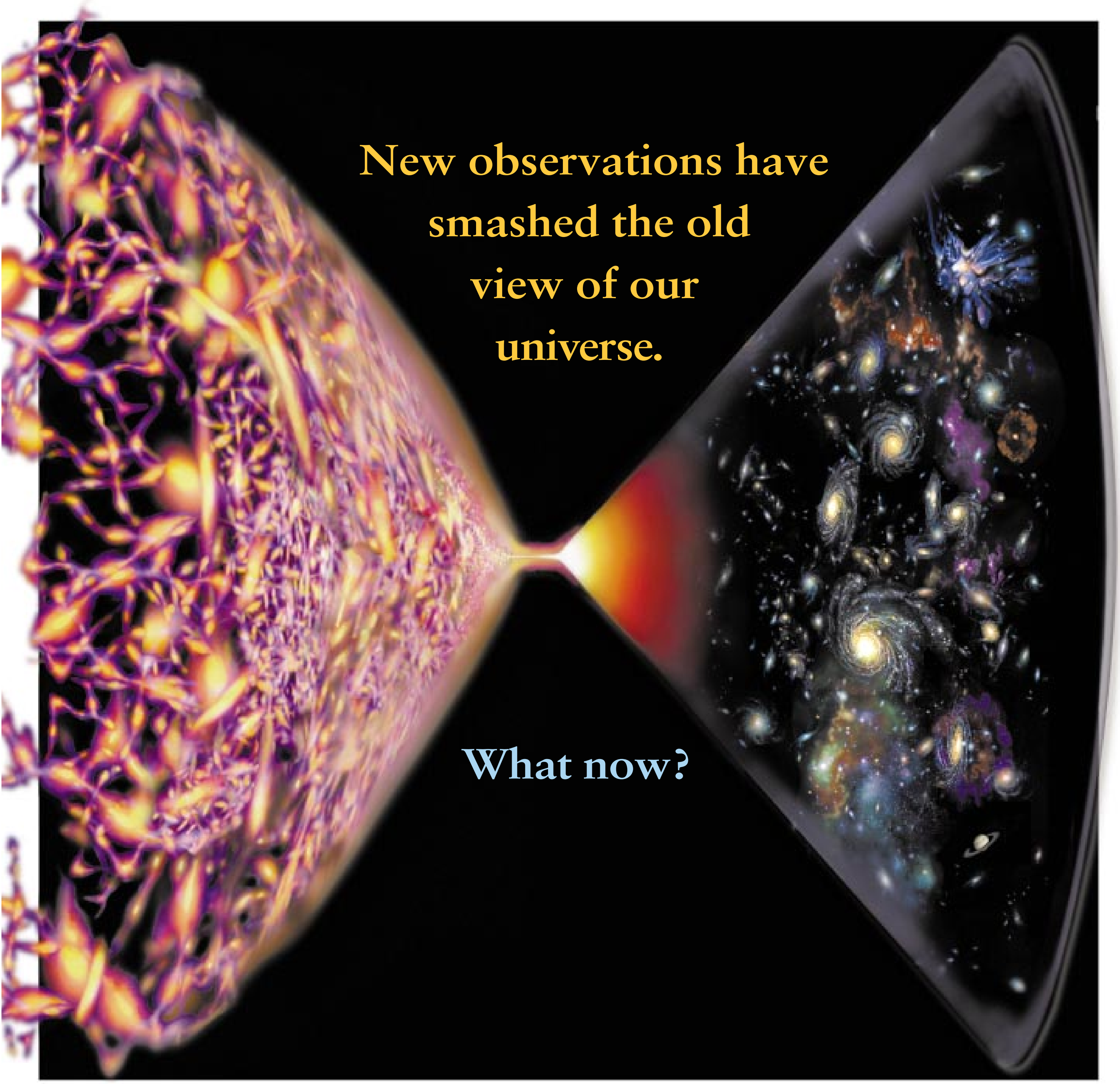
THE FLU: NEW DRUGS BEAT KILLER VIRUSES • Y2K BUG: HOW TO FIX IT, WHAT TO EXPECT

SCIENTIFIC AMERICAN

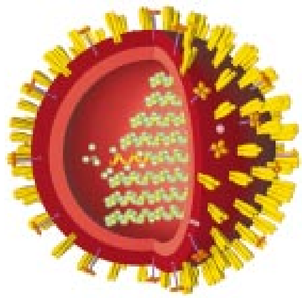
JANUARY 1999 $4.95

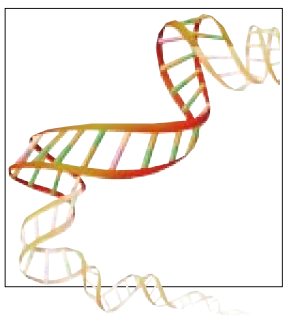
SPECIAL REPORT: Revolution in Cosmology

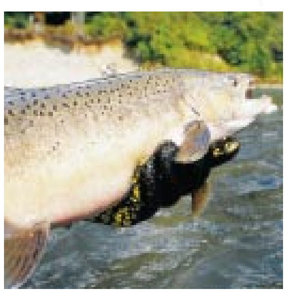


SCIENTIFIC AMERICAN

January 1999 Volume 280 Number 1





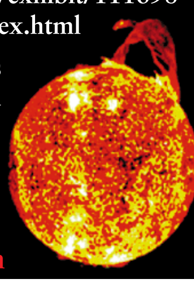


About the Cover
Quantum particles fluctuating in and out of existence might, on a cosmic scale, counterbalance gravity's tug on ordinary matter and push the universe outward. Painting by Don Dixon.

THE SCIENTIFIC AMERICAN WEB SITE

Watch a solar flare re-created in the laboratory:
www.sciam.com/exhibit/111698sun/index.html

Then browse this month's articles and departments linked to other science resources on the World Wide Web.



www.sciam.com

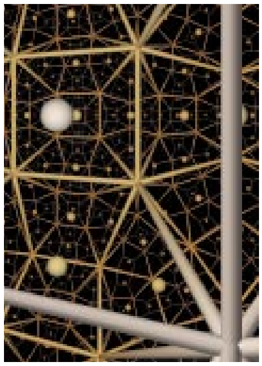
Scientific American (ISSN 0036-8733), published monthly by Scientific American, Inc., 415 Madison Avenue, New York, N.Y. 10017-1111. Copyright © 1998 by Scientific American, Inc. All rights reserved. No part of this issue may be reproduced by any mechanical, photographic or electronic process, or in the form of a phonographic recording, nor may it be stored in a retrieval system, transmitted or otherwise copied for public or private use without written permission of the publisher. Periodicals postage paid at New York, N.Y., and at additional mailing offices. Canada Post International Publications Mail (Canadian Distribution) Sales Agreement No. 242764. Canadian BN No. 127387652RT; QST No. Q1015332537. Subscription rates: one year $34.97 (outside U.S. $49). Institutional price: one year $39.95 (outside U.S. $50.95). 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; 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.**

From the Editors

Getting Complicated

The fate of the universe used to be so simple. It was either fire or ice. Either the combined gravity of the universe would bring its expansion to a halt, compelling the cosmos to replay the big bang in reverse, or else gravity would steadily weaken and the universe would expand forever, slowly and inexorably pulling planets, stars and galaxies apart until it became a barren, frigid void.

Now cosmologists realize that things aren't so straightforward. The universe may not be governed by the gravity of ordinary matter after all. If the latest observations of the distant universe (as discussed in our special report, beginning on page 45) are borne out, matter has little say in its own fate. Instead the universe may be controlled by the so-called cosmological constant, a surreal form of energy that imparts a gravitational repulsion rather than the usual attraction.



Geometry of space-time?

The idea of the constant has been embraced and renounced more than once since Albert Einstein initially proposed it 80 years ago. This time it may be here to stay. At first glance its shadowy reinforcement of cosmic expansion suggests that, as the ultimate fate, ice will have to suffice. But that judgment is premature. Because physicists know so little—"nothing" would be a fair approximation—about the constant, the fate of the universe is back where it started: in the realm of uncertainty.

One implication is that science writers who have been using Robert Frostian fire-and-ice allusions will have to find a new metaphor. Another is that the cosmos might be undergoing a second round of "inflation," a resurgence of the process that, 12 billion or so years ago, caused space to go bang. Just as that earlier period of explosive growth ended—giving form and light to what had been void—so, too, might the rekindled inflation. If so, the universe will expand to unimaginable proportions, the constant will fade away and physical possibilities will unfold that are only dimly perceived in today's theories.

If there is a story to be seen in cosmic history, it is the march from the utter simplicity of the big bang to ever increasing complexity and diversity. The near-perfect uniformity of the primordial fireball, and of the laws that governed it, has steadily given way to a messy but fertile heterogeneity: photons, subatomic particles, simple atoms, stars, complex atoms and molecules, galaxies, living things, artificial things.

Understanding how this intricacy is immanent in the fundamental laws of physics is one of the most perplexing philosophical puzzles in science. The basic rules of nature are simple, but their consummation may never lose its ability to surprise. A perpetual trend toward richness, the outcome of which cannot be foreseen, may be the true fate of the universe.

GEORGE MUSSER, *Staff Editor*
editors@sciam.com

STUART LEVY *National Center for Supercomputing Applications* AND TAMARA MUNZER *Stanford University*

LETTERS TO THE EDITORS

Parents responded in large numbers to Russell A. Barkley's "Attention-Deficit Hyperactivity Disorder" in the September issue. Many echoed the sentiments of Lynne Scholl from Cincinnati, Ohio, who wrote that "my daughter has many wonderful qualities that are a direct result of her ADHD. She is incredibly creative and inquisitive. Yes, there are differences in how she responds to her environment—some are good; some are not. Does this mean I should try to change her? I would not want to do that, so I choose to help her deal with her ADHD-related difficulties, just as I would teach her to cope with any of life's difficulties." Kathleen G. Nadeau, director of the Chesapeake Psychological Services of Maryland, suggested that "the high activity level, low boredom tolerance and impulsivity of people with ADHD bother teachers and challenge parents. These same traits, once the school years are through, translate very often into creativity, entrepreneurial capacity and high energy." Additional comments are included below.

ATTENTION, PLEASE

It was a pleasure reading Russell A. Barkley's article, "Attention-Deficit Hyperactivity Disorder," in your September issue. There is no doubt that Barkley's work has been tremendously valuable. But I have trouble with Barkley's assertion that ADHD is purely a neurological "disorder," best treated with stimulant medication. Before we rush to the pharmacist, I propose we cast a broader perspective.

The criteria for a diagnosis of ADHD are so general that any one of us could be diagnosed with it at some point in our lives. This accounts for a high false-positive rate when we attempt to classify people with many common complaints of impulsivity, agitation and difficulty focusing. Compounding this is the tremendous overlap between the criteria used for an ADHD diagnosis and those for other, more common disorders such as anxiety and depression.

We must recognize that no previous generation of children has been as hurried, overstimulated and subjected to powerful social challenges as today's American children have. It would be great if there were a magic pill to provide the cure—but the reality is that our clinics remain crowded with ADHD children, already on medications, who continue to seek help with navigating the demands of a stressful, frenetic childhood.

ANTHONY V. RAO
Department of Psychiatry
Harvard Medical School

Barkley disappointed this reader in several ways. The article implies a complete picture but ignores attention-deficit disorder (ADD), without hyperactivity. There are those who have quiet trouble channeling their attention; because they may procrastinate, forget or hyperfocus but don't disrupt, their difficulties are often overlooked. Barkley also doesn't discuss the originality and energy that can characterize ADD/ADHD. Often creativity is ADD/ADHD gone right.

