visser is crawling under a table. Visser is the head of the institute's animal-population biology department. Right now he is stalking a small chickadeelike bird that entered the room through an open window. But before the scientist can grab the bird or shoo it out, the frightened animal takes flight and leaves of its own accord.

Coincidentally, the intruder was a great tit, the subject of a long-term study that Visser heads. The study was not designed to investigate global warming. When Visser's predecessors instituted the research in the 1950s, it was intended to contribute to the general understanding of bird populations. Under Visser's leadership, however, the research has become one of the only studies in the world examining the cascading effects of global warming on a food chain.

Great tits at De Hoge Veluwe National Park, a large wooded area near Visser's office, engage in their annual nesting rituals in April and May. At the same time, Visser engages in his own rites of spring, monitoring these birds and their offspring. The scientist and a crew of assistants record the activities and health of every breeding pair in about 400 wooden birdhouses in the park. One mid-May morning, carrying a short aluminum ladder on his shoulder, the ornithologist mounts a sturdy yellow bike and sets off over the park's unpaved paths.

After several minutes of pedaling, he parks his bike by a nest box mounted a few feet above eye level in a spindly oak. With help from the ladder, he scales the tree, installs a metal trap in the box and waits a short distance away. Within moments the trap catches one of the two adults that nest in the box. Visser carefully swings the top open and removes the bird. It has a gray back, a black-and-white head and a pale yellow breast with a black stripe and weighs only about as much as a AA battery. Gently he locks the bird's head between his pointer and middle finger and makes some measurements with a ruler. He weighs the tiny animal in a sandwich bag. Multiplied thousands of times a season, these statistics are critical raw material for Visser's research. He and his colleagues visit each box weekly except as hatching (and, later, fledging) approaches, when they make daily checks.

What Visser has found sounds harmless enough: the tits laid their eggs at almost the same time last year as they did in 1985. But over this same period, spring temperatures in the area have climbed, especially during the mid-spring (between April 16 and May 15), which has seen a warming of two degrees C. And although the chronology of the tits has not changed with this warming, that of winter moth caterpillars—which (along with other less abundant species) tits feed their chicks—has [*see illustration on page 89*]. Cater-



FIRST FLOWERING DATES of 385 plant species from 1991 to 2000—the warmest decade on record—were an average of 4.5 days ahead of the 1954-to-1990 mean. For clarity, two species with extreme deviations, *Lamium album* [−55 days] and *Buddleja davidii* [+36 days] are omitted. A graph of the first flowering for each decade compared with the long-term mean (1954–2000) underscores the change that occurred in the 1990s (*inset*).

WOLFGANG KAEHLER *Corbis* (*left*); DANIEL GROSSMAN (*right*); SOURCE: ALASTAIR AND RICHARD FITTER, *Science*; NADIA STRASSER (*graph*)

GROWTH RECORDS (*above*) made of the great tit, repeated thousands of times a season by Marcel E. Visser (*left*) and his team at the Netherlands Institute of Ecology, shed light on the local ecosystem's reaction to climate change. The research, begun in 1955 to help understand the species' population dynamics, has become one of the best examples of how climate change can disturb the links of a food chain.

pillar biomass—or the total meat available to the birds—peaks two weeks earlier today than it did in 1985. Back then, it occurred almost precisely when the tit hatchlings needed it most. Now, by the time most chicks have hatched, the caterpillar season is on the wane and food is becoming scarce. Only the earliest chicks get the worms.

It is not just the birds and the moths in this food web that are getting out of synchrony, or "decoupled," as Visser likes to say. The scientist also looks lower down the food chain to the relation between the moth and its food—young, tender oak leaves. To survive, the moth's caterpillar must hatch almost precisely at "bud burst," when the oaks' leaves open. If the insect hatches more than about five days before bud burst, it will starve. It will also starve if it hatches more than two weeks too late, because oak leaves become infused with inedible (to the caterpillar) tannin. Visser has discovered that at the De Hoge Veluwe park, oak bud burst now occurs about 10 days earlier than it did 20 years ago. Caterpillars hatch 15 days earlier, overcompensating by five days for the change in the oaks. The caterpillars were already hatching several days before bud burst in 1985, so now they must wait on average about eight days for food.

Visser's research shows that the winter moth population at De Hoge Veluwe is declining, but he has not collected moth numbers long enough to be sure this is not part of a natural cycle. The gap between the schedules of the caterpillars and the birds has had no demonstrable effect so far on tit numbers. The scientist says that could be because normal year-to-year fluctuations caused by various factors such as the availability of winter food are greater so far than the impact of warming. In a system where "timing is everything," however, Visser observes that the decoupling between links in the food chain cannot continue growing without consequence. "It's only a matter of time before we see the population come down," he says of the birds.

According to Visser, what is most worrisome about his research is not that tits at De Hoge Veluwe could be on the verge of a decline but that the decline suggests that many other species are also in danger. "I'm sure if we go to other food chains we'll find the same thing," he states. His findings suggest vulnerabilities to climate change that are universal to all

THE AUTHOR

DANIEL GROSSMAN is a journalist who has been covering environmental topics for 17 years. His recent focus, which has brought him to Greenland and Antarctica in a single year, is global warming and its impact on ecosystems.

DANIEL GROSSMAN

Organisms have come to synchronize their life cycles using certain cues. But old rules do not work anymore.

ecosystems. Some scientists say that these weaknesses are best understood using a vocabulary invented in the 1960s by marine biologist David Cushing, formerly the deputy director of the Fisheries Laboratory in Lowestoft, England.

Match and Mismatch

IN ATTEMPTING to explain year-to-year variations in herring stocks, Cushing looked at phytoplankton, the food of the herring larvae. He showed that when the hatching of the herring larvae coincided with phytoplankton blooms, a high proportion of that year's offspring survived to adulthood. This happy state of affairs, which leads to elevated larvae numbers, he called a match. A mismatch, in contrast, is when larvae hatchlings are out of sync with their food, causing a poor year.

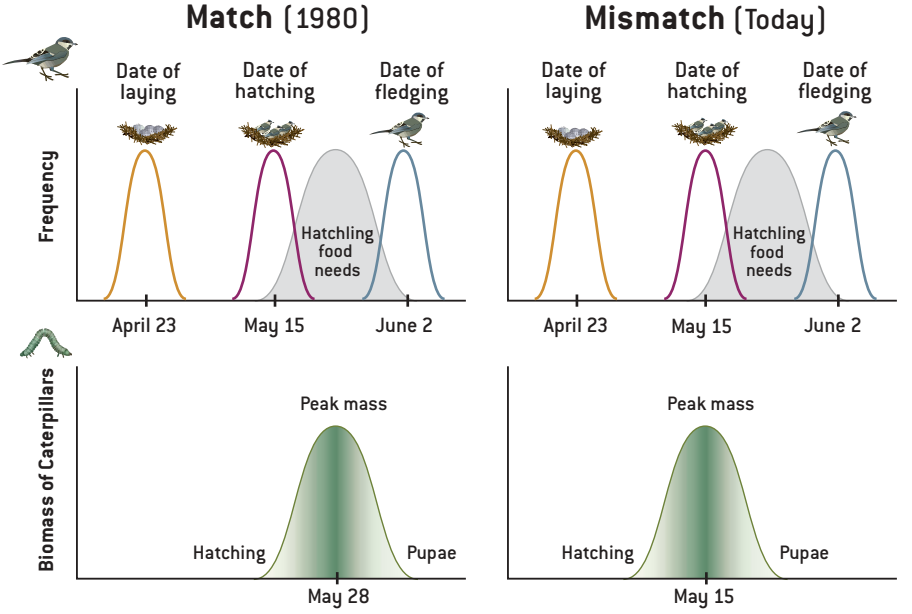
The conceptually simple match-mismatch hypothesis is now being applied by a number of researchers to explain the impact of climate warming. The power of the idea comes in part from the fact that a match can refer to different kinds of relationships. For instance, it can describe temporal relationships between predator and prey—as in the case of the tits and the caterpillars—or animal and plant, as in the case of the caterpillars and the oaks. It can be applied to the temporal relationship between different plants. For example, the Fitters found that recent changes in the flowering times of plants are not uniform across species.

Alastair Fitter says that this kind of discrepancy means, among other things, that competition for sunlight, nutrients and water will be altered, with, as he and his father wrote in *Science,* "profound ecosystem and evolutionary consequences." Finally, the match-mismatch concept can be applied to the relation between animals or plants and their physical surroundings. One example is research in Colorado showing that American robins that migrate to high-altitude summer habitats arrive early and wait longer for the winter snow to melt before mating.

Because matches often require synchrony between disparate species, it is no surprise that climate change creates mismatches. Some species are influenced by average temperatures, whereas others respond only to extremes, such as cold snaps.

Tits, oaks and moths, for instance, all seem to respond to temperature in some fashion, though each in a different way. The hatch date for tit eggs appears to be determined about a month earlier, when the eggs are laid. Visser says the birds base the laying date at De Hoge Veluwe on early spring temperatures, which, in contrast to late spring temperatures, have not changed in the past 30 years. Visser has discovered that the hatch date of moth eggs appears to be related to a combination of two factors: the number of winter and early-spring frost days (days when temperatures dip below freezing) and temperatures in late winter and early spring. The winter and early spring temperatures at De Hoge Veluwe have increased in recent decades, although there has been no change in frost days. Finally, the oaks appear to adjust the timing of bud burst depending in part on late spring temperatures, which have risen by two degrees C since 1980. Over thousands of years of evolution, these three organisms have come to synchronize their life cycles using these cues. But climate warming has decoupled the cues so that the old rules do not work anymore.

Long-distance migrants face special challenges. They may need to use cues in one habitat to determine when to depart for another. In the case of migrating birds, arrival at the summer breeding site may require precise timing of departure from wintering grounds. But wintering-site cues will not necessarily shift in synchrony with changes in nesting areas, especially if summer and winter sites are separated by thousands of kilometers, as is the case with many birds. One reason is that climate is not changing uniformly across the globe. Compared with temperate regions, for example, the tropics are hardly warming at all. The varied effects



INTERDEPENDENT SPECIES can be "uncoupled" under stress from global warming. In De Hoge Veluwe National Park in the Netherlands, changes in weather patterns have caused oak buds to burst into leaf sooner. As a result, winter moth caterpillars—an important food that great tit chick hatchlings need to reach fledging size—peak in total biomass earlier today (*right*) as compared with two decades ago (*left*). Egg-laying time has not shifted.

SOURCE: MARCEL E. VISSER, CHRISTIAAN BOTH AND MARCEL M. LAMBRECHTS; NADIA STRASSER

Many species are already responding to global warming by moving their ranges. Less nimble ones will be left behind.

of warming on El Niño and other such climate phenomena confuse the picture further. Moreover, because temperatures in the tropics are not strongly correlated with those of temperate regions, many birds do not even use climate cues to decide when to depart tropical wintering sites. Instead they regulate their travels by the length of the day. Of course, global warming has no influence on day length. Thus, such birds are now in danger of arriving at temperate breeding grounds on a date that no longer makes sense.

Christiaan Both of the University of Groningen in the Netherlands says that the pied flycatcher, which migrates 5,000 kilometers from tropical West Africa to De Hoge Veluwe, appears to be suffering from such a mismatch between nesting and wintering-site conditions. Like great tits, the flycatchers feed their hatchlings caterpillars, which reach peak abundance half a month earlier than they did 20 years ago. But the flycatchers are arriving on virtually the same date today as they did in 1980. In a 2001 *Nature* article, Both and his co-author, Marcel Visser, observe that the flycatchers' cue to leave Africa is day length, which explains why the arrival time has not changed. "They have a decision rule that has become maladaptive," Both notes.

The birds have compensated for their late migration by shortening their rest time after they reach the Netherlands. The interval between arrival and breeding has shrunk by 10 days since 1980. But even that is not enough to produce hatchlings in synchrony with the caterpillar peak. Today only the earliest flycatchers have healthy chicks. The rest have underweight offspring, most of which do not return to breed the following year. The birds have stopped nesting in forests with the earliest caterpillar season. Both says that to date, flycatcher numbers do not appear to have dropped off, but that could change if warming continues, because the birds probably cannot reduce the time between arrival and mating any further: "Now they have used up all the safety margin." Both and Visser speculate in *Nature* that the same mechanism that is affecting the pied flycatchers may be one of the factors behind why a number of other European migrants have declined in recent years.

Penguin in a Coal Mine

THERE ARE VARIOUS warning signs but so far few cases in which a mismatch caused by climate change has actually had a serious impact on a plant or animal population. Montana State University ecologist William Fraser, however, says that he has evidence that climate change is causing the extinction of Adélie penguins living on the western coast of the Antarctic Peninsula. Over the past 30 years, Fraser has documented a startling 70 percent decline in the number of Adélie penguins nesting on a number of islands in the vicinity of Palmer Station, one of three U.S. research bases in Antarctica. He believes that climate change is affecting the birds through a mechanism not previously suspected, proving the difficulty of anticipating how nature will respond to warming temperatures.

The Antarctic Peninsula has experienced more warming than almost any other place on the earth. In the past 50 years, winter temperatures in this part of the Antarctic Peninsula have climbed by almost six degrees C. Counterintuitively, this warming has increased snowfall. That is in part because sea ice, which forms an impermeable cap on the ocean, has decreased with rising temperatures, permitting more moisture to escape into the air. That moisture falls down as snow. Fraser says that colonies sustaining the worst losses are located on the southern slopes of hillsides in the rocky nesting grounds. These areas, he has discovered, are also sustaining the greatest impact from the extra snow falling along the Antarctic Peninsula. Southern slopes, in the lee of the prevailing winds of this region's winter storms, are the last to melt in the spring because they collect snowdrifts in winter and receive relatively little of the sun's warmth (southern slopes are sunniest in the Northern Hemisphere; the opposite is true in the south). But Adélies are "hardwired," Fraser explained in an interview at Palmer Station, "to need the timing of conditions to occur in a very precise chronology," and the extra snow is altering that chronology.

ADÉLIE PENGUIN is from a population near the U.S. research base at Palmer Station in Antarctica that has dropped 70 percent in three decades.

JOHN CONRAD *Corbis*

Shortly after they arrive at these islands to breed in October, the birds need bare ground to build their pebble nests. If the snow does not melt in time, they will occasionally try nesting on top of it, but that does not work. When the snow eventually melts, the eggs in such nests become waterlogged, yielding "addled" eggs, not live chicks. Actual precipitation varies from year to year, but over time increased snow has gradually cut the number of new members added to the colonies, which are dying by attrition. Fraser calls the penguins' dilemma a "mismatch between physics and biology." He predicts that as the process continues, more of each island will be affected, extirpating Adélies from the region within 15 years. He says that the Adélies are "extremely sensitive indicators of climate-induced perturbations," suggesting that important changes may be afoot elsewhere. They are

NESTING ADÈLIE PENGUINS suffer from increased snowfall, a consequence of climbing temperatures on the Antarctic Peninsula's western coast. Melting snow "addles" eggs, so they do not hatch.

"one more piece of evidence that our planet is changing."

Researchers say that in many cases, plants and animals will be able to adapt to changing conditions, avoiding the fate of Fraser's Adélie penguins. Christiaan Both is doing genetic studies of pied flycatchers to see if certain families migrate earlier, raising the possibility that evolution will produce an early migrating subspecies. In at least some instances, however, it is unlikely that an evolutionary trick will be able to solve the problem climate warming has created. For example, it might seem that evolution could produce an early-laying great tit that could remain in synchrony with caterpillars in the oak trees in the De Hoge Veluwe park.

But there is a catch: before adult female tits can produce eggs, they must nourish themselves. They do so by foraging for insects in a different group of trees, primarily larch and birch, which open their leaves before the oaks. These trees have not changed the date of bud burst nearly as much as the oaks have. If this trend continues and if the insects in these trees remain in sync with larch and birch bud burst, the tits could not possibly be ready to breed any earlier in the season.

Many species, such as certain birds and insects, are already responding—and will most likely continue to respond—to global warming by moving their ranges farther north or, in mountainous areas, to higher elevations. Less nimble species, such as trees, will be left behind. Terry L. Root, a biologist at the Center for Environmental Science and Policy at Stanford University, warns that these differential responses will cause existing ecosystems to be "torn apart," leaving altered, more impoverished ones in their place. The problem is compounded by human uses of land, such as cities, farms and highways, which fragment the landscape. Root published a paper in *Nature* last year with evidence of "fingerprints" of global warming on wild plants and animals. She says pollen studies of the end of the last ice age suggest what might happen. Such research shows that as the ice sheet that covered most of North America retreated north, forests were not far behind. But complete communities did not simply move north in synchrony with rising temperatures. Instead the blend of plants and animals changed as the forests moved. Root fears that global warming could produce similar results, creating ecosystems unknown today and casting off species that no longer fit in.

There is, of course, much more research left to do to determine how serious and widespread the problem is. Certain conundrums will also need to be resolved. For example, why are the great tits in the De Hoge Veluwe park failing to remain in sync with winter moth caterpillars, while birds of the same species at a research site near Cambridge, England, just 400 kilometers away, have altered life cycles in step with the same insects? One thing, says Alastair Fitter, is clear: "The natural world is paying attention to what's going on in the climate." And, he adds, "It's going to get worse." SA

MORE TO EXPLORE

How Climate Change Alters Rhythms of the Wild. Bernice Weuthrich in *Science,* Vol. 287, No. 5454, pages 793–795; February 4, 2000.

Adjustment to Climate Change Is Constrained by Arrival Date in a Long-Distance Migrant Bird. Christiaan Both and Marcel E. Visser in *Nature,* Vol. 411, pages 296–298; May 17, 2001.

Climate Change and Biodiversity. Edited by Habiba Gitay, Avelina Suárez, Robert T. Watson and David Jon Dokken. IPCC Technical Paper, 2002. Available at **www.ipcc.ch/pub/techrep.htm**

Rapid Changes in Flowering Time in British Plants. A. H. Fitter and R.S.R. Fitter in *Science,* Vol. 296, pages 1689–1691; May 31, 2002.

Wildlife Responses to Climate Changes: North American Case Studies. Edited by Stephen H. Schneider and Terry L. Root. Island Press, 2002.

GALEN ROWELL *Corbis*

THE CURIOUS HISTORY of the First POCKET CALCULATOR

It was called the Curta, and it proved lifesaving when its inventor was trapped in a Nazi concentration camp

By Cliff Stoll

Johannes Kepler, Isaac Newton and Lord Kelvin all complained about the time they had to waste doing simple arithmetic. Foolscap covered with numbers obscured answers; elegant equations led to numerical drudgery.

Oh, for a pocket calculator that could add, subtract, multiply and divide! One with digital readouts and memory. A simple, finger-friendly interface.

But none were available until 1947. Then, for a quarter of a century, the finest pocket calculators came from Liechtenstein. In this diminutive land of Alpine scenery and tax shelters, Curt Herzstark built the most ingenious calculating machine ever to grace an engineer's hand: the Curta calculator.

Advertisements in the back pages of *Scientific American* in the 1960s promised an arithmetic panacea: "The Curta is a precision calculating machine for all arithmetical operations. Curta adds, subtracts, multiplies, divides, square and cube roots … and every other computation arising in science and commerce … Available on a trial basis. Price $125."

With its uncanny resemblance to a pepper grinder, this device—still owned by some lucky people—does everything that your $10 pocket calculator can do. Except that it's entirely mechanical—no battery, no keypad, no liquid-crystal display. You turn a crank to add numbers.

A windup adding machine? You bet. Today I'm holding a Curta in my left hand, grinding out answers with my right. To add, I enter numbers with little sliders, spin the crank, and the result appears in small windows circling the top. I'm literally crunching numbers.