PRISCILLA L. VAIL
Bedford, N.Y.

Barkley's article on ADHD was very informative; however, he does parents a disservice when he characterizes special education as an option of last resort. Parents should not be discouraged from seeking special education services at an early point in their child's education, as it could help the child avoid years of frustration. Barkley seems to be under the impression that special education means separate classes. This is not the case: more and more students across the U.S. receive special education services while being included in regular classes. Information centers in each state can help parents learn about special education support for students with ADHD. Call the Technical Assistance Alliance for Parent Centers at 888-248-0822 for the location of the center nearest you.

DEBORAH LEUCHOVIUS
PACER Center
Minneapolis, Minn.

MANAGING ADHD often requires a combination of medication and special education.

YAN NASCIMBENE

Barkley replies:

People with ADHD undoubtedly have many fine personal attributes, as parents of ADHD children can testify. But no study has ever shown that having ADHD results in enhanced creativity, intelligence and the like. In fact, some studies have shown that ADHD can reduce IQ scores by an average of seven to 10 points and diminish certain forms of creativity.

Rao is mistaken—I did not suggest medication alone as a treatment for ADHD. Rather it is part of a package that should include special education and other accommodations as needed, as Leuchovius's letter emphasizes. Twin studies show that environmental factors such as a hectic society do not cause ADHD. Furthermore, the criteria for ADHD are surely not so broad as to diagnose everyone with the disorder: studies find that only 3 to 7 percent of children meet the diagnostic criteria.

Vail mentions ADD involving only inattention, without hyperactivity; space constraints on the article precluded my addressing this possibly distinct disorder. Inattentiveness in ADD patients may be qualitatively different, resulting in a lower risk for antisocial behavior and a different pattern of cognitive deficits.

SPACE CONSTRAINTS

Ronald White is correct that NASA research on "Weightlessness and the Human Body" [September] is good news for denizens of the International Space Station and for future interplanetary missions. The evidence suggests, however, that such research is of limited value to those of us here on Earth. Twenty years of microgravity research on Mir and earlier space stations have yielded some interesting insights but no major breakthroughs. Space medicine cannot justify its enormous cost when the National Institutes of Health can fund only about 25 percent of the deserving research applications it receives. The staggering $1.3-billion annual cost of keeping four American researchers in orbit could pay for 5,000 or more grants for cutting-edge research at laboratories and universities. Surely the best way to study aging and to improve medical care is to spend our limited resources on Earth.

DALE BUMPERS
U.S. Senator, Arkansas

White replies:

The decision to build the International Space Station and to continue human space flight was based on many factors, not just on the benefits that might result from biomedical research in space. Thus, it is grossly misleading to weigh the benefits of such research against the entire cost of the human space flight program. I hope my article did not lead readers to believe that space biomedical research would somehow replace NIH-supported research on aging, osteoporosis or anything else. The NASA biomedical research program is much smaller than the NIH program. The two approaches are complementary, not mutually exclusive. Judging by our experience so far, I believe that space biomedical research will make unique and important contributions to health on Earth at the same time as it improves the health of space travelers.

Letters to the editors should be sent by e-mail to editors@sciam.com or by post to Scientific American, 415 Madison Ave., New York, NY 10017. Letters may be edited for length and clarity.

ERRATA

"The Asymmetry between Matter and Antimatter" [October] contains an error on page 77 regarding the handedness of neutrinos. The article should have stated that as far as we know, there are no right-handed neutrinos: they are always left-handed. In "Cryptography for the Internet" [October], the screen shots shown were from the program QuickMail Pro. We apologize for the confusion.

OTHER EDITIONS OF SCIENTIFIC AMERICAN

Le Scienze
Piazza della Repubblica, 8
20121 Milano, ITALY
tel: +39-2-29001753
redazione@lescienze.it

Spektrum der Wissenschaft
Verlagsgesellschaft mbH
Vangerowstrasse 20
69115 Heidelberg, GERMANY
tel: +49-6221-50460
redaktion@spektrum.com

Investigacion y Ciencia
Prensa Científica, S.A.
Muntaner, 339 pral. 1.ª
08021 Barcelona, SPAIN
tel: +34-93-4143344
precisa@abaforum.es

Pour la Science
Éditions Belin
8, rue Férou
75006 Paris, FRANCE
tel: +33-1-55-42-84-00

Majallat Al-Oloom
Kuwait Foundation for
the Advancement of Sciences
P.O. Box 20856
Safat 13069, KUWAIT
tel: +965-2428186

Swiat Nauki
Proszynski i Ska S.A.
ul. Garazowa 7
02-651 Warszawa, POLAND
tel: +48-022-607-76-40
swiatnauki@proszynski.com.pl

Nikkei Science, Inc.
1-9-5 Otemachi, Chiyoda-ku
Tokyo 100-8066, JAPAN
tel: +813-5255-2821

Svit Nauky
Lviv State Medical University
69 Pekarska Street
290010, Lviv, UKRAINE
tel: +380-322-755856
zavadka@meduniv.lviv.ua

Ke Xue
Institute of Scientific and
Technical Information of China
P.O. Box 2104
Chongqing, Sichuan
PEOPLE'S REPUBLIC OF CHINA
tel: +86-236-3863170

50, 100 AND 150 YEARS AGO

SCIENTIFIC AMERICAN

JANUARY 1949

NUCLÉAIRE—"The first self-sustaining chain reaction to be produced outside of the English-speaking nations has just been achieved by French physicists. Frederic Joliot-Curie, director of the French Atomic Energy Commission, announced that a uranium pile went into operation last month at Fort de Châtillon, on the outskirts of Paris. To U.S. workers, who have taken great pains to refine the uranium used in their reactors, the ability of the Châtillon pile to sustain itself on impure uranium (uranium oxide) is something of a surprise."

OEDIPUS COMPLEX—"Freud knew the Oedipus myth from Sophocles' tragedy *King Oedipus*. The question is whether Freud was right in assuming that this myth confirms his view that unconscious incestuous drives and resulting hate against the father-rival are an intrinsic part of any male child's equipment. If we examine the myth more closely, however, doubts arise. There is no indication whatsoever in the myth that Oedipus is attracted by or falls in love with Jocasta. The myth has to be understood as a symbol, not of the incestuous tie between mother and son, but of the rebellion of the son against the authority of the father in the patriarchal family; the marriage of Oedipus and Jocasta is a symbol of the victory of the son who takes over his father's place and with it all the privileges.—Erich Fromm"

Oedipus, king of Thebes, with Jocasta, his queen

JANUARY 1899

POLONIUM AND RADIUM—"Two of us have shown that, by purely chemical processes, a strongly radio-active substance can be extracted from pitchblende. We therefore came to the conclusion that pitchblende might contain a new element, for which we proposed the name of polonium. Subsequently, we have met with a second substance, strongly radio-active, and entirely differing from the first body in its chemical properties. The new radio-active substance has the properties of almost pure barium; its chlorides, however, having a radioactivity 900 times greater than that of uranium. We believe that the new radio-active substance contains a new element, to which we propose to give the name of radium.—M. P. Curie, Mme. P. Curie, and M. G. Bémont"

BAD AIR—"Dr. G. B. Grassi for a long time had doubts on the connection between mosquitoes and malaria, owing to the absence of malaria from certain districts where mosquitoes abound. A careful classification of the various species of gnat has now led him to the conclusion that the distribution of certain kinds coincides very closely with the distribution of the disease. The common Culex pipiens is to be regarded as perfectly innocuous. On the other hand, a large species (Anopheles claviger, Fabr.) known in Italy as 'zanzarone,' or 'moschino,' is constantly found associated with malaria, and is most abundant where the disease is most prevalent."

JANUARY 1849

BIOCIDES FOR AGRICULTURE—"The London Lancet mentions a practice which is common among the English farmers, of steeping their wheat in a solution of arsenic before sowing it, to prevent the ravages of the worm on the seed, and of birds on the plant when grown. The plan is stated to have proved eminently successful, and of course exerts no deleterious effects on the plant. In Hampshire, Lincolnshire, and many other districts where the practice prevails, numbers of partridges and pheasants have been found dead in the wheat fields, poisoned by eating the seed. This is certainly a practice to be condemned. We can afford to feed both men and birds."

MAINSTREAM NICOTINE—"Prout, in his Treatise on Disease, says about tobacco, 'Although confessedly one of the most virulent poisons in nature, yet such is the fascinating influence of this noxious weed, that mankind resorts to it in every mode that can be devised to insure its stupefying and pernicious agency. The severe and dyspeptic symptoms sometimes produced by inveterate snuff-takers are well known; and I have seen such cases terminate fatally with malignant diseases of the stomach and liver. Surely, if the dictates of reason were allowed to prevail, an article so injurious to the health and so offensive in its mode of employment would speedily be banished from common use.'"

A GREAT DIAMOND—"Koh-i-noor—or, 'mountain of light.' A diamond of inestimable value has been taken by the British troops in India, from one of the native princes. It is proposed to insert it in the centre of Queen Victoria's diadem."

The 1998 Nobel Prizes in Science

AP/WIDE WORLD PHOTOS

Here follow explanations of the mechanisms and processes that underlie the world's top awards for physics, chemistry and physiology—and an excerpt from a Scientific American *article by the economics laureate*

PHYSICS

HOW ELECTRONS SPLIT

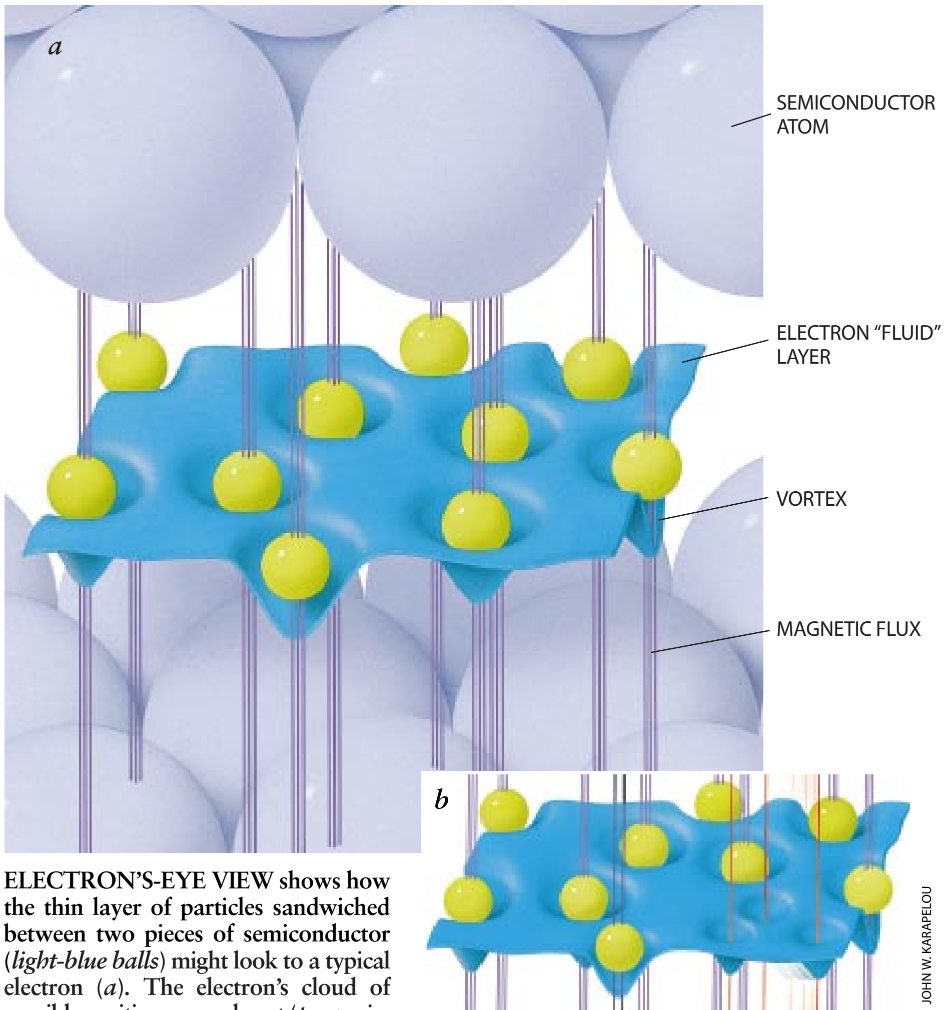
HORST L. STÖRMER
Bell Laboratories

DANIEL C. TSUI
Princeton University

ROBERT B. LAUGHLIN
Stanford University

The humming, beeping, well-lit modern world could not have been built without the knowledge that electric current is a parade of electrons and that those particles are not ricocheting billiard balls but fuzzy clouds of probability that obey odd rules of etiquette as they maneuver in a dance of mutual repulsion. Discoveries about how electrons behave can thus have far-reaching consequences, although they may seem little more than curiosities at the time. Superconductivity was one example. One day it may turn out that the discovery for which Horst L. Störmer of Bell Laboratories, Daniel C. Tsui of Princeton University and Robert B. Laughlin of Stanford University received the 1998 Nobel Prize in Physics is another.

Störmer and Tsui tortured electrons for their secrets. They squeezed electrons into a layer so thin that particles could move neither up nor down. They zapped the electrons with high magnetic flux. And they chilled the whole assembly to within a hair of absolute zero. Then physicists saw something unexpected. The electrical resistance across the thin current of electrons rose in steps rather than a straight line as they turned up the magnetic field. The plateaus suggest that



JOHN W. KARAPELOU

ELECTRON'S-EYE VIEW shows how the thin layer of particles sandwiched between two pieces of semiconductor (*light-blue balls*) might look to a typical electron (*a*). The electron's cloud of possible positions spreads out (*turquoise sheet*) like a liquid to fill the layer except for spots where bits of magnetic flux (*violet lines*) zip through the ceiling. The electron avoids those spots, so vortices in its cloud open there. Other electrons in the area (*green balls*), repulsed by the first electron and by one another, naturally drift into the holes. As they do, they become bound to the lines of magnetism. If an electron is bumped out of the layer, it leaves behind an unoccupied vortex that can then split into smaller holes (*b*). Three rays of magnetic flux anchoring a single electron can thus become three separate "quasiparticles" (*red lines*), each carrying one third of the original charge. Similarly, if the magnetic field is reduced slightly, a ray of flux may disappear, causing one vortex to shrink (*black lines*) and creating an apparent excess of one-third electron charge at that point.

some new kind of particle was carrying fractions—$^{1}/_{3}$, $^{2}/_{5}$, $^{3}/_{7}$ and so on—of a single electron charge. Electrons are fundamental particles: they do not split. So what was going on?

It took Laughlin a year to work out a theoretical explanation, which further experiments have since supported. Cramped and cold, the usually frenetic electrons condense into a kind of fluid. Vortices in the fluid match up bits of magnetism with electrons. If there are not enough magnetic lines to share equally, some of the vortices can separate from their electrons and dance about independently, carrying fractions of positive charge (*opposite page*).

The fractional quantum Hall effect, as it is called, occurs in rare conditions. But that does not mean it will lack applications. When quantum wells were discovered, they were equally rare and curious. Today they are built into nearly every compact-disc player sold.

CHEMISTRY

REACTIONS ON A COMPUTER

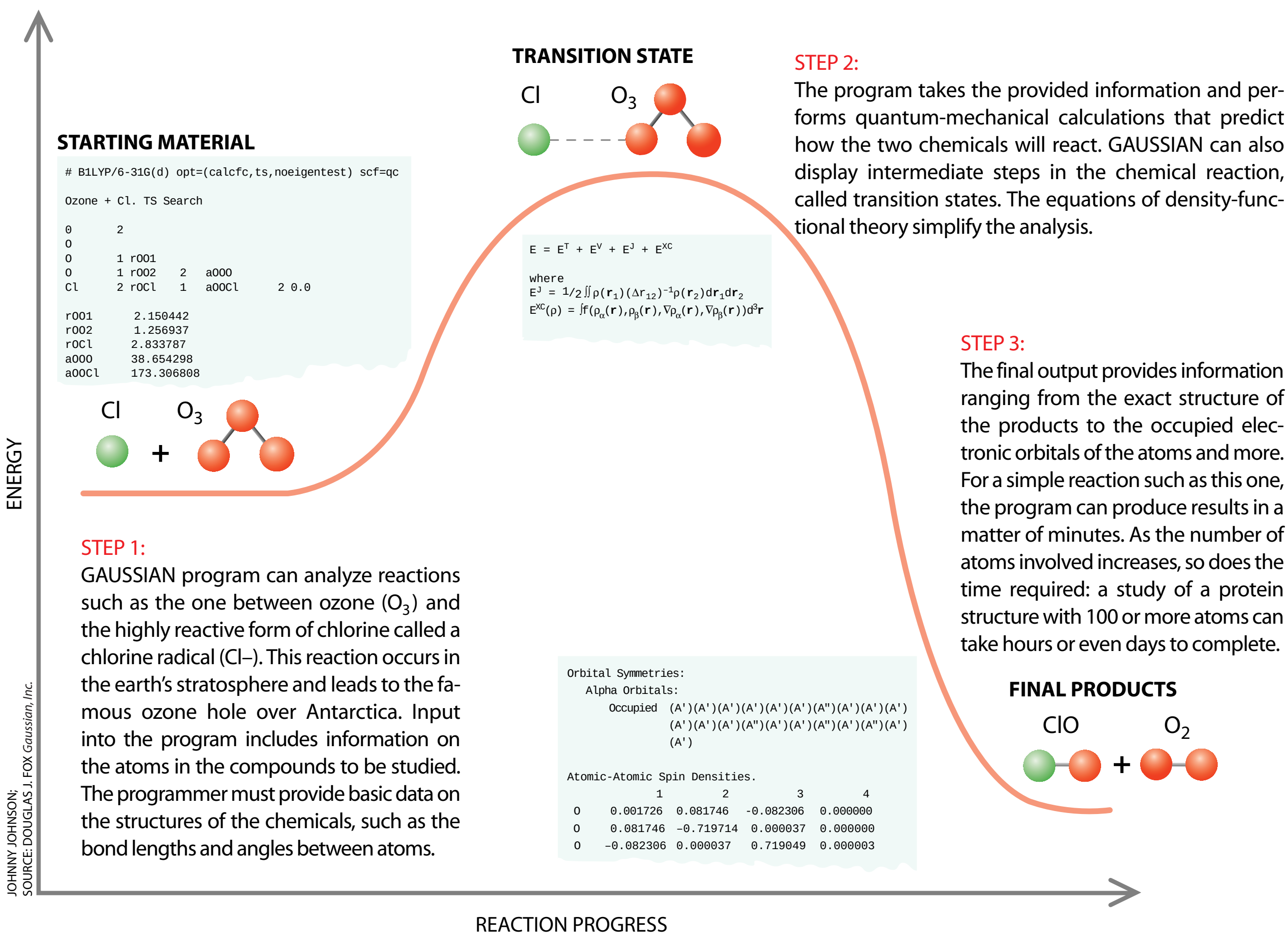
WALTER KOHN
University of California, Santa Barbara

JOHN A. POPLE
Northwestern Unversity

Predicting how chemicals will react is not an easy business, even for computational chemists, who study virtual reactions on computers rather than mixing chemicals in beakers. Chemical reactions involve the breaking and reformation of bonds between atoms; whether or not a bond will form depends on the position and energy of the atom's electrons. This year's Nobel Prize in Chemistry recognizes advances in computational techniques that predict reactions more quickly and accurately.

Nobel recipient Walter Kohn developed the computational method known as density-functional theory. It can be used to determine a molecule's structure and other properties; more important, it greatly simplifies essential calculations. Instead of tracking the motion of each individual electron in a given molecule (large molecules can contain hundreds or even thousands of electrons), Kohn's technique uses quantum mechanics to consider the overall density of electrons throughout the molecule. With density-functional theory, chemists today can often perform structure calculations on desktop computers instead of mainframes.

One program popular among chemists that incorporates Kohn's density-functional theory in addition to many other computational techniques was developed by the co-recipient of this year's prize, John A. Pople. He designed the program GAUSSIAN, first released in 1970. More than 10,000 scientists now use the latest version of it.



PHYSIOLOGY OR MEDICINE

A VERSATILE GAS

ROBERT F. FURCHGOTT
State University of New York
Health Science Center at Brooklyn

LOUIS J. IGNARRO
U.C.L.A. School of Medicine

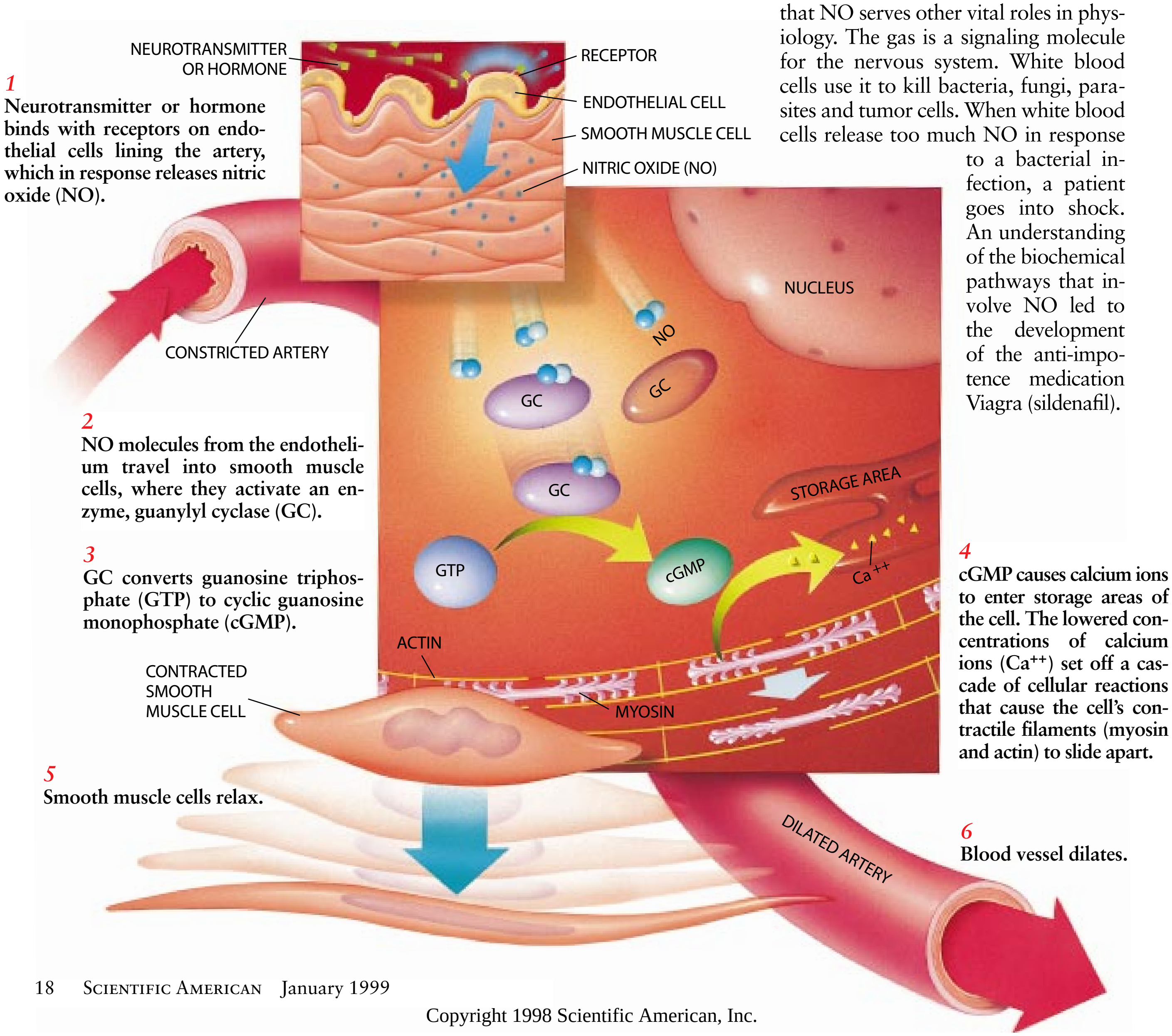
FERID MURAD
University of Texas Medical School
at Houston