And yes, it multiplies and divides—although I have to spin that crank 10 or 20 times to find the product of two big numbers. There's no on/off switch, but a handy finger ring clears the memory. As for square and cube roots, well, you carry along special tables and remember a few shortcut algorithms.

This is no slide rule that approximates an answer to three or four places. Through the windows on the top, 11 numbers click into place. Hey—your electronic calculator probably can't deliver 11 digits of precision.

Okay, it does arithmetic. So why has the Curta been called "a treasure of our civilization" and "a marvel of technology"? Why do collectors cherish these devices when any cheap calculator works much faster?

Because along with its impressive arithmetical abilities comes the sensation of mechanical elegance and certainty: You set numbers by sliding dials that slip into place with little curtsies. The crank turns with the smoothness of a fine pocket watch. Digits click into position with neither slop nor drag. Each number is engraved in magnesium, and steel gears handle the computation. The Curta purrs as you calculate.

Then, too, this machine was designed to make calculating easy. To avoid errors, separate displays show the entered number, how many times you turn the crank, and the result. De-

IS IT A PEPPER MILL? A camera lens? A pencil sharpener? The handle of a fishing reel? No, it's a mechanical calculator more precise than many electronic varieties. The example depicted on the opposite page is only about four inches tall and two inches in diameter.

TINA WEST

tents—small catches—tell your fingers when each digit is entered and when the answer is ready. It's easy to undo an error, but a ratchet won't let you clash the gears by going backward. And you won't erase your answer by mistake, because the clearing ring can't be accidentally activated.

The Curta calculator combines the precision of a Swiss watch, the craftsmanship of an old Nikon F camera and the elegance of a tango—all in a compact cylinder. In 1950 the Curta's portability startled engineers: a calculator that you could carry! All the more astonishing, then, that this device arose from the nadir of civilization, the Buchenwald concentration camp.

Serious Shrinkage Problems

JUST AS TODAY'S technicians crave featherweight laptop computers, engineers and accountants long yearned for a portable mathematics machine. Thomas de Colmar built an adding machine the size of a piano for the 1855 Paris Exhibition. Fifty years later the Millionaire Calculator could not only add and subtract, it could directly multiply and divide. Yet it weighed more than 60 pounds. For a real pocket calculator, civilization would wait for Curt Herzstark.

> "I looked at everything backwards. What does this machine have to look like so that someone can use it?"

Born in 1902, Herzstark grew up around calculators. His father sold Remington and Burroughs office machines in Vienna. Within a few years his family built a factory to make calculators. It thrived, and young Curt found himself demonstrating adding machines across Austria.

During World War I, his family's factory made war material. Afterward, with the factory's machines worn out or destroyed, his father took to selling used calculators until the factory could be rebuilt. At the same time, new competition appeared, including Fritz Walther, who had made automatic pistols but now found himself stymied by postwar disarmament. Seeing opportunity in office equipment, he converted his gun factory to one that made electric adding machines.

In the 1930s the calculator business multiplied. "But something was missing in the world market," Herzstark later recalled. "Wherever I went, competitors came with wonderful, expensive, big machines. I'd talk with a building foreman, an architect or a customs officer who'd tell me, 'I need a machine that fits in my pocket and can calculate. I can't travel 10 kilometers to the office just to add a row of numbers.'"

Manufacturers such as Monroe, Friden and Marchant tried shrinking big desk models, like a watchmaker miniaturizing a clock into a wristwatch, but without much success. Visit an antique store, and you may bump into a miniature calculator of yesteryear. The Marchant "lightweight" adding machine weighed 34 pounds, with nine columns of buttons and a carriage sporting 18 mechanical readouts. Two big cranks sprout from the side, reminiscent of a Model T Ford. Accountants lugged them around in suitcases. That's what portable meant in 1935.

Having seen failed attempts to shrink desktop adding machines, Herzstark, then in his 30s, started anew. "I looked at everything backwards. Let's pretend that I have already invented everything. What does this machine have to look like so that someone can use it?

"This machine can't be a cube or a ruler; it has to be a cylinder so that it can be held in one hand. And holding it in one hand, you would adjust it with the other hand, working the sides, top and bottom. The answer could appear on the top."

Like a good software engineer, Herzstark began with the user interface, rather than letting the mechanism control the design. Instead of using a typewriter-style keyboard, he decided to wrap sliders around a cylinder so you would enter numbers by sliding a thumb or finger. This approach would create an area for the results register around the top of the cylinder, as well as a convenient site for a crank to power the calculator.

Other mechanical calculators used a separate mechanism to calculate every digit in the answer. For instance, the Friden calculator had 10 columns of keys to enter a number, and it had 10 separate sets of calculating gears—expensive and heavy. Herzstark realized that he needed only one calculating mechanism if it could be consecutively used by each input digit. His calculator might have eight sliders to enter digits, but the teeth (or steps) on a single central drum would handle the arithmetic. The drum would allow him to trim the size and weight of his machine.

By 1937 Herzstark understood how to perform arithmetic using a single rotating step drum. Everything would work for addition and multiplication. But his design stumbled over subtraction and division. He couldn't subtract by turning the crank backward, because adding two digits may create a carry condition after the operation, but subtraction requires a borrow be-

Overview/*The Curta*

- The first precise calculator you could hold in one hand was mechanical, not electronic. It could add, subtract, multiply, divide and help give square roots. Unlike analog slide rules, it solved problems to at least 11 digits.
- Its inventor, Curt Herzstark, completed the design while imprisoned in Buchenwald concentration camp during World War II. He adopted novel mathematical and mechanical techniques to save space and weight.
- Widely used by scientists, engineers, surveyors and accountants during the 1950s and 1960s, the Curta lost favor only when electronic pocket calculators hit the market in the 1970s.

forehand. A single arithmetic step drum couldn't properly look ahead to see what might be coming.

"Traveling in a train through the Black Forest, I sat alone in a compartment. Looking out the window, I thought, 'Good grief! One can get the result of a subtraction by simply adding the complement of a number.'"

To find the nines complement of a number, just subtract each digit from nine. By adding a number to the complement of another number, you can simulate subtraction. For example, to calculate 788,139 minus 4,890, first find the nines complement of 004,890: 995,109. Now add 788,139 and 995,109 to get 1,783,248. Remove the highest-order digit to arrive at 783,248. Finally, add one to find the answer: 783,249. Sweet—the same technique is used in computers today.

Herzstark's calculator would retain the single rotating step drum but would have two sets of teeth: one set dedicated to addition, the other to subtraction. Lifting the crank three millimeters would engage the subtraction teeth to perform nines-complement addition. Subtraction would be as easy as addition.

Multiplication and division would be handled by repeated addition and subtraction. And since the results register could be rotated in relation to the input sliders, several shortcuts would speed these operations. For instance, to multiply by 31,415, you don't spin the crank 30,000-plus times; the movable carriage cuts this to 14 turns: five turns for the 5, once for the 10, four times for the 400, and so on.

By late 1937 Herzstark was ready to build a handheld, four-function calculator. Then came Hitler.

CURT HERZSTARK, inventor of the Curta, became familiar with mechanical calculators as a child, because his family produced them. At age eight (*right*), in 1910, he displayed his skill at five- and six-place multiplication problems at the International Office Exhibition in Vienna, earning the label "miracle child." In 1985, at 83 (*below*), he adopted a similar pose.

COURTESY OF RICK FURR

Brought Forth in Buchenwald

IN MARCH 1938 the German army entered Austria. As the son of a Catholic mother and a Jewish father, Herzstark faced trouble. "The first weeks were dreadful. The mob came, then anti-Semites and all terrible things."

German military officers arrived to inspect the factory. To his surprise, they asked him to make precision items for the army. After a one-sided negotiation, his factory began producing gauges for Panzer tanks.

It went well for a few years. "But in 1943," Herzstark said, "two people from our factory were arrested. They had listened to English radio stations and transcribed the broadcasts on a typewriter. The typewriter was identified, and the owner was one of our mechanics. He was beheaded. The second one was imprisoned for life, which was much worse. I tried to intervene for them with the Gestapo. The officer threw me out, saying, 'What impudence, that a half-Jew dares to speak on the behalf of these people!'

"I was invited to testify for these people and arrested—nice, no? My house was searched, and, of course, I never had a trial. I was accused of supporting Jews, aggravation, and having an erotic relationship with an Aryan woman. All fabricated crimes. Later I found that a dozen others were arrested under similar circumstances."

The SS threw him into the infamous Pankratz prison, where torture of Jews was routine. "I shared a cell with 50 others, without anything at all—no beds, no lavatory, nothing. And I was even lucky, as I was sent to the Buchenwald concentration camp.

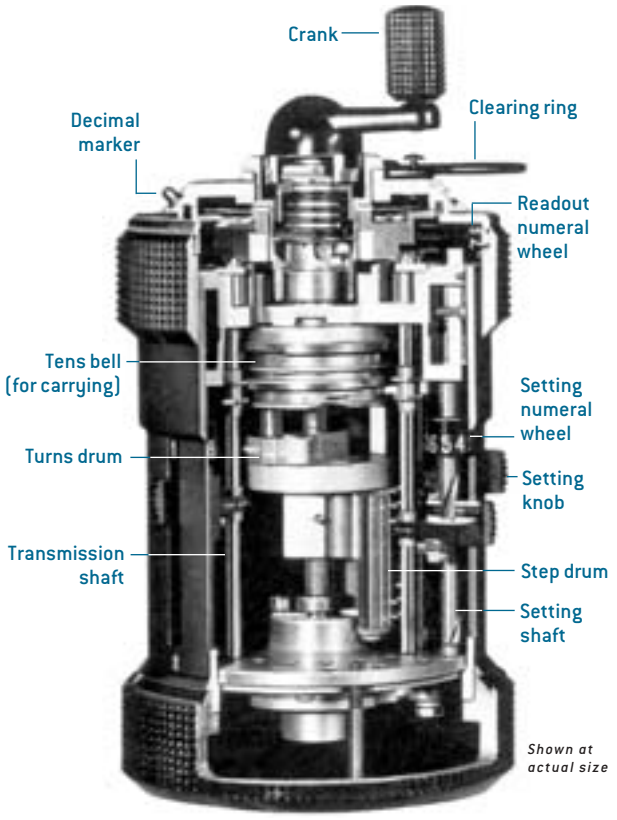
"Once there, I was put in a work unit where I believed I would be buried. It was November, and all I had was a shirt, a pair of prisoner's pants, wooden shoes and a knitted cap. I worked gardening and was completely exhausted.