Careers that seek to counter the conventional wisdom may either founder in obscurity or garner the highest accolades. Robert F. Furchgott, Louis J. Ignarro and Ferid Murad received the 1998 Nobel Prize in Physiology or Medicine for discoveries related to the biological function of a molecule that was once primarily known as an air pollutant. The three were heralded for elucidating nitric oxide's role in initiating cellular events that dilate blood vessels. "Signal transmission by a gas that is produced by one cell, penetrates through membranes and regulates the function of another cell represents an entirely new principle for signaling in biological systems," noted the Nobel Assembly at the Karolinska Institute in Sweden. (Actually, a letter published in *Science* subsequent to the Nobel announcements pointed out that ethylene gas had been recognized as a signaling molecule in plants since 1934.)

Nevertheless, many scientists originally dismissed the notion that a gas like nitric oxide (NO) could be an intercellular messenger. The typical signal molecules are proteins, peptides or smaller organic molecules. NO, a highly reactive gas, is so unstable that reactions with oxygen or water will convert it into nitrites or nitrates within 10 seconds.

But as Furchgott, Ignarro and Murad showed, NO is essential to keeping blood vessels wide open to maintain blood flow and pressure (*below*). In atherosclerosis, in which plaque occludes the coronary arteries, the cells lining the blood vessels produce less NO. The work that led to the Nobel explains why patients with chest pain (angina pectoris) caused by atherosclerosis get relief from pills containing nitroglycerin: the compound, once it has entered the smooth muscle cells, releases NO. Ironically, dynamite, invented by Alfred Nobel, the founder of the prizes, contains nitroglycerin as its active ingredient.

In recent years, scientists have found that NO serves other vital roles in physiology. The gas is a signaling molecule for the nervous system. White blood cells use it to kill bacteria, fungi, parasites and tumor cells. When white blood cells release too much NO in response to a bacterial infection, a patient goes into shock. An understanding of the biochemical pathways that involve NO led to the development of the anti-impotence medication Viagra (sildenafil).



JOHN W. KARAPELOU

ECONOMICS

THE ETHICAL DIMENSION

AMARTYA SEN
University of Cambridge

An Indian newsweekly featured Amartya Sen on the cover of a late October 1998 issue with the headline "The Prophet We Ignore." The scholar of poverty has spent decades devising novel approaches to solving India's woes—and the government of his native country has often chosen to forgo his advice, the magazine contends.

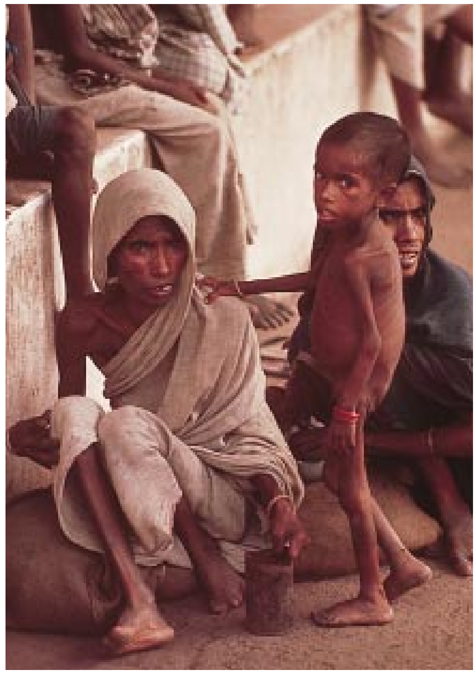
Nevertheless, Sen's work has not gone unnoticed. The Royal Swedish Academy of Sciences chose to award Sen a Nobel—more formally, the Bank of Sweden Prize in Economic Sciences in Memory of Alfred Nobel—for his contributions to welfare economics, the study of the way societies make fair choices about allocating resources. His work deals with fundamental questions such as how income inequality should be measured and what the conditions that lead to famines are. The academy noted that Sen's melding of tools from philosophy with economics "restored an ethical dimension to the discussion of vital economic problems."

In May 1993 Sen wrote an article in Scientific American called "The Economics of Life and Death." In the following excerpt, Sen discusses the genesis of a famine:

Economic explanations of famine are often sought in measures of food production and availability. And public policy is frequently based on a country's aggregate statistics of the amount of food available per person, an indicator made prominent by Thomas Robert Malthus in the early 1800s. Yet contrary to popular belief, famine can result even when that overall indicator is high. Reliance on such simple figures often creates a false sense of security and thus prevents governments from taking measures to avert famine.

A more adequate understanding of famine requires examining the channels through which food is acquired and distributed as well as studying the entitlement of different sections of society. Starvation occurs because a substantial proportion of the population loses the means of obtaining food. Such a loss can result from unemployment, from a fall in the purchasing power of wages or from a shift in the exchange rate between goods and services sold and food bought. Information about these factors and the other economic processes that influence a particular group's ability to procure food should form the basis of policies designed to avoid famine and relieve hunger.

The Bangladesh famine of 1974 demonstrates the need for a broader appreciation of the factors leading to such a calamity. That year, the amount of food available per capita was high in Bangladesh: indeed, it was higher than in any other year between 1971 and 1976. But floods that occurred from late June until August interfered with rice transplantation ... and other agricultural activities in the northern district. Those disruptions, in turn, caused unemployment among rural laborers, who typically lead a hand-to-mouth existence. Bereft of wages, these workers could no longer buy much food and became victims of starvation....



HUBERT LE CAMPION SYGMA

BANGLADESH FAMINE of 1974 took place even though the amount of food available per person that year was high.

[The situation was exacerbated by precautionary hoarding and speculative stockpiling, which caused prices to rise and hurt the food-buying ability of poor Bangladeshis.] When food prices peaked in October, so also did the death toll....

The occurrence of this famine illustrates how disastrous it can be to rely solely on food supply figures. Food is never shared equally by all people on the basis of total availability. In addition, private and commercial stocks of produce are offered to or withdrawn from the market in response to monetary incentives and expectation of price changes....

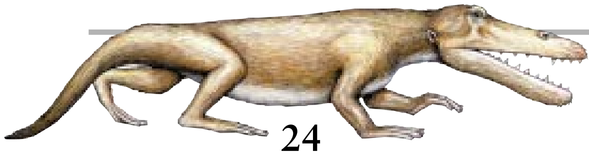
There are several ways to prevent famine. In Africa and Asia, growing more food would obviously help, not only because it would reduce the cost of food but also because it would add to the economic means of populations largely employed in producing food.... Augmenting food production, however, is not the only answer. Indeed, given the variability of the weather, concentrating too much of a nation's resources on growing more food can increase the population's vulnerability to droughts and floods. In sub-Saharan Africa, in particular, there is a strong need for the diversification of production, including the gradual expansion of manufacturing....

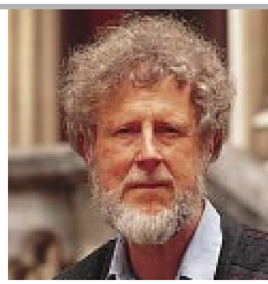
No matter how successful the expansion of production and diversification may be in many African and Asian countries, millions of people will continue to be devastated by floods, droughts and other disasters. Famine can be averted in these situations by increasing the purchasing power of the most affected groups—those with the least ability to obtain food. Public employment programs can rapidly provide an income. The newly hired laborers can then compete with others for a share of the total food supply. The creation of jobs at a wage does, of course, raise prices: rather than letting the destitute starve, such practice escalates the total demand for food. That increase can actually be beneficial, because it brings about a reduction in consumption by other, less affected groups. This process distributes the shortage more equitably, and the sharing can deter famine.

Reporting for the section by W. Wayt Gibbs, Sasha Nemecek and Gary Stix.

See also www.sciam.com/explorations/1998/1019nobel/index.html on the World Wide Web.

NEWS AND ANALYSIS





IN FOCUS

THE NET EFFECT

The Internet can be a powerful tool for political dissidents and "hacktivists." But the medium has yet to reach the grassroots level

NOEL QUIDU *Gamma Liaison*

***INDONESIAN PROTESTERS** mobilize first on the Internet, then on the streets.*

The Internet has dramatically altered the way many people perform numerous tasks—communicating with one another, shopping, banking, making travel arrangements, keeping abreast of the news. Now add to the list political and human-rights reform. Proponents in those fields assert that the Internet and the World Wide Web have become essential tools for effecting change. But critics contend that the medium is often least available where it is most needed.

The ongoing struggle for democracy in Indonesia underscores the power of the Internet. Last spring protesters bypassed the state-controlled media there by posting a Web site containing a database that kept track of the corruption of then president Suharto. People across the country were continually adding information about the accumulated wealth of the president and his children, knowledge of which fueled an already inflammatory situation. Students also relied on the Internet to coordinate their demonstrations, which eventually led to Suharto's resignation.

Indeed, political dissenters and human-rights organizations around the world have taken advantage of the Internet's ability to disseminate information quickly, cheaply and efficiently. The Zapatista rebels have exploited it to garner support among international journalists and sympathizers against the Mexican government. The Free Burma Coalition uses its Web site to encourage consumers to boycott companies doing business in Myanmar. And the Digital Freedom Network routinely posts on the Web the writings of political dissidents, such as Raúl Rivero of Cuba, who are censored in their homelands. "To build up on-line communities with such limited resources is amazing," notes Xiao Qiang of Human Rights in China, a group based in New York City, which uses the Internet to organize letter-writing campaigns. Adds William F. Schulz, New York executive director of Amnesty International USA, "the Web is a critical new tool that we now have. It has radically increased our ability to funnel information."

CHINESE SOCIETY FOR HUMAN RIGHTS STUDIES

ANTIONLINE

GOVERNMENT WEB SITES
have become the targets of "hacktivists." When a Chinese agency recently tried to proclaim its efforts in human rights, the home page **(top)** ***was quickly—and unofficially—replaced*** **(bottom).**

For their part, governments face a quandary: How do they cobble together restrictive policies that will help them maintain the status quo *without* stifling the Web's many business benefits? Because of Indonesia's solid economic growth before the recent downturn, the country had a hands-off policy toward the Internet, which many companies had used to communicate with suppliers and customers across the sprawling archipelagic nation. But the same medium that enabled firms there to monitor the status of their factories and inventories also allowed dissidents to mobilize.