"Spiritually at zero, I thought, 'I have to die.' I was called before Buchenwald's commanding SS officer. He had my life history in his hand and said, 'You have delivered gauges and instruments to the army. Listen closely. If you follow our commands obediently, you may find life bearable. I order you to work in the factory connected with the concentration camp. If you do well there, you may be able to live.'"

Alongside Buchenwald, the Nazis had built a slave-labor factory to make machinery for secret military projects. The managing engineer placed Herzstark in charge of precision parts to be shipped to Peenemünde—launch site for ballistic missiles. For the next two years, he made components for V2 rockets.

Being responsible for the section that made mechanical parts, Herzstark visited different places in the factory. At first, other

A LOOK INSIDE



AS AN OLD MANUAL SAYS, "The little Curta is a precision instrument." The step drum that performs addition and subtraction (*above* and *at right*) has 37 layers, each half a millimeter thick. A separate component—called the tens bell—participates in carrying. Each digit in the numeric readouts at the top (*see diagram at bottom left in box on opposite page*) is three millimeters tall—easily readable at arm's length. Early Curtas had eight setting knobs and 11 digits in the results register. Later versions had 11 input digits and gave 15-digit answers.

prisoners thought he was an informer. "They soon found out that I was no spy. For example, I talked to a machine operator: 'You are making this part well, my friend. You are industrious, but you've been told to do a simple process on an expensive machine. I will report that the machine is not being used efficiently, even though the prisoner is doing model work.' In this way, I became acquainted with people from Luxembourg, France, Denmark and many other places.

"Naturally, comrades came to me and said, 'Curt, you have a certain influence. Can't you bring this or that prisoner into the factory? He will die otherwise.' So I would set up an inspection station in a factory hall, seat a [captive] lawyer there and give him a micrometer."

"The SS guards checked our operations, and if there was really an inspection, there would be a sudden concert of coughing. Then the lawyer knew that danger was coming and to look industrious. But I was anxious because the comrades always wanted more from me. I knew if this came out, I'd be under the cold ground the next day. But fate helped me again.

"As the Germans retreated from Italy, they took production machines with them. One day in Buchenwald, we received two truckloads of office machines. I unloaded them, and local factory owners came to inspect them. One person kept looking at me as if he knew me. 'Herzstark?' 'Yes, Herzstark,' I answer. 'Walther,' he answers."

Fritz Walther. Herzstark's old competitor was now back to making guns. "He laid a pack of cigarettes on a lathe for me. 'Now it's all over,' I think. Accepting a gift is strictly forbidden, no? But my guard saw it and didn't want to see it. I was allowed to put the cigarettes in my pocket."

Within wartime Germany, Walther was a celebrity. He recognized the prisoner Herzstark as more important than any Italian booty and informed the concentration camp commander of his prize.

Soon after, the managing engineer took Herzstark aside and told him, "'I understand you've been working on a new thing, a small calculating machine. I'll give you a tip. We will allow you to make and draw everything. If it really functions, we will give it to the Führer as a present after we win the war. Then, surely, you will be made an Aryan.'

"'My God!' I thought to myself," Herzstark said. "'If I can make this calculator, I can extend my life.' Right there I started to draw the calculator, the way I had imagined it."

The SS didn't lighten Herzstark's workload, but he was allowed to spend his spare time on the calculator. "I worked on the invention Sunday mornings and in the evenings after lights-out. I worked in the prison, the workroom and where we ate. I drew up the machine in pencil, complete with dimensions and tolerances."

Meanwhile the Allies bombed Germany. "We'd leave the factory and go outside during lunch. Always we saw the American planes in Christmas tree formation and not one defending aircraft. Afterwards the bombs came, one saw the flashes and counted eight, nine, 10 seconds. You'd calculate the distance by multiplying by 333 meters. But one day the Christmas tree flew towards us. Now we knew this was coming to us and were terribly afraid. I ran into a small forest, hid my nose in the moss and covered my ears. It started the next moment, banging and roaring.... When I put my head up, everything was smoky and I could barely breathe.

"Several hundred prisoners were hurt that day, terrible when one sees such a thing. Of course we saw equally horrible things in the daily camp. When they hung someone, we had to watch until he finally died. Terrible. They hung people so they died slowly, a wretched death.

RICK FURR

THE CURTA IN ACTION

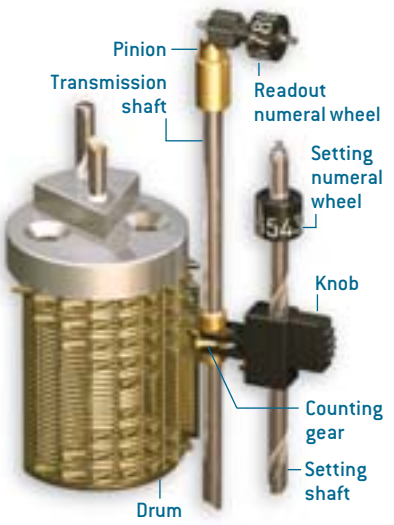
A COUPLE OF ARITHMETIC PROBLEMS show how the Curta works. To learn how to multiply, divide, and find square roots or to play with a Curta simulator, follow the links listed in the More to Explore box on page 99.

ADDITION

Add 32 + 41 + 49.

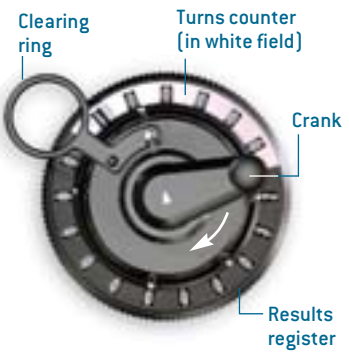
1. Spin the clearing ring once to set the results register and the turns counter to all zeros.
2. Input 32 by sliding down the two setting knobs at the far right until the numbers 3 and 2 appear in the windows under the CURTA label (*right*).

As you move a knob, you cause a grooved setting shaft (*at bottom right of diagram*) to turn an attached numbered dial—the setting numeral wheel—to show the number you select. As you slide the knob, you also drag down a counting gear encircling a second axle—the transmission shaft. In this example, the gear for the tens place gets positioned so it can engage with a layer of the step drum that has three teeth protruding from it, and the gear for the ones place gets positioned next to a drum layer sporting two teeth.



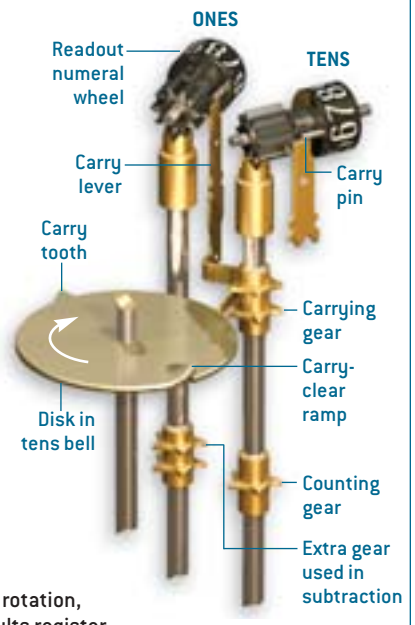
3. Turn the crank one full rotation.

This maneuver spins the step drum once, causing the teeth to turn the counting gears, which spin their transmission shafts. Pinions at the top of those axles spin readout numeral wheels, which make a 3 and a 2 appear in the results register (*right*).



4. Set 41 by using the ones and tens knobs again.
5. Turn the crank to add 41 to the results register, which then shows 73.
6. Set 49, turn the crank and view the answer (122) at the top.

This last addition requires the Curta to carry a 1 to the tens place. When the readout numeral wheel atop a transmission shaft passes 9, a carry pin extending from the wheel depresses a rod (the carry lever) that pushes down on a carry gear encircling the transmission shaft of the next higher place (here, the tens place), positioning the gear to interact with the tens bell. As the bell revolves, a tooth on a disk in the bell turns the gear, causing the readout numeral wheel on the second shaft (the tens-place axle) to advance by one digit.

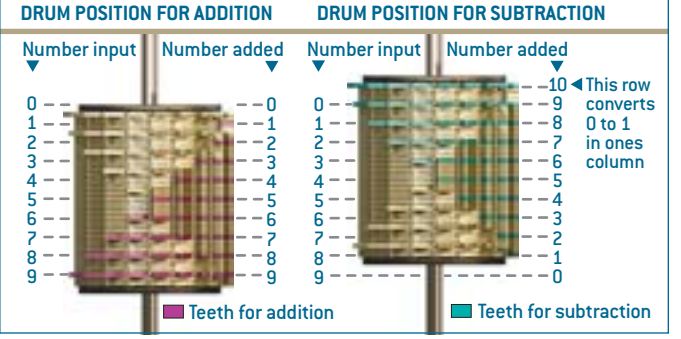


SUBTRACTION

Calculate 139 – 78.

1. Clear the machine.
2. Set 139 as you would for addition.
3. Turn the crank one full rotation, placing 139 in the results register.
4. Set the number 78.
5. Lift the crank to its upper position, which aligns the gears of the transmission shafts with the "subtraction" parts of the drum (*right side of diagram below*). Turn the crank once.
6. Read the answer in the results display: 61.

The Curta subtracts by adding complements of 9. When the crank lifts the drum, a setting of 7 aligns the counting gear on the transmission shaft with a drum layer having two teeth, an 8 aligns with one tooth, and a 0 aligns with nine teeth. A turn of the crank adds 21 preceded by a series of 9's to 139. On paper, a 1 would appear at the far left of the sum, giving a wrong answer. But the leftmost transmission shaft of the Curta has no carry lever, so the extra 1 disappears. Lifting the crank also serves another purpose. The ones-place transmission shaft has two counting gears. The extra gear has no role in straight addition, but when the crank is lifted, this gear engages the teeth in the layer above the other gear. As the drum turns, the upper layer adds 1 to the ones column, making the result here 61 instead of 60.



GRIFF WASON; SOURCES: CLIFF STOLL, RICK FURR AND WWW.VCALC.NET

"Some guards were not so bad. If an older SS was there, he often said to me, 'Ha, what's new? What kinds of machines will we look at today?' The young SS were the most dangerous. If they found an opportunity, they would be very cruel. If a prisoner annoyed them, they would shoot him, because it was necessary, wasn't it?"

Herzstark had pretty much completed his drawings on April 11, 1945, when he saw jeeps coming from the north. A soldier in the front seat called out, "You're all free." They were Americans; some were Jewish boys who'd fled before Hitler came to power. Because they could speak German, they had been assigned to the forward area.

Buchenwald was the first concentration camp freed by Western forces. Some American soldiers vomited at the sight of bodies stacked 10 deep. Looking back, Herzstark shook his head at the experience. "It was incomprehensible. If I'd been a lawyer or something, I would have died miserably. They would have sent me to a quarry, and in two days I would have a lung infection and it's all over. A thousand died like this. God and my profession helped me."

PAGE BACK 34 years and seven months to see the Curta advertised in the June 1969 issue of *Scientific American.*

The Crown Prince Calculates

A FEW DAYS AFTER the Americans liberated Buchenwald, Herzstark walked to the city of Weimar with his plans folded in his pocket. He brought his drawings to one of the few factories still standing, where machinists examined them. He remembered the technicians' response: "It was like scales falling from their eyes. The solution was clear, and there was nothing more to think about." Though penciled in the concentration camp, Herzstark's designs were so clear that it took only two months to make three prototype calculators.

But just as contracts were being drawn up, the Russian army arrived. Herzstark knew the score: he grabbed the prototypes and headed for Vienna, taking the machines apart and putting the pieces in a box. "If someone had looked in, it was like a toy," he said. "The whole thing was disassembled."

He traveled to Austria by walking, sleeping on floors, and bartering cigarettes for a train ride. His family's old factory was unusable. With nothing but his three models, Herzstark filed for patents and tried to get someone to invest in his idea. Remington-Rand, the American office machine firm, displayed some interest but never called back. The government of Austria turned him down. Europe was in cinders, without the infrastructure to start new projects.

Yet the prince of Liechtenstein had been thinking about developing industry in his country. At the time, Liechtenstein was almost entirely agricultural; its major industry was the manufacture of false teeth. Invited to the court, Herzstark showed his models to royalty, ministers and patent specialists. "In his castle the prince himself calculated with it. Family members watched as well as professionals. The prince was immediately enthusiastic and said this project was the right one for the country. He received me charmingly, and we had a four-hour conversation."

All went well at first. Liechtenstein created a company, Contina, and then floated loans and issued stock. Herzstark served as technical director, received a third of the stock and was to receive a royalty for each machine sold.

Herzstark advertised in the Swiss newspapers for mechanics willing to begin a new career. Contina rented a hotel ballroom where Herzstark's machinists built the first 500 Curta calculators. They went on sale in 1948 and were promoted at trade shows and in technical magazines. Six months later an American department store tried to order 10,000, with an option for more. Instead of latching onto this order, the finance director decided it was beyond the company's capability, dooming the Curta to mail-order sales and an occasional specialty store.

The demand was there, however, and Contina expanded from the ballroom to a proper factory, ramping up production to several hundred per month. With this progress, the financiers behind the company pulled the rug out from under Herzstark—reorganizing the firm and annulling his stock. Like Edison, Tesla and so many other inventors, Herzstark would be squeezed out of the profits from his own creation.

"Then came a stroke of good luck that I could not have imagined," Herzstark said. "The patents were still in my name."

THE AUTHOR

CLIFF STOLL is best known for breaking up a ring of hackers during the early days of the Internet, as described in his best-selling book *The Cuckoo's Egg*. His other books include *High Tech Heretic: Why Computers Don't Belong in the Classroom* and *Silicon Snake Oil*. Although his Ph.D. is in planetary science, he now rebuilds mechanical calculators, makes Klein bottles and occasionally teaches physics. He lives in Oakland, Calif., with his wife, two children and a pair of cats he pretends to dislike. Stoll thanks Rick Furr, Jan Meyer, Jack Christensen and Chris Hamann, experts on the Curta, for their generous assistance, and the Charles Babbage Institute at the University of Minnesota for making available computer historian Erwin Tomash's 1987 interview of Curt Herzstark. The quotations in this article come from that transcript.

Early on the trustees hadn't wanted to take over the patents in case of litigation—they wanted Herzstark to take the heat if someone challenged his invention. Because the company had never acquired the patent rights, Herzstark forced them to come to terms. Throughout the 1950s and 1960s, he actually made money from his invention.

After the success of the first calculator, Herzstark designed a slightly bigger model, increasing its capacity from 11 to 15 digits. But thereafter the only thing that changed significantly was the shape of the carrying case. Setting a rare record in the computing industry, Herzstark got the design right the first time.

For two decades, the Curta calculator sold steadily, touted as "the Miniature Universal Pocket Size Calculating Machine with reliability derived from rational, robust construction." As Herzstark predicted, engineers used the miracle machines to find satellite orbits, surveyors to keep track of transit positions, and traveling accountants to balance books. One New York bank manager was amazed when an auditor appeared without a briefcase-size calculator yet tallied the books down to the penny.

Curiously, sports car enthusiasts around the world adopted Curta calculators, reckoning speeds and distances in rallies. Toggling the numbers by feel, navigators would quickly calculate their ideal driving times without taking their eyes from the road. The Curta's small size fit the confined quarters of a sports car, and—unlike early electronic calculators—the device was unfazed by bumps, vibrations or voltage spikes. Even now, vintage car rallyists enjoy the challenge of mechanically calculating their travel times.

Just as battery-operated quartz watches pushed aside windup watches, electronic calculators eclipsed Herzstark's invention. After a run of 150,000, the last Curta calculator was sold in the early 1970s. Hasn't been a mechanical calculator made since.

Herzstark left Contina in the early 1950s, afterward consulting for Italian and German office machine makers and living in a modest apartment in Liechtenstein; back then, technowizards didn't buy million-dollar spreads. The government of Liechtenstein recognized his accomplishments only after he turned 84; he died not long after, in 1988.

Still Working after All These Years

YOUR ELECTRONIC POCKET calculator will solve problems faster than the Curta. And a desktop computer does wizardry compared to it. Perhaps the Curta's only use is to balance your checkbook during a blackout.

Yet as I hold Herzstark's lilliputian calculator, passed down from my first astronomy professor, I'm acutely aware that this machine has outlived its first owner and doubtless will live beyond its second. As the instruction manual says, "Your Curta will last you a lifetime, and remain an indispensable aid always ready to hand. You can be entirely confident of its precision; the little Curta is born of long experience in the field of calculating machines. It is manufactured… by international specialists in fine mechanics, with superior quality metals. No artificial materials whatsoever are used in its construction." (I can't imagine reading, "Your Excel program uses no artificial materials" or "Your Pentium microprocessor will last a lifetime," although both statements are true.)

I figure that I don't own something until I understand it, and I can't understand it until I see how it works. So armed with a magnifying glass, tweezers and jeweler's screwdrivers, I unscrew the barrel to uncover 600 parts: gears, shafts, pawls and pinions.

I delicately remove eight setting shafts, each machined with a spiral groove and designed without collaborators, assistants or even drafting tools. I see the ingenious step drum mechanism, first penciled under impossibly wretched circumstances. I touch lightweight alloys, revolutionary in their day. I feel a tactile finesse, which transcends half a century of calculating progress. Entirely confident in the Curta's precision? Absolutely.

> "'My God!' I thought… 'If I can make this calculator, I can extend my life.' Right there I started to draw…."

On my now reassembled 50-year-old calculator, I divide 355 by 113. With my thumb, I slide the setting knobs on spiral axles, then turn the crank to stash the first number in the machine. I enter the second number, lift the handle and rotate it again. The Curta's counting gears engage the nines-complement cogs of the step drum. Steel transmission shafts translate this motion through right-angle pinions and then into the results register. As I turn the handle, control, logic and digits rotate around the crankshaft. Two dozen spins later my answer snaps into tiny windows.

Before me is an approximation to pi and more. At once I'm holding the lineal descendant of the first calculating machines, the acme of Western mechanical craftsmanship and a monument to one man's vision overcoming a wall of hostility. SA

MORE TO EXPLORE

Antique Office Machines: 600 Years of Calculating Devices. Thomas A. Russo. Schiffer Publications, 2001.

The Universal History of Computing: From the Abacus to the Quantum Computer. Georges Ifrah. John Wiley, 2001.

Information on the Curta, including a manual with operating instructions, can be found at Rick Furr's Web site: **www.vcalc.net**

Try Jan Meyer's Curta simulator at **www.vcalc.net/curta_simulator_en.htm**

Greg A. Saville will show you the details of a Curta's disassembly at **http://home.teleport.com/~gregsa/curta/disas**

Find calculating algorithms at **www.curta.org**

For road rallies where you could use a Curta, visit **www.VintageRally.com** or **www.scca.com**

VIRTUAL 1ST DOWN MARKER

Phantom Gain

Football fans just call it "the yellow line." It appears out of nowhere on the TV screen, marking how far up the field the offensive team must advance to gain a first down. The stripe looks as if it is painted on the turf: when the camera angle changes, it stays in proper skew; as the camera zooms in and out, it widens and narrows; and when players run by, it disappears behind them. But once the play ends, the streak vanishes, only to rematerialize when the team lines up for the next play.