Meanwhile the Internet's role in political and human-rights reform has been evolving beyond mere information dissemination and calls for action. On Mexican Independence Day, thousands of people staged a "virtual sit-in" to protest the government's treatment of Zapatista rebels in Chiapas. The digital demonstrators tried to overwhelm targeted Web sites, including those of Mexican president Ernesto Zedillo, reportedly by using an automated software program to issue repeated phony requests to download information.

Other groups have gone further, breaking into systems and defacing Web pages. Last October, soon after the Chinese government had launched a new Web site to proclaim its efforts in human rights, hackers replaced the home page with one containing a diatribe: "China's people have no rights at all, never mind Human Rights." Other "hacktivists" have plied their craft to protest conditions in various areas—among them East Timor, Indonesia; Kashmir, India; and Kosovo, Serbia—knowing all too well that attacking a government is usually much easier electronically than physically. And often the main reason for such electronic rabble-rousing is not the actual acts themselves but the follow-up media attention that can garner quick, worldwide publicity for a cause.

John Vranesevich, founder of AntiOnline, a Web site that tracks hacker activities, predicts that the number of such electronic exploits will escalate in the future as the first generation of young hackers matures. "These hackers are becoming politically minded," Vranesevich says. "They are starting to vote, and they are starting to take a look at the world around them. Now they are using the skills they've honed to make their opinion heard." Bronc Buster, a pseudonym for the 26-year-old who led the attack on the Chinese human-rights server, recalls that when he first saw that Web site he was outraged. "Two years ago, when I was a freshman, I had to do a huge paper on China for one of my political science classes, so I knew what was happening over there," he says. "When I went to that site and read what was on it, I got extremely mad. It reminded me of the Nazis saying the Holocaust never happened."

Yet while some people have proclaimed the dawning of a new age in electronic activism, others caution that the Internet's effect may be grossly exaggerated. Of a total worldwide population of about six billion people, only a tiny fraction is wired, and most of that is in North America, Europe and Japan, geographic areas not particularly known for political tyranny or egregious human-rights violations. For this reason, critics say the view of the Internet as a juggernaut for implementing sweeping reforms is an overblown, North-centric perspective. "How many people in the world have never even made a phone call? Maybe a third to a half. And how much impact do you think the Web's having on them?" asks Patrick Ball, senior program associate for the Science and Human Rights Program of the American Association for the Advancement of Science.

The North-South dichotomy could worsen as the experiences of countries such as Indonesia and China make other nations wary of going on-line. In Saudi Arabia, for example, Internet service providers must apply for a license through the government, which requires that Web traffic be filtered through state-controlled proxy servers. And a host of governments have stepped up their efforts to make certain activities illegal, if for no other reason than to instill a chilling effect among the general populace. Last spring a Shanghai software engineer was arrested for allegedly sending a list of the e-mail addresses of thousands of Chinese to a U.S.-based dissident publication. Such acts notwithstanding, countries have also been loath to pull the plug on the Internet, fearing that the medium will be essential for their future economic success.

But the greatest value of the Internet certainly goes far beyond the actual numbers of people on-line, asserts Jagdish Parikh of Human Rights Watch in New York City. "How many people in China have Internet access? Not many," he notes. "But then why is the government there rushing to make laws restricting access? It's because the Internet makes people realize that they should have the legal, codified right to information." —*Alden M. Hayashi*

ANTHROPOLOGY

OUT OF AFRICA, INTO ASIA

Controversial DNA studies link Asian hunter-gatherers to African pygmies

Scientists may have pinpointed direct descendants of the first humans to migrate out of Africa into Asia. They could be the aboriginal inhabitants of the Andaman Islands in the Bay of Bengal, who have long been noted for their resemblance to African pygmies. Some convergence of features—dark skin and small, gracile form—is to be expected in peoples who have evolved in the tropics. But a recent DNA study of hair from Andamanese individuals, collected in 1907 by British anthropologist Alfred R. Radcliffe-Brown, suggests a closer connection.

Carlos Lalueza Fox, a postdoctoral fellow at the genetics laboratory of Erika Hagelberg at the University of Cambridge, had extracted DNA from 42 out of 70 hair samples and amplified a short segment of DNA from the mitochondria. Known as mtDNA, such DNA is less directly related to physical characteristics than chromosomal DNA and is therefore believed to be less sensitive to the pressures of natural selection. Fox and Hagelberg found that the sequences of base pairs in the mtDNA fragments clustered closer to African populations—especially southern African pygmies—than to Asian ones.

If substantiated, the findings will lend support to the Out of Africa theory of human descent. Proponents hold that the first humans left Africa some 100,000 years ago, reaching Asia around 60,000 years ago. According to Peter Bellwood of Australian National University in Canberra, some of these hunter-gatherers moved southward to New Guinea and Australia during the ice ages 40,000 years ago. At the time, glaciers had sucked water out of the oceans, lowering the sea level and expanding Asia into a vast region known as Sundaland. As a result, much of the southward migration occurred on foot.

Archaeological evidence of human occupation of the Andamans, excavated most recently by Zarine Cooper of Deccan College in India, dates back at most 2,200 years. But Bellwood guesses that the Andamanese reached their islands during the first wave of human migration at least 35,000 years ago. Eventually the seas rose, cutting them off. The seas were to fall and rise many more times, most recently about 10,000 years ago. Andamanese mythology describes violent storms and deluges that drowned the islands, forcing the survivors to repair to the former hilltops.

Almost all the first humans in Asia were wiped out by waves of later migrants; survivors persisted only in isolated, embattled pockets. The Andamanese ensured their own survival—at least until modern times—by determined opposition to all seafarers who attempted to land. To this day, one group of Andamanese, inhabiting tiny North Sentinel Island, attacks with arrows any approaching boats.

The Out of Africa theory has also received recent support from an extensive survey of Chinese DNA conducted by Li Jin of the University of Texas at Houston and his colleagues. The researchers examined DNA markers called microsatellites from 28 ethnic groups across China, including four from Taiwan. They found only minor genetic variations among the populations, suggesting that these groups had had little time to diverge from one another. Possibly, they all arose from recent African migrants.

A rival scenario derives from the Multiregional hypothesis, which holds that humans evolved separately in different parts of the world from populations of *Homo erectus* that dispersed (also from Africa) one to two million years ago. These groups of humanoids managed to develop into a single species—*H. sapiens*—by exchanging genes with one another. To some anthropologists, fossils excavated in China suggest a continuum between *H. erectus* and modern Chinese peoples. Milford Wolpoff of the University of Michigan has pointed out that interbreeding could have ensured that the descendants of different humanoids ended up being genetically similar.

Wolpoff is likewise skeptical of the Andaman study, which cannot be properly critiqued until it is published. An unfortunate dispute regarding the hair has held up publication. Robert A. Foley, director of the Duckworth Collection at Cambridge, which owns the hair, has complained that permission was never obtained for its use. Hagelberg protests that Foley knew about the study for at least a year before voicing this objection when the results were reported at a conference in August. Matters became so unpleasant that Hagelberg has packed up her lab and moved to the University of Otago in Dunedin, New Zealand.

The research will be difficult to replicate, because fresh materials from the Andamanese are scarce. Access to blood, hair and other human samples is restricted by many countries (in this case, India) for fear that the genetic information they contain will be misused—specifically, put to commercial use. So it will be a while before the intriguing links between Andamanese and Africans strengthen into familial bonds. *—Madhusree Mukerjee*

MADHUSREE MUKERJEE

ANDAMANESE MALE ***from the dense forests of Middle Andaman Island belongs to a group that has recently been emerging to meet with settlers in peace.***

IN BRIEF

Growing Stem Cells
Last November saw major advances in cultivating human embryonic stem cells—a "holy grail" of biotechnology. Such cells can become any of the body's tissues, so the cultivation of them could lead to organs on demand. James A. Thomson and his colleagues at the University of Wisconsin described in *Science* how they coaxed days-old embryo cells to grow indefinitely in their undifferentiated state while retaining their ability to become specialized tissue cells. Days later John D. Gearhart of Johns Hopkins University and his co-workers reported a similar feat, using primordial germ cells (cells that would eventually become sperm and eggs). In unpublished work, researchers at Advanced Cell Technology in Worcester, Mass., say they fused nuclei from adult human cells with cow eggs that had their nuclei removed. The human nuclei commandeered the bovine cells, turning them into embryonic stem cells.

Cosmic Forecast
Processes in deep space, it seems, can influence the earth's climate. Henrik Svensmark of the Danish Meteorological Institute found that during the last 11-year activity cycle of the sun, the earth's cloud cover was more closely correlated with the flux of cosmic rays coming from the rest of the galaxy than with the sun's radiance. Apparently, the solar magnetic field interacts with the cosmic rays: when strong, the sun's field blocks more cosmic rays, which ionize air molecules in the lower atmosphere and in this way are thought to contribute to cloud cover and other weather-related phenomena.

Created by cosmic rays

JIM CORWIN *Photo Researchers, Inc.*

What Friends Are For
Playing in front of a home crowd may not be so advantageous. Jennifer L. Butler of Wittenberg University and her colleagues showed that individuals performing difficult tasks, such as doing stressful arithmetic, were less likely to succeed in front of a supportive audience than in front of a neutral or adversarial one. The reason? In front of unfriendly faces, people do not concern themselves with disappointing the audience and therefore tend to perform with greater concentration.

More "In Brief" on page 30

PALEONTOLOGY

CETACEAN CREATION

New fossils leave researchers wondering where whales come from

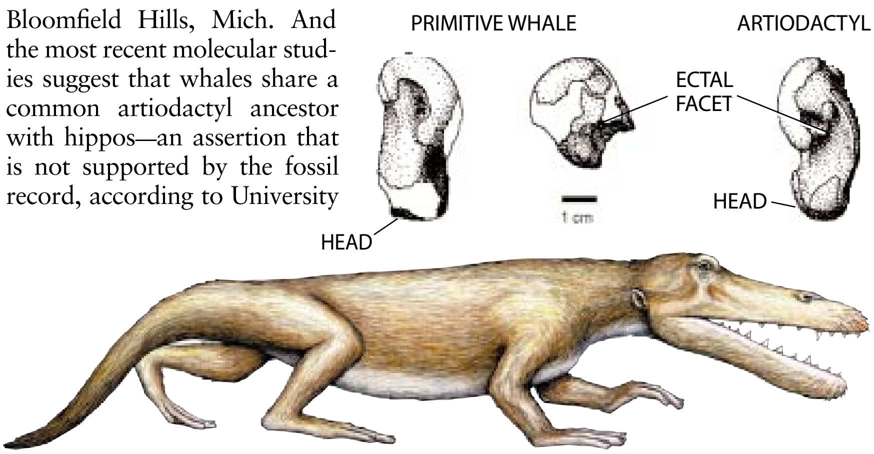
From four-legged landlubbers to streamlined ocean dwellers, whales represent one of the most dramatic evolutionary transformations. But what their terrestrial ancestors were and how whales are related to other living mammals have eluded scholars for over a century. Paleontologists have long held that whales are most closely related to extinct, wolflike creatures called mesonychians, based on striking dental similarities. A few years ago, however, molecular biologists weighed in with DNA data suggesting that whales are actually highly specialized artiodactyls (the group that includes hippopotamuses, camels, pigs and ruminants) and are closer to one of those living subgroups than to mesonychians.