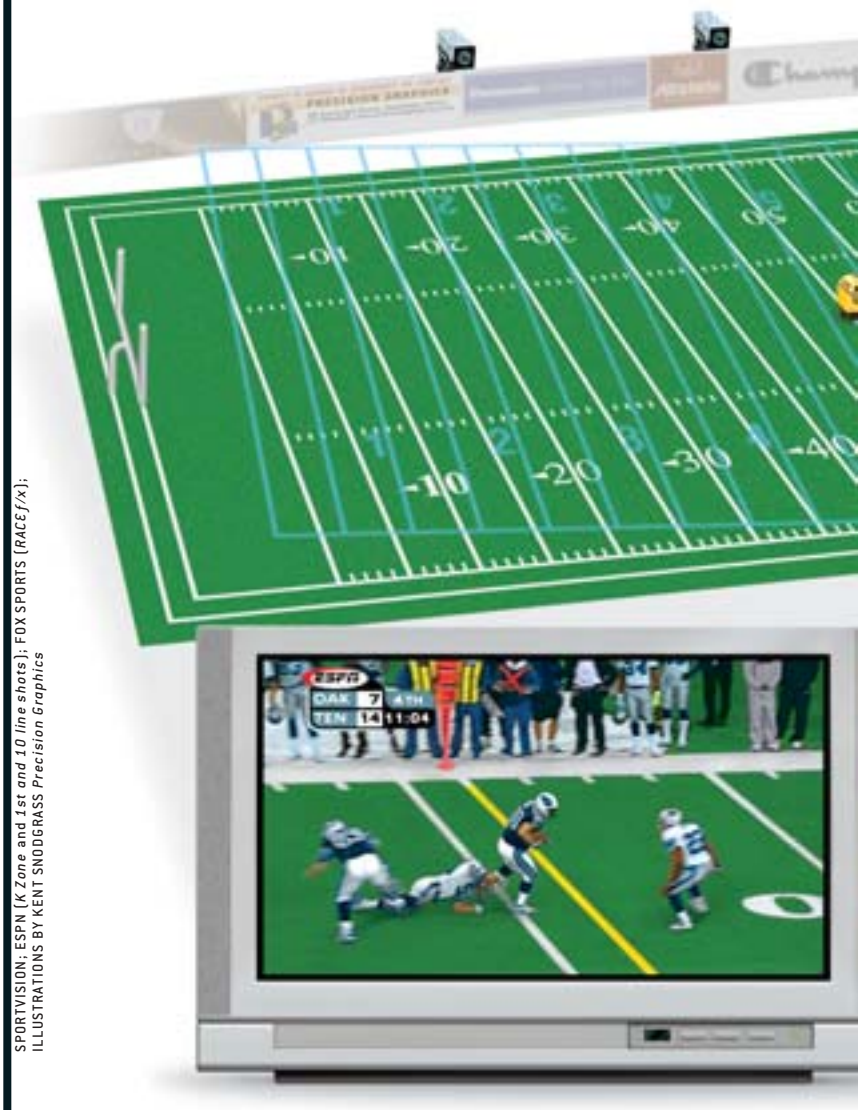
Sportvision in Chicago produces the "1st and 10" line for the ESPN, ABC, Fox and Turner networks, and Princeton Video Image in Lawrenceville, N.J., does it for CBS. An operator arrives at the host stadium with the network's video crew one or two days before game time and proceeds to digitally map the gridiron, calibrate the three main game cameras at the 25, 50 and 25 yard lines, and customize the videographics software so that the golden apparition will appear on every play from scrimmage.

"When we began in 1998 we needed four people, five racks of computers and our own video-production van to produce the line," says Marv White, chief technical officer for Sportvision. "Now it takes only a half-rack of computers and one operator who works from the network's regular broadcast truck."

The technician keys the line to the physical first-down marker that game officials stake at the sideline for each series of downs. But actually drawing the stripe—a winner of multiple Emmy Awards for technical achievement—is a high-tech trick somewhat akin to the "blue screen" used to superimpose computer-generated maps behind television weather forecasters. To make the yellow streak look as if it is on the field, the operator continually refines the colors in the broadcast feed over which the computer should paint the line (grassy greens) and those it should not (players' skin and uniforms).

"This is the hardest part, operationally," White says, noting that lighting and turf conditions steadily change during the course of a contest. "We use a lot of technology, but it's also an art. The operator is pretty busy all game." —*Mark Fischetti*

PREGAME: A laser surveyor (*center*) maps the field, and a computer in the stadium's video-processing truck transforms the data into a digital grid (*blue*), which is sent to the three primary cameras used during a broadcast. The computer adjusts the grid to match the actual side and yard lines as seen through each lens, then calculates the distance from each camera's optical center to reference points on the blue field.



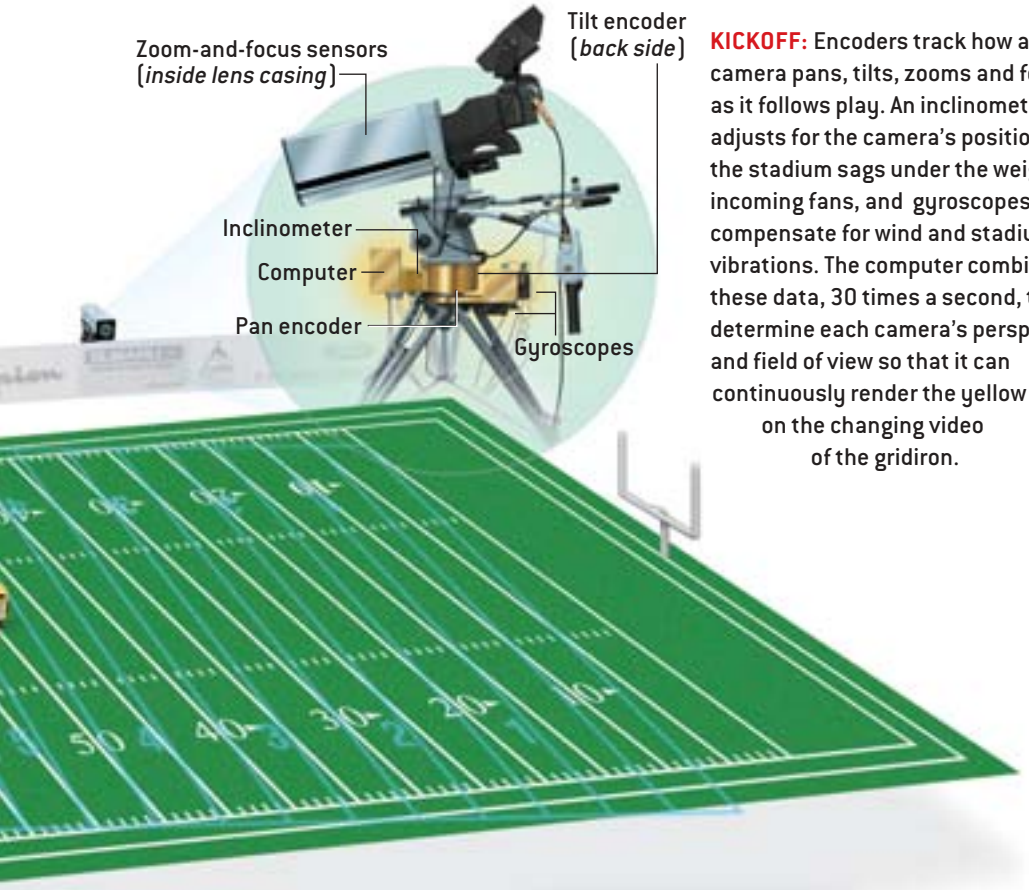
SPORTVISION; ESPN (*K Zone* and *1st and 10 line shots*); FOX SPORTS (*RACE f/x*); ILLUSTRATIONS BY KENT SNODGRASS *Precision Graphics*

DID YOU KNOW . . .

➤ STRIKE THREE! Broadcasters have enhanced baseball games with an adaptation of the yellow first-down line system. K Zone from Sportvision, used by ESPN, replays a pitch as seen from just behind the pitcher (*below, right*). It outlines the strike zone above home plate and freezes the ball as it passes the batter, so viewers can see if the pitch was a strike or not. Three stadium cameras use optical tracking to follow the ball and compute its trajectory. Fans love the view; umpires might not be thrilled if it becomes used for critiques. By the way, the changing billboards that appear on the wall behind home plate—and the transient corporate logos that seem to be painted on football fields—are created by both Sportvision and Princeton Video Image with versions of their respective yellow-line technology.

➤ FLOATING DASHBOARD: How fast is car 18 racing around the speedway? This and other stats pop up in virtual bubbles that follow cars around tracks at NASCAR Winston Cup events televised by Fox and NBC. Sportvision's RACEf/x system (*below, right*) relies on the Global Positioning System and inertial measurement units in the cars.

➤ ALMOST LIVE: A "live" broadcast reaches your television as much as 1.5 seconds later than real time. Satellite transmission delays are a quarter second for each up and down link, and video compression known as MPEG can take even longer. Processing football's 1st and 10 line adds another 100 milliseconds. Want a test? Bring a TV to the stadium; you'll see a punt sooner than it appears on your screen.



KICKOFF: Encoders track how a camera pans, tilts, zooms and focuses as it follows play. An inclinometer adjusts for the camera's position as the stadium sags under the weight of incoming fans, and gyroscopes compensate for wind and stadium vibrations. The computer combines these data, 30 times a second, to determine each camera's perspective and field of view so that it can continuously render the yellow stripe on the changing video of the gridiron.

K Zone

RACEf/x

FINAL IMAGE: The virtual line is superimposed to the nearest tenth of a yard. The yellow tone can be adjusted to look like chalk on grass or paint on artificial turf.

DURING PLAY: To appear real, the stripe must disappear behind players who cross it. An operator in the video broadcast truck chooses pixel colors the line should and should not paint over—covering the greens in grass but not flesh tones or colors in uniforms. Because sun, clouds, stadium light and field muddiness change throughout the game, the operator alters the palette minute by minute—in this case, selecting blues to complement yellows to better cover off-green patches of turf in bright sunlight.

Send topic ideas to workingknowledge@sciam.com

A Great Echelon of Birds

HALF A MILLION SANDHILL CRANES STOP ALONG A STRETCH OF NEBRASKA'S PLATTE RIVER EVERY SPRING

BY MARGUERITE HOLLOWAY

They first appear as phrases written in the sky. A line of cursive coalesces in the air, then fragments. Hundreds of shifting lines with words moving between them; language forming, breaking up, reconfiguring. That is how it appears to me. To my friend, Patricia Wynne, an artist, the blue sky seems filled with black lace that is coming unraveled, being rewoven. We pull our car to one side of the road and stand next to a cornfield to watch a lattice weave itself above us and to hear the air fill with the twanging, plucked-rubber-bandlike calls of thousands of migrating sandhill cranes.

These tall red-crowned birds arrive every spring and pack into fields along a 60-mile section of the Platte River in southeastern Nebraska. For several weeks, the Platte is a crucial stopover along a migration path called the Central Flyway for about 500,000 sandhill cranes, the largest grouping of cranes in the world, and for roughly 10 million other birds—northern pintails, snow geese, greater white-fronted geese and, we soon learn, an occasional surprise. As we stand mesmerized, watching wave upon wave of birds come out of the south and circle, Pat glimpses a flash of white. We chortle to ourselves, thrilled to see one of the rare whooping cranes that sometimes fly with the sandhills. A farmer passes without a second glance; he had seen hundreds like us.

Pat and I had arrived in Omaha the night before so we could roust ourselves in the dark morning and drive two hours to Grand Island to see the birds at early light. We are part of the annual migration of birders—just two of the thousands who come to this region of Nebraska every year to view the profusion of cranes. That first day we see cranes in all configurations: flying in, feeding, and roosting in the river. In the fields alongside virtually every road, they eat corn left after the harvest. Some of the cranes travel 12,000 or so miles each year, coming from Texas, New Mexico and Mexico, stopping along the Platte to eat enough for the journey to nesting grounds in Canada, Alaska and Siberia. Three of the six subspecies stop here: the Canadian, greater and lesser—and it is the lesser that comes in greatest number, about 330,000 of them.

We watch them perform what orni-

MIDDLE OF MARCH is thick with sandhill cranes in Nebraska's Rainwater Basin region. About 80 percent of the world's sandhills stop here to feed.

thologists call dancing—wings outstretched, springing up and backward, legs straight—in the fields. Standing four feet tall, with burgundy skin on their heads (if they are mature) and a six-foot wing span, the sandhills seem as though they have arisen from the fallow ground beneath them: gray, rust, brown. As she sketches them, Pat says they look primal, like small bipedal dinosaurs, which, in some distant way, they are.

The most dramatic scenes take place not in the fields but in the Platte at dawn and dusk, and so in the chilly late afternoon we settle on a bridge in the Crane Meadows Nature Center (www.cranemeadows.org) with guide Blake Hatfield. From every direction, clusters of cranes arrive and land along the banks. Then, one or two at a time, they fly into the shallow water for the night, where they can hear the approach of predators such as coyotes and bobcats. The Platte used to provide many hundreds of miles for feeding and refuge, but in the past century, 70 percent or more of the river's flow has been diverted for agriculture. The Platte is no longer "a mile wide and an inch deep," as some 19th-century pioneers described it, and about 80 percent of the original associated wetlands have disappeared. As the waters have waned, so have the birds and cranes—particularly the whooping crane, of which there remain fewer than 200 in the U.S.

Within 15 minutes, 300 sandhills are standing in the water about a quarter of a mile from the bridge. And that stretch of river soon would have filled in completely, but for a bald eagle. One sight of the raptor alighting in a nearby willow, and the entire flock of cranes cries out and rises and wheels into the fields. It is too dark to see the river when they finally return.

Early the next day we attach ourselves for a few hours to a five-day excursion that ornithologist Kim Eckert is leading for Victor Emmanuel Nature Tours (www.ventbird.com/). Eckert graciously lets us tail the group's two minivans as they head north to see prairie-chickens doing their morning courtship displays and, as it turns out, one sharp-tailed grouse that is living with them, most likely because there isn't another one of these threatened birds in the vicinity. We watch and listen to morning larks and a horned lark as we try to catch sight of the far-off chickens. Eckert teaches us everything we now know about sandhills and then something we didn't want to know: we had not seen a whooping crane but rather an albino sandhill—a common mistake, he explains.

White or gray, red-capped or juvenile, dancing or feeding, soaring or wading, these amazing birds can be seen from many vantage points between late February and early April: from blinds, from cars or vans, from a warm viewing center, on foot, from bridges. And there are many sites from which to watch: centers, sanctuaries and fields between Kearney and Grand Island, the stretch of river where the birds are most concentrated. And innumerable ways to try to capture what they look like: In *A Sand County Almanac,* Aldo Leopold describes them as "a great echelon of birds." But, he continues, "our ability to perceive quality in nature begins, as in art, with the pretty. It expands through successive stages of the beautiful to values as yet uncaptured by language. The quality of cranes lies, I think, in this higher gamut as yet beyond the reach of words."

SUNSET brings the cranes into shallow water for the night, out of the reach of at least some predators.

For more information about the migration, browse these Web sites: www.ngpc.state.ne.us/wildlife/migrate/intro.htm, http://platteriver.unk.edu and www.rowesanctuary.org. In addition to Victor Emmanuel Nature Tours, multiday trips are often organized by Wings (http://wingsbirds.com/) and Field Guides (www.fieldguides.com/). For general information on cranes, see the International Crane Foundation's site (www.savingcranes.org). Among the many helpful books are *The Birds of Heaven,* by Peter Matthiessen; *Crane Music,* by Paul A. Johnsgard; *The Cry of the Sandhill Crane,* by Steve Grooms; and *Birding Crane River,* by Gary R. Lingle. SA

ILLUSTRATION BY PATRICIA J. WYNNE; LAYNE KENNEDY *Corbis (top left* and *top right)*; KEVIN SCHAFER *Corbis (bottom left)*; JAMES L. AMOS *Corbis (bottom right)*

Metaphorical Suns . . .

. . . BUT VERY REAL DESTRUCTION **BY PHILIP MORRISON**

100 SUNS
by Michael Light
Alfred A. Knopf, 2003 ($45)

Text-free, portrait-large photographs—many in dramatic full color, mainly crimson and black by land, clouded skies by sea—are the hundred metaphorical suns promised. Rather more than half of them disclose the proverbial mushroom cloud, luminous or vapor-borne. Each one is a prompt, distant shot of an American nuclear weapon explosion, made during the years from 1945 to 1962, until the Limited Test Ban Treaty quelled both public witness and most fallout through burial underground.

The meticulous compiler—photographer Michael Light, whose book *Full Moon* drew wide praise—ordered his portraits here for visual effect. A contextual look discloses much of weapon development amid the politics of unbridled state power. Since 1945, with the first test and the two calamitous attacks on Japanese cities, the explosive energy ranged from Little Feller I, a test of a midget atomic rocket suited for one-man launch, up to H-bomb Mike, shown in five striking views from 1952. Mike, the first large American thermonuclear device, raised the ante as measured in tons of TNT, from a 10-ton truckload to a fanciful TNT-laden boxcar train 2,000 miles long, rattling past at full speed during two nights and one day.

Numbers do not convey everything. The image that most compels a viewer is one from 1946 itself, the first postwar year. The U.S. Navy felt the need for a demonstration of the new atomic threat against warships (no H-bombs as yet). The Bikini Atoll test was duly prepared in the summer of 1946. One fast daylight snapshot from the air shows something near human scale. Against the huge foamy tower of seawater thrown upward, a few tiny black splinters are dwarfed. The furious waters reached and ruined them. Are they kayaks? They were in fact among the largest battleships ever sent to sea, Japan's naval pride, anchored empty as targets.

H-bomb tests are observed from 50 miles off; their images here are mostly colorful and complex layers of cloud formations out to the horizon. A few plates show witnesses, some of them troops set closer to the fireball than we would so casually plan today.

The documentation is admirable. And Michael Light has put his own views briefly but clearly at the end of the book, recognizing that photographs tell only how things look: "When it's all we have, however, it's enough to help understanding. It exists. It happened. It is happening. May no further nuclear detonation photographs be made, ever." SA

Philip Morrison, emeritus professor of physics at M.I.T., wrote the book review column for this magazine for more than 30 years. He was a member of the Manhattan Project and a witness of the first test.

WATCHING the test of an 81-kiloton bomb on Enewetak Atoll, 1951

BAKER, 21 kilotons, Bikini Atoll, 1946

Verifying Your Circuits

BY DENNIS E. SHASHA

You have just received a large collection of digital circuits from a not too trustworthy supplier. You know which wires are connected to which circuit elements, and you are told what the elements are supposed to be. The question is: Have you been told the truth? You want to use as few tests as possible to determine if the supplier has actually put in the correct elements. These circuits use just two possible elements: AND and OR logic gates. Each can be characterized by a truth table relating its two inputs to its output [*see tables below*]. The output of an AND gate is 1 only when both inputs are 1, and the output of an OR gate is 1 when either input is 1.

As a warm-up, assume that the circuit is in the three-element configuration shown at the bottom left. You suspect that the OR gate (element 3) may actually be an AND gate and that one or both of the AND gates (elements 1 and 2) may actually be OR gates. You can test the circuit by putting a binary value—either 1 or 0—in each of the inputs (A, B, C and D) and seeing which values appear in the outputs (E and F). Only one test is needed to determine whether the logic gates in this circuit are correctly labeled. But what inputs should you use for the test, and what outputs should you expect?

Here's the answer: Put 1 in input A, 0 in B, 1 in C, and 1 in D. If element 2 is truly an AND gate, its output will be 0. If element 1 is truly an AND gate, output E will be 0, and if element 3 is truly an OR gate, output F will be 1. If any of the elements is mislabeled, outputs E and F will not be 0 and 1, respectively.