Now key fossils—50-million-year-old whale ankle bones from Pakistan—have been unearthed. But instead of shedding light on whale origins as expected, they have left researchers even more puzzled than before.

Paleontologists agree that among living mammals, artiodactyls are the closest relatives of whales and that they share a common ancestor in the distant past, but saying that an artiodactyl was an ancestor to whales is "a really different, much more specific hypothesis," explains Mark D. Uhen of the Cranbrook Institute of Science in Bloomfield Hills, Mich. And the most recent molecular studies suggest that whales share a common artiodactyl ancestor with hippos—an assertion that is not supported by the fossil record, according to University of Michigan paleontologist William J. Sanders. He points out that the earliest known fossil branching of hippos was 15 to 18 million years ago and the earliest whales more than 50 million years ago in the Eocene epoch. Thus, if whales and hippos shared a common ancestor, it would have to have persisted for at least 32 million years—but there is no fossil evidence for such a creature spanning that immensity of time. And Sanders is not persuaded by the proposed hippo ancestors that might bridge that gap. "In terms of fossils in the right time, in the right place and in the right form," states Philip D. Gingerich, also at Michigan, "[mesonychians] are the only things that we know so far that are candidates for the ancestry."

For their part, the molecular biologists are confident that the DNA data show conclusively that whales share a special relationship with hippos. "Frankly, I think the issue is settled," declares Michel C. Milinkovitch of the Free University of Brussels. "The molecular data smoke the morphological evidence." Complaints from paleontologists that the DNA evidence is "noisy"—that is, the similarities reflect convergent evolution rather than common ancestry—have recently been addressed: Norihiro Okada of the Tokyo Institute of Technology and his colleagues have unpublished analyses of snippets of noncoding DNA called SINEs (short interspersed elements), which are purportedly noise-free, and the results support the whale-hippo link.

Still, paleontologists point out that the molecular analyses include data only from extant animals. Because most of the group of interest is extinct, the DNA data



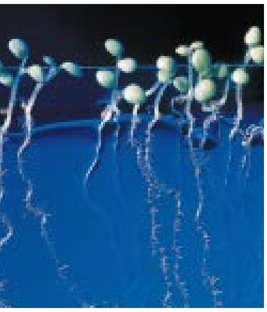
ANCIENT ANKLE FRAGMENTS FROM BEASTS like* Ambulocetus, *a primitive whale, add to the mystery of whale origins. The whale bone head is not rounded, arguing against a descent from artiodactyls. But similarities in other joint surfaces, such as the ectal facet, support artiodactyl ancestry.

JOHN KLAUSMEYER

In Brief, continued from page 26

Neuroweeds

Weeds appear to use the same kind of neurotransmitting system that humans do. Gloria Coruzzi and her colleagues at New York University found that the weed *Arabidopsis* has genes that encode for glutamate receptors. Glutamate is one of the neurotransmitters the human brain relies on for several functions, including memory formation and retrieval; faulty glutamate systems have also been linked to mental illnesses. Coruzzi speculates that the glutamate receptor in the weed could be an ancestral method of communication common to both plants and animals.



Arabidopsis *seedlings in culture*

DAMIEN LOVEGROVE *Science Photo Library*

Bacterial Turn-ons

Some kinds of deadly bacteria—including those that cause tetanus, tuberculosis, syphilis and botulism—remain innocuous until something triggers their insidious activity. Dagmar Ringe of Brandeis University and his co-workers report in *Nature* that they have found the genetic on-off switch for diphtheria, a complex called DtxR. Latched tightly to bacterial DNA, DtxR acts as a repressor; when the host harboring the bacteria experiences an iron deficiency, however, DtxR falls off, allowing the expression of the genes that tell the bacteria to attack the host cells. In principle, a new class of antibiotics could be developed to which bacteria would not become resistant, because the drugs would not kill the bacteria but simply keep them from becoming virulent.

Proton Armageddon

According to physics theories, most everything in the universe decays—including protons. Sooner or later, matter as we know it will cease to exist. The proton's lifetime is still not known, but a new, more stringent lower limit has been found by the Super-Kamiokande underground detector in Japan. The device, which last year found that neutrinos have a slight mass, looked for by-products of proton decay (principally, positrons and pi mesons) but found none. The research team therefore concludes that protons persist for at least 1.6×10^{33} years—far longer, by 100 billion trillion years, than the current age of the universe.

More "In Brief" on page 32

may provide skewed results, warns Maureen A. O'Leary of the State University of New York at Stony Brook. O'Leary's own research demonstrates that leaving the fossil data out of morphology-based analyses yields results similar to those of the molecular biologists, thus calling into question the DNA results. She concedes that "the molecular signal is very strong" but wonders how the molecular results would differ if DNA data from mesonychians were available.

Mesonychian fossils are far too old to contain intact DNA, but researchers thought that finding an ancient whale ankle bone would settle the debate. Artiodactyls are characterized by certain features on one of their ankle bones (the astragalus), which increase mobility. If whales are artiodactyls, primitive whales (those that had not yet adapted to life in the sea) should exhibit these ankle features. In the October 1, 1998, issue of *Nature,* J.G.M. Thewissen, a paleontologist at the Northeastern Ohio Universities College of Medicine, and his colleagues announced their discovery of two ancient whale astragali; intriguingly, the bones do not support either hypothesis.

The fossils are fragmentary, but Thewissen believes that together they provide a complete picture of what either bone would have looked like in its entirety. This composite exhibits a perplexing combination of features: it lacks the rounded head seen in all artiodactyl astragali, but two of its other joint surfaces match a specialized condition found in artiodactyls but not in mesonychians. "Our whale astragalus doesn't look like an artiodactyl," Thewissen observes. "Unfortunately, it also doesn't look like a mesonychian."

Despite the ambiguity of the new fossils, paleontologists hope to recover additional astragali from even older whales, which may be more diagnostic. For now, Thewissen emphasizes that all the data should be considered. He suspects that convergence is confounding the morphological evidence but is impressed with the molecular evidence linking whales and hippos. "Previously I was convinced that whales came out of this mesonychian group," he confesses. "Now I'm on the fence." *—Kate Wong*

SPACE SCIENCE

JOHN GLENN'S EXCELLENT ADVENTURE

Sure, it was a publicity stunt, but science was served, too

The participation of Senator John Glenn of Ohio in shuttle mission STS-95 made it the most ballyhooed space flight since the Apollo moon landings. Millions of television viewers watched the liftoff of the shuttle *Discovery* and avidly followed the progress of the nine-day mission. Glenn even made a guest appearance, via radio link, on the *Tonight Show.* The public was clearly delighted to see the former Mercury astronaut—the first American to orbit the earth—return to space at the age of 77. And the publicity was a much needed shot in the arm for the National Aeronautics and Space Administration, which is now starting work on the controversial International Space Station.

But the stated goal of STS-95 was not publicity; Glenn's primary role was to serve as a guinea pig in a barrage of medical experiments, most of them designed to study the connections between space flight and aging. The results of those tests won't be released for several months, but scientists already know that the studies will not yield any conclusive findings. The problem with the experiments is that they involved just one elderly subject: Glenn himself. To draw reliable conclusions, researchers must be able to compare Glenn's data with tests on other senior citizens in space. But NASA has no plans to send any more septuagenarians into orbit.

The scientists involved in the medical experiments admit that they would have included more subjects if they had had the chance. They maintain, however, that the Glenn studies will prove useful by helping them determine where to focus their future research. "It's a fishing expedition," says Lawrence R. Young, director of the National Space Biomedical Research Institute. "We know there's fish in the pond, but we don't know what we're going to catch." There are intriguing parallels between the symptoms of space flight and aging: both astronauts and the elderly suffer from loss of muscle and bone mass, sleep disturbances and impairment of balance. But researchers have no idea whether the same bodily mechanisms are at work in both cases.

The shuttle experiments involving Glenn were more like a doctor's exami-

In Brief, *continued from page 30*

A Weapon against MS

Positive results are in from the most extensive clinical trial of a drug to treat a form of multiple sclerosis, in which the body's immune system attacks the coatings of nerve cells. The study, which involved more than 500 patients in nine countries, looked at interferon beta 1a. Derived from genetically modified hamster cells, the drug is identical to the human body's own interferon beta, which acts to suppress wayward immune responses. As reported in the November 7, 1998, *Lancet,* the drug reduced relapse rates by up to one third, slowed the progression to disability by 75 percent and decreased brain lesions—all without substantial side effects.

Tag-Team Voting

The Minnesota gubernatorial election of former pro-wrestler Jesse "The Body" Ventura, nemesis of Hulk Hogan, may not have been democratically fair, argues Donald G. Saari, a mathematician at Northwestern University. In the three-way race, Ventura won but did not receive more than half of all votes. Saari says such plurality elections are akin to ranking a student who earned three As and two Fs higher than one who got two As and three Bs. Elections using weighted votes (two for the first choice, one for the second, zero for the third), first proposed by French mathematician Jean-Charles Borda in 1770, can more accurately reflect an electorate's wishes.

AP PHOTO

Governor of Minnesota

Where the Money Goes

The National Science Foundation recently issued a report describing trends in venture-capital spending. In the U.S., such investments reached $9.4 billion in 1996; the biggest recipient was the computer-technology business, which got 32 percent of the funds. Medical/health care and telecommunications companies were other big winners. In Europe, which invested an equivalent of $8.6 billion in 1996, the focus was on industrial equipment, high-fashion clothing and consumer products, which received more than 30 percent of the money; computer-related companies took in only 5 percent. In both the U.S. and Europe, seed money for new firms accounted for only 3 to 6 percent of the total; the bulk, more than 62 percent, went to back company expansions. *—Philip Yam*

SA

SHELLY KATZ *Gamma Liaison*

***JOHN GLENN SUITS UP** at the Kennedy Space Center in preparation for his nine-day shuttle flight.*

nation than a scientific study. The senator wore a cardiovascular monitor during the flight to measure his heart rhythms and blood pressure and a sleep monitor to gauge his brain waves and eye movements while he was slumbering. He also provided blood and urine samples to determine how quickly his bones and muscles were deteriorating in zero gravity. When researchers analyze the data, they will look for unusual results that may justify full-scale studies in space. "We won't get any answers from these experiments," says Andrew Monjan, chief of the neurobiology branch of the National Institute on Aging. "But we may get some interesting questions."

Overlooked in the media frenzy over Glenn's return to space were the more significant scientific accomplishments of the mission. The shuttle crew successfully released and retrieved the Spartan 201 satellite, which provided striking images of the sun's corona. The crew also tested a platform of instruments that will be installed on the Hubble Space Telescope in 2000. In addition, dozens of experiments were conducted in the shuttle's Spacehab laboratory, including a study to determine whether near-perfect crystals of human insulin can be grown in zero gravity.

After the flight, Glenn was a little wobbly on his feet, but after a good night's sleep he said he was back to normal. When the seven crew members returned to Houston—home of the NASA Johnson Space Center—1,000 people gathered at the airport to welcome them. Houston Mayor Lee Brown said the flight had "renewed an American love affair with space travel." The question now is: Will the love last? *—Mark Alpert*

PHYSICS

COSMIC POWER

Superenergetic cosmic rays could reveal the unification of the forces of nature

Meteorites have been called the poor man's space probe—cheap samples of the beyond. In that case, cosmic rays must be the poor man's particle accelerator. A cosmic-ray particle coming from the direction of the constellation Auriga, detected by an instrument in Utah in 1991, had an energy of 3×10^{20} electron volts—more than 100 million times beyond the range of present accelerators. Such natural largesse achieves what purpose-built machines have long sought: a probe of physics underlying the current Standard Model.