Now look at the four-element circuit shown at the bottom right. What is the smallest number of tests needed to verify the circuit, and what inputs would each test entail? As a second challenge, answer the same question for a circuit in an identical configuration but containing four AND gates. Finally, go back to the first illustration and consider all the possible combinations of AND and OR gates that could be arranged in this three-element configuration. Which combinations can be verified using just one test? SA

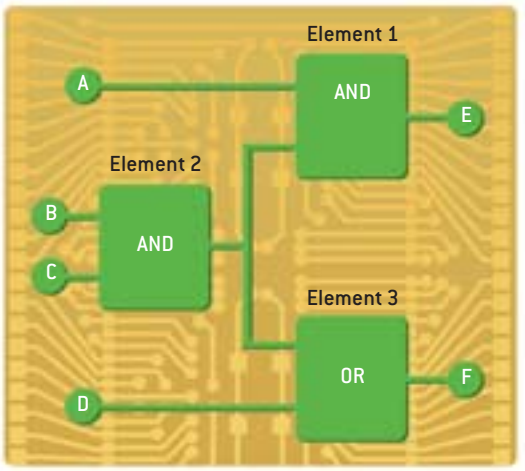
Truth table for AND gate

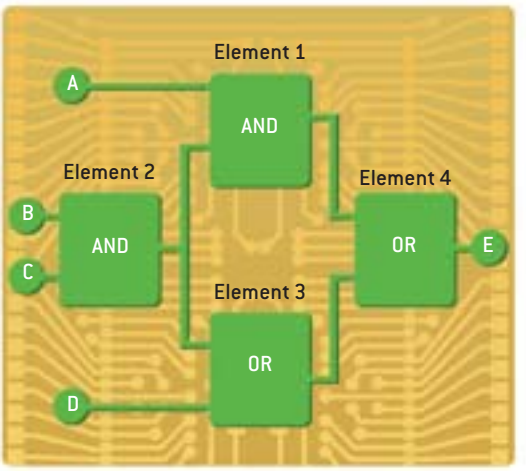
Input	Input	Output
0	0	0
0	1	0
1	0	0
1	1	1

Truth table for OR gate

Input	Input	Output
0	0	0
0	1	1
1	0	1
1	1	1

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





Answer to Last Month's Puzzle

The longest 2-surprising sequence composed from the first five letters of the alphabet is 12 letters long. An example is AABCDECBDBAE. A long 2-surprising sequence from the first 10 letters of the alphabet is BDGCJHHFACAEJGFIEDIBHCGBFJ. A long 2-surprising sequence from all 26 letters of the alphabet is HLYOBAZQCRLDSGMVKDWXTEEPXIKNUJBURFTGWIPVJIMSMKCVYHDOQRNGZLEXAWHZBY. A long 3-surprising sequence built from five letters is BCEDACDADAECCBBDEA.

Web Solution

For more discussion of last month's problem and a peek at this month's solution, visit www.sciam.com

MEGHAN GERETY

Check Those Figures

HEY, IT'S PLAYTIME! ICONS—YOU CAN, TOO! BY STEVE MIRSKY

Superman could fly faster than a speeding bullet, presumably a lot faster. But even more powerful than the caped Kryptonian was Albert Einstein, who limited Superman's flight speed to that of light. So it makes sense that Einstein, too, is now available as an action figure—think G.I. Joe with a rumpled sweater, using his kung fu grip to smash paradigms with a single equation. Or, as the catalogue sales copy has it, "Dressed for intense classroom action, this 5″ tall, hard-plastic Einstein Action Figure stands with a piece of chalk in his hand, poised to explain relativity.... Features realistic disheveled hair."

The Einstein Action Figure (it really should be a No-Action-at-a-Distance Figure, the editor in chief of this magazine helpfully pointed out) is brought to you by Archie McPhee & Co. in Seattle, which bills itself as "Outfitters of Popular Culture. Since 1980." The McPhee action hero line includes Sigmund Freud as well as Ben Franklin, a first-rate scientist when he wasn't revolting. But no other scientists are represented. I therefore propose the creation of the following additional science action figures to round out the collection:

Barbara McClintock: "You'll jump genes with reckless abandon as McClintock plays her corny joke on the scientific establishment. Includes Nobel Prize, which arrives much, much later."

Isaac Newton: "Pelt your genius action figure with apples until he figures out the gravity of the situation. Equal-and-opposite reaction figure sold separately."

Archimedes: "'Hey, Mom, I found him!' Whether taking a bath, looking for that elusive lever long enough to move the earth or just plain screwing around, Archy metes out good times. Keep the Syracuse smarty away from Roman soldiers, and he might come up with the calculus two millennia early!"

Carl Djerassi: "He put the 'action' in action figure when he invented the birth-control pill. Now you can put him on your shelf—but don't forget to take him down the morning after!"

Edward Teller: "Ed goes fission—and fusion!—through political minefields for five decades pitching nukes. From mass destruction to nudging away incoming comets and asteroids, Edward tells ya, 'No nukes is bad nukes.' "

Lorenzo Romano Amedeo Carlo Avogadro: "What a mouthful! But you've got his number now—you'll have 6.02×10^{23} hours of fun with the man who figured out that equal volumes of all gases at the same temperature and pressure contain the same number of molecules. No, really, you will."

The Big Box o' Steves: "What's more fun than a scientist-writer? A scientist-writer named *Steve!* You'll get Steven Weinberg, Stephen Jay Gould, Stephen Wolfram, Steven Pinker and Stephen Hawking. Hi-ho, Steverino! (Keep Pinker and Gould figures separate to avoid spontaneous combustion.)"

Antoine Lavoisier: "Out with the bad phlogiston, in with the good oxygen. You'll lose your head (he did, in that scary Reign of Terror!) over the man behind the law of conservation of mass. Thirteen-year-old wife not sold in the U.S."

Franz de Waal: "I see primates, I see Franz! Stare at de Waal as he observes chimps and picks up clues about why we humans do the wacky things we do. You'll have a Goodall time!"

Werner Heisenberg: "Uncertain where you left your Heisenberg action figure? So is he! Check the desk drawers. But don't be surprised if you find Schrödinger's cat. Which may be dead. Or alive!"

Charles Darwin: "This action figure loves coming along for boat rides and bird-watching but can get down at the Down house in Downe for 40 years at a time. Comes complete with his faithful bulldog, Huxley. Warning: some school boards may try to stick pins in him, but he'll survive those pricks."

I was going to suggest B. F. Skinner, but what good is an action figure you don't take out of the box? SA

ILLUSTRATION BY MATT COLLINS; PHOTOGRAPH BY FRANK VERONSKY

How does spending prolonged time in microgravity affect astronauts?

—A. Kokacy, Newton, Tex.

Jeffrey Sutton, director of the National Space Biomedical Research Institute, and Nitza Cintrón, chief of NASA's Space Medicine and Health Care Systems Office, explain:

Space affects the body in many ways. A partial list of the consequences of long stays in microgravity (where the pull of Earth's gravity is virtually unnoticeable to humans) includes bone loss at a rate of 1 to 1.5 percent a month, producing changes similar to osteoporosis; an increased risk of kidney stones and bone fractures, both associated with bone demineralization; and loss of muscle mass, strength and endurance, especially in the lower extremities. Other changes are diminished cardiac function and the possible occurrence of heart rhythm disturbances, redistribution of body fluids away from the extremities and toward the head, and alterations in the neurovestibular system that often lead to disorientation and decreased neuromuscular coordination on return from prolonged missions. Disruptions of circadian rhythms because the 24-hour day-night cycle is absent result in sleep loss and stress, and the body experiences reduced blood volume, immunodeficiency and transient postflight decreases in levels of red blood cells, despite adequate nutritional intake.

Space also presents health risks in the form of radiation, normally blocked by Earth's atmosphere. The space environment contains galactic cosmic rays, heavy ions such as iron, trapped electrons and protons, and neutrons. Such radiation can induce cataracts and cancer and adversely affect physiological processes.

To counter these dangers, mission planners have developed a variety of strategies. During prolonged missions, exercise is employed to minimize large-muscle atrophy. Certain tasks, such as extravehicular activities (spacewalks), are not performed routinely until bodily fluid redistribution stabilizes and astronauts have had an opportunity to acclimatize to space for several days. Medications have proved effective in treating motion sickness and orthostatic hypotension (low blood pressure when standing), and some drugs are potentially useful in reducing bone loss. Different lighting intensities and wavelengths are also being studied and implemented as a way to maintain astronauts' normal circadian cycle. To protect against space radiation, special shielding is installed on spacecraft.

How do geckos' feet unstick from a surface?

—S. Beres, Trumbull, Conn.

Kellar Autumn of Lewis & Clark College studies gecko adhesion and provides the following discussion:

The adhesive on the gecko's toes is quite different from a conventional tape. Instead of tacky polymers, geckos have arrays of millions of microscopic hairs—setae—on the bottom of their feet. Each seta ends in a smaller array of nanostructures, called spatulae, permitting intimate contact with surfaces.

Last year research colleagues and I discovered that setae adhere by weak intermolecular van der Waals forces—a function of the geometry of these nanostructures rather than their surface chemistry. This finding suggested that splitting any surface into small protrusions can make the surface sticky. With Ronald S. Fearing and Robert J. Full, both at the University of California at Berkeley, I have used this principle to make the first synthetic versions of the gecko adhesive.

Control of attachment and detachment in geckos is also a function of geometry, not chemistry. All 6.5 million setae on a gecko attached at once could lift 133 kilograms. This impressive gripping power raises the question of how geckos detach their feet in just 15 milliseconds. We learned that simply increasing the angle of the seta to 30 degrees causes detachment. Setae detach easily because the setal shafts act as levers to peel the spatulae away from the wall. The gecko's unusual toe-peeling behavior may also aid in reducing detachment forces by removing only a small number of setae at a time. SA

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

MATT COLLINS

FUZZY LOGIC