For years, people thought the 1991 ray and a few similar ones—registered, for example, by the Akeno Giant Air Shower Array (AGASA) west of Tokyo—might have been flukes. But last summer Masahiro Takeda of the University of Tokyo and the rest of the AGASA team reported five more such events. Roughly one is seen by the array each year, and there is no indication of any limit to their energy.

Current theories say that is impossible. If these cosmic rays are protons or atomic nuclei, as the experiments hint, they must be moving almost at the speed of light. At that clip, the cosmic microwave background, a tenuous gas of primordial radiation that fills space, looks like a thick sea. Particles wading through it lose energy until they fall below 5×10^{19} eV, known as the Greisen-Zatsepin-Kuzmin cutoff. After traveling 150 million light-years, no ordinary particle could still have the observed energies.

Yet astronomers have seen no plausible source within that distance. Exploding stars can propel particles up to only about 1 percent of the required energy. And the mightiest known cosmic slingshots—quasars and active galactic nuclei, the by-products of a massive black hole at lunch—are all too far away, as Jerome W. Elbert and Paul Sommers of the University of Utah showed in 1995. Researchers are forced to one of two equally bizarre conclusions: either the cosmic rays evade the cutoff, or their source is not a normal astronomical object.

In favor of the former, Glennys R. Farrar of New York University and Peter L. Biermann of the Max Planck Institute for Radioastronomy in Bonn recently matched the five most powerful rays with the directions of rare young quasars. The distance of these quasars ranges from four billion to 13 billion light-years. If cosmic rays traverse such lengths, they must be a type of particle that is barely affected by the cosmic microwave background. A neutral and heavier relative of the proton would do the trick. No such stable particle is predicted by the Standard Model, but enhanced theories—drawing on the concept of supersymmetry—do predict one: the so-called S^0 particle.

Another idea, proposed by Thomas J. Weiler of Vanderbilt University, invokes energetic neutrinos that smack into other neutrinos milling about the Milky Way and spill debris particles in the earth's direction. The only requirement is that the neutrinos have a slight mass—which again extends the Standard Model. It is also conceivable that there is no Greisen-Zatsepin-Kuzmin cutoff after all, as Sidney Coleman and Sheldon L. Glashow of Harvard University speculated in August. But if so, special relativity does not apply at high energy.

What if the correlation seen by Farrar and Biermann turns out to be pure chance? Then cosmic rays must emanate from some unidentified celestial phenomenon. The enigmatic sources of gamma-ray bursts might be responsible. More exotic candidates include kinks in the fabric of space and time, such as monopoles and cosmic strings. Tucked within their folds is a sample of the hot early universe in which the forces of nature are unified. As they decay, a miniature big bang ensues, and particles are created with energies up to the unification scale of 10^{25} eV and names like crypton and vorton. The cosmic rays may be these particles or their decay products, as first suggested

ANTI GRAVITY

Taste Matters

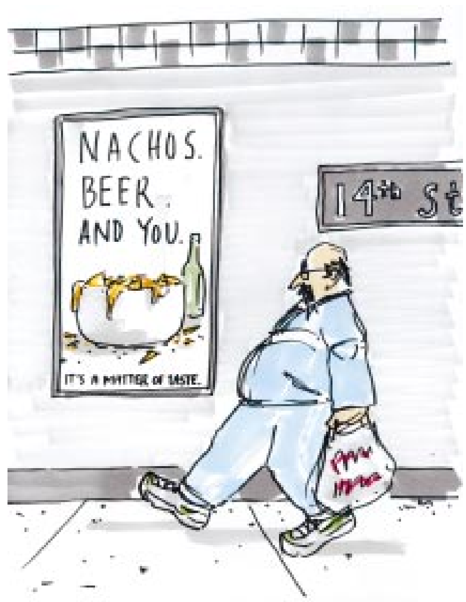
If we are indeed what we eat, then Americans can rest assured that they actually have something that some commentators have often doubted: good taste. According to a study published in the October 1998 *Journal of the American Dietetic Association,* taste is the primary factor that motivates people's choices of what to stick in their pieholes. Previous studies have also revealed that most of us prefer the delectable comestible over the foul-tasting dining experience. As the authors sum up, "People are most likely to consume foods that they evaluate as tasty." I know, I know, you're shocked—shocked! But the study does have a serious message about what we eat and how perhaps to modify those choices better.

The researchers examined four variables in addition to taste—nutrition, cost, convenience and weight-control concerns. They also noted the subjects' other health behaviors, such as exercise patterns, smoking and drinking, and looked at how all those affected food choices. The almost 3,000 subjects were classified according to their overall health profiles. Some of these groupings, seven in all, received alliterative appellations by the resourceful researchers.

For example, one group was labeled the "physical fantastics." They were the most health-oriented individuals, who don't smoke, don't drink much, eat healthfully, exercise routinely and watch their weight. In short, they can still get into the pants they wore in college. Another group, the "active attractives," have some interest in their overall health but mostly because of a concern with their looks. They tend not to smoke, but they do like to experience firsthand the effects of ethanol. They mean to work out, eat right and keep their weight down, but they're not quite doing it. In other words, they still have their college pants, but they're in the bottom drawer. The "decent dolittles" don't smoke or drink, but they don't exercise or eat healthfully. Their college pants have been taken out more than they have. Finally, the "noninterested nihilists" smoke, eat anything and don't exercise. Their college pants can be heard on any staircase at the college.

The bottom line: all the groups rated taste as being the most important factor in food choice. As the authors point out, "Taste, therefore, can be considered a minimal standard for food consumption." The other factors, however, varied widely depending on which group you looked at. Nutrition and weight control were almost as important as taste for physical fantastics but far less important for noninterested nihilists and even active attractives.

With all these data in hand, the authors make what seems to be quite a reasonable suggestion. Health experts are always trying to get people to eat better in this country but do it by harping on the nutritional value to be found in those wholesome foods. Basically, can that idea right along with any vegetables you want to see again in the spring.



"A more promising strategy," they write, "might be to stress the good taste of healthful foods." After all, if Madison Avenue can still figure out ways to convince millions of Americans that smoking is charming, they can probably come up with a plan to make us crave vegetables. Picture the ad campaign: "Brussels sprouts. Not as bad as you remember them." Or maybe: "Broccoli. Not as bad as Brussels sprouts." Well, they might want to start by comparing apples. And oranges. —*Steve Mirsky*

MICHAEL CRAWFORD

in 1987 by Christopher T. Hill and David N. Schramm of the Fermi National Accelerator Laboratory and Terrence P. Walker of Ohio State University.

It is probably not a very good sign that the number of models exceeds the number of data points. "When you have so many speculations," declares James W. Cronin of the University of Chicago, "it shows we really don't understand much at all."

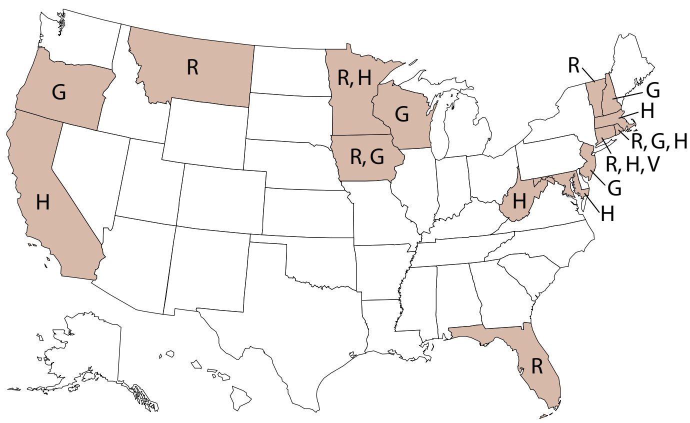
To tilt the balance in favor of data, Cronin and Alan A. Watson of the University of Leeds are heading the Pierre Auger project, an international effort to build two huge cosmic-ray observatories, one south of Salt Lake City and the other near San Rafael, in the wine country of western Argentina. Each will have 50 times the sensitivity of AGASA and should detect rays at a proportionately greater rate. Meanwhile an upgraded version of the Utah experiment—the High Resolution Fly's Eye—should start scanning the skies later this year. Theorists will soon need to be more parsimonious.—*George Musser*

BY THE NUMBERS

Privacy in the Workplace

The U.S. Constitution gives substantial protection to privacy in the home but not where Americans make a living. A 1998 survey of 1,085 corporations conducted by the American Management Association shows that more than 40 percent engaged in some kind of intrusive employee monitoring. Such monitoring includes checking of e-mail, voice mail and telephone conversations; recording of computer keystrokes; and video recording of job performance. Random drug testing is done by 61 percent of those surveyed. Psychological testing, which often attempts to probe intimate thoughts and attitudes, is done by 15 percent of corporations. Genetic testing, which creates the potential for discrimination on a vast scale, is practiced by only 1 percent but, in the absence of a federal law preventing the practice, could become far more widespread if the cost continues to decline.

According to a 1996 survey by David F. Linowes and Ray C. Spencer of the University of Illinois, a quarter of 84 Fortune 500 companies surveyed released confidential employee information to government agencies without a subpoena, and 70 percent gave out the information to credit grantors. Paradoxically, about three fourths of companies barred employees from seeing supervisors' evaluations of their performance, and one fourth forbade them from seeing their own medical records.



R RANDOM DRUG TESTS BANNED

G GENETIC TESTS BANNED

H HONESTY TESTS BANNED

V VIDEO SURVEILLANCE IN LOCKERS, BATHROOMS BANNED

SOURCE: Privacy Journal, *Providence, R. I. Data are the latest available as of late 1998. Some states with random drug-test laws make exceptions for workers in "safety-sensitive" jobs.*

Employers are understandably concerned with raising worker productivity, preventing theft, avoiding legal liability for the actions of employees and preventing corporate espionage. These concerns have largely been given far more weight by the courts than the privacy rights of workers, reflecting the reality that federal laws generally do not give strong protection to workers. One of the few exceptions is the Employee Polygraph Protection Act of 1988, which bars polygraph testing except in certain narrow circumstances. Many scientists consider polygraph testing to be unreliable, yet it has been used as the basis for firing employees.

To make up for federal inadequacy, some states have enacted their own privacy statutes. Federal law takes precedence, but where state laws provide greater protection, employers are usually subject to both. The map shows states that ban various activities, including paper-and-pencil honesty tests, which have not been scientifically validated. No state gives strong privacy protection to workers using e-mail, voice mail or the telephone, nor does any state prohibit intrusive psychological testing. The map illustrates that state laws provide only spotty overall support for worker privacy. Surprisingly, it also shows that worker protection from state laws is weak in the seven states stretching from New York to Missouri, where unions are strongest.

Can the legitimate concerns of employers be reconciled with the privacy concerns of workers? In the early 1990s Senator Paul Simon of Illinois and Representative Pat Williams of Montana attempted to do just that with the Privacy for Consumers and Workers Act. Key provisions require that employers clearly define their privacy policies, refrain from monitoring personal communication, refrain from video monitoring in locker rooms or bathrooms, and notify workers when telephone monitoring is in progress (except for quality control). The act, which represented a compromise by unions, employees and civil-rights organizations, was shelved after the Republicans took over Congress in 1994.

A leading privacy activist, Robert Ellis Smith, publisher of *Privacy Journal,* believes the bill is still worthy of passage but would add more provisions, such as stipulating that employers would have to spell out in advance the reasons for monitoring, discontinue it when the reasons no longer apply and destroy tapes of any innocent employee who was monitored. Linowes and Spencer suggest that any new law regulating data privacy be backed up with the threat of punitive damage awards. Lewis Maltby of the American Civil Liberties Union suggests that unless or until a national workplace privacy law can be passed, corporations try to be less intrusive. For example, they could discontinue video surveillance in locker rooms and bathrooms and end secret monitoring of employees unless there is suspicion of severe misconduct.

—*Rodger Doyle (rdoyle2@aol.com)*

RODGER DOYLE

PROFILE

Flynn's Effect

Intelligence scores are rising, James R. Flynn discovered—but he remains very sure we're not getting any smarter

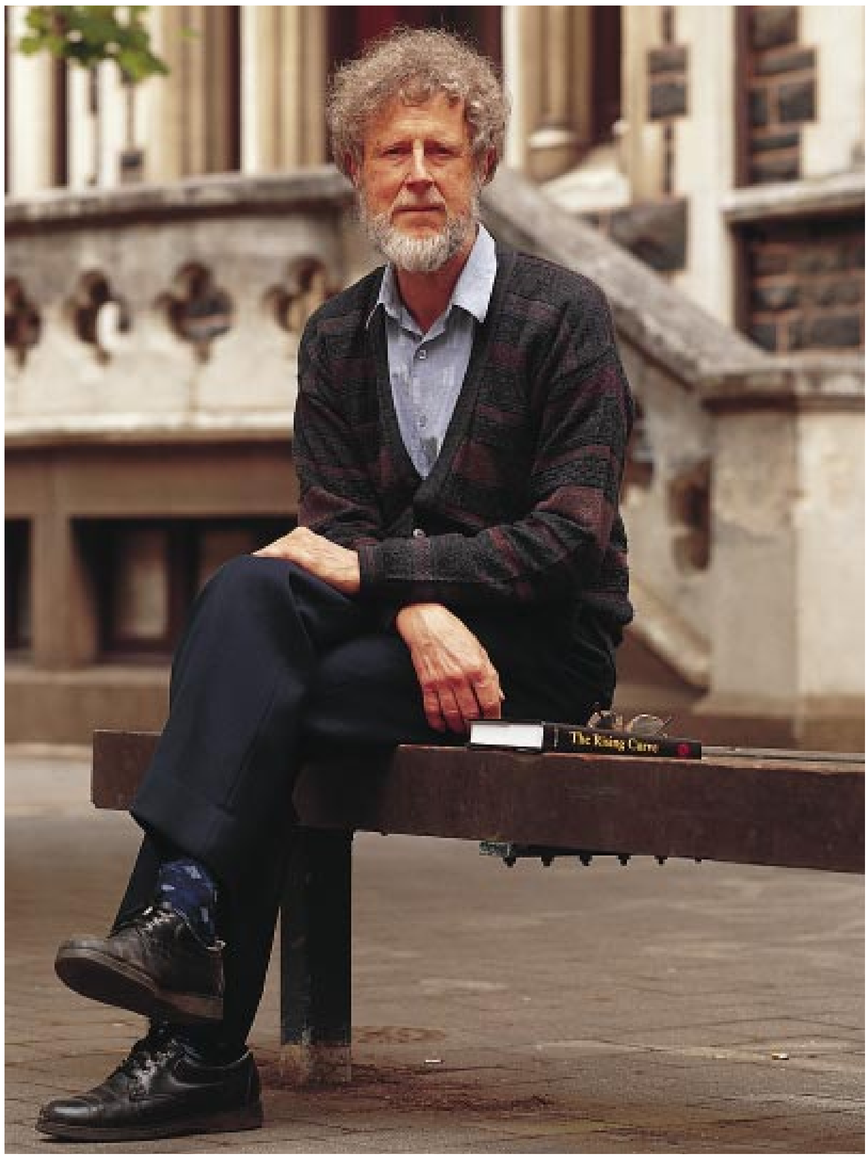
Just back from teaching, James R. Flynn darts into his office to write down a revelation about Marx, free will, Catholicism and the development of the steam engine that came to him in the midst of his lecture. Busily scribbling, the professor of political science at the University of Otago in Dunedin, New Zealand, declares that extemporaneous talking leads to creative thinking and new ideas. His pronouncement made, Flynn—who, it should be noted, talks for a living—is ready to discuss the insight that made him famous: the observation that intelligence quotients, as measured by certain tests, have been steadily growing since the turn of the century.

Flynn's carefully documented findings have provoked a sort of soul-searching among many in the psychological and sociological communities. Before Flynn published his research in the 1980s, IQ tests had their critics. In general, however, the tests were viewed as imperfect yet highly helpful indicators of a person's acuity and various mental abilities or lack thereof. But after the widespread discussion of the eponymous Flynn effect, nothing has been the same. Debates roil about what the tests really measure, what kinds of intelligence there are, whether racial differences persist and, if IQ truly is increasing, why and what the political and social implications are [see "Exploring Intelligence," the winter 1998 issue of Scientific American Presents].

"It is transforming work," comments Ulric Neisser of Cornell University, editor of *The Rising Curve*. The recent book, which emerged from a 1996 American Psychological Association symposium, reviews the Flynn effect and the various explanations for it—including better nutrition and parenting, more extensive schooling, improved test-taking ability, and the impact of the visual and spatial demands that accompany a television-laden, video-game-rich world.

Flynn himself doesn't particularly cotton to any of these explanations. Sitting in his office amid swells of books and papers, he looks very much like a wiry, irreverent Poseidon: gray curls, white beard, pale blue eyes and a kindly, contrary demeanor. A trident poses no challenge to the imagination. If the gains in intelligence are real, "why aren't we undergoing a renaissance unparalleled in human history?" he demands, almost irritably. "I mean, why aren't we duplicating the golden days of Athens or the Italian Renaissance?"

Flynn's own humanist beliefs led him to investigate IQ in the first place. During the 1950s, he was a civil-rights activist in Chicago, where he was political action co-chairman for the university branch of the NAACP while getting his doctorate. After that, he taught at Eastern Kentucky University and chaired the Congress of Racial Equality in Richmond, Ky. "As a moral and political philosopher, my main interest is how you can use reason and evidence against antihumane ideologues," he explains. "Prominent among these are racial ideologues because racism has been one of the chief chal-



JOHN CRAWFORD *Gamma Liaison*

***CIVIL-RIGHTS ACTIVISM LED JAMES R. FLYNN** to discover that IQ scores increase with each generation—a strong argument for environmental factors, rather than genetic ones.*

lenges to egalitarian ideals over the ages."

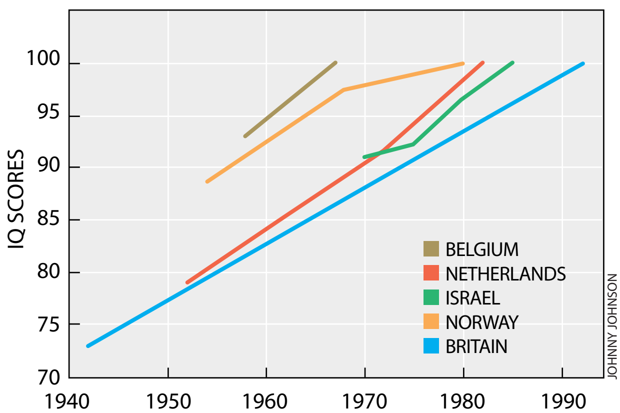
Flynn claims his civil-rights involvement did not prove helpful to a young academic's job search. He and his wife, Emily—whose family had been active in the Communist Party and who, Flynn says, was no stranger to persecution—decided to find a country where they could feel comfortable. They decided on New Zealand: "It seemed to me much more like the sort of social democracy that I would want to live in."

Once they settled into their new home and had started raising their two children, Flynn continued to fight American racism from afar. "I thought that in order to argue effectively with racist ideas, I had to look at the race-IQ debate, the claims that blacks, on average, are genetically inferior." He set out to refute Arthur R. Jensen of the University of California at Berkeley, one of the main proponents of that view. In 1980 Flynn published *Race, IQ and Jensen,* and the duel was on. He decided to follow up with a short monograph on military intelligence tests, because he had a hunch the data had been mishandled and that, in fact, black recruits were making large IQ gains on whites—a trend that would support Flynn's conviction that IQ was linked more to environmental factors than to genetic ones.

Sure enough, Flynn says he found a mistake in the way that some of the military data had been analyzed. But as he investigated further, he realized that Jensen and others would dismiss his findings on the grounds that military intelligence tests were—in contrast to other IQ tests—heavily educationally loaded. In other words, education played a big role in performance. Because black recruits were better educated in the 1950s than they were in the 1920s, any rise in their scores could be attributed to education, not to "real" IQ gains.

Flynn was undeterred. It would be a simple matter, he thought, to find a test measuring "genuine" intelligence that correlated with the military tests, thereby allowing him to use the data from the latter. There was no such correlation to be found, but in the process Flynn unearthed a gold mine. He discovered that certain IQ tests—specifically, the Stanford-Binet and Wechsler series—had new and old versions and that both were sometimes given to the same group of people. In the case of one of the Wechsler tests, for instance, the two versions had been given to the same set of children. The children did much better on the 1949 test than they did on the 1974 one. Everywhere Flynn looked, he noticed that groups performed much more intelligently on older tests. Americans had gained about 13.8 IQ points in 46 years, Flynn reported in 1984.

Although other researchers had noticed different aspects of the phenomenon, they had always explained it away. Flynn did not. "I think the main reason was that since I wasn't a psychologist, I didn't know what had to be true," he muses. "I came as an outsider and didn't have any preconceived notions." (Or, as psychologist Nathan Brody of Wesleyan University points out, there is always the explanation that Flynn, quite simply, "is a very good scholar with a very critical mind.")



WORLDWIDE IQ SCORES
have been rising for more than 50 years.

Critics, including Jensen, responded by saying that the tests must have higher educational loading than previously suspected. So Flynn looked at performance changes in a test called Raven Progressive Matrices, which measures what is called fluid *g:* on-the-spot problem solving that is not educationally or culturally loaded. These tests use patterns instead of, say, mathematics or words. "Polar Eskimos can deal with it," Flynn notes. "Kalahari bushmen can deal with it." Amazingly, it turned out that the highest gains were on the Raven. Flynn observed that in 14 countries—today he has data from at least 20—IQ was growing anywhere from five to 25 points in one generation. "The hypothesis that best fits the results is that IQ tests do not measure intelligence but rather correlate with a weak causal link to intelligence," Flynn wrote when he published the data. "So that was the 1987 article," he says, laughing, "and it, of course, put the cat among the pigeons."

Flynn has recently discovered another dramatic and puzzling increase in the scores of one of the Wechsler subtests—one that measures only verbal ability. Before this new finding, Flynn points out, the explanation that the Raven scores were rising because of video games or computer use had some plausibility. But now, he says, the mystery has only deepened.

Despite two decades of jousting with Jensen, Flynn says he has the deepest regard for the scholar and his scholarship. "There is a temptation on the liberal left not to want to look at the evidence," he remarks. "The fact is that if Arthur Jensen is right, there is a significant truth here about the real world to which we must all adapt." Flynn says he wants humanitarian egalitarian principles to reign "where I have the guts to face up to the facets of the real world. And if one of the facets is that blacks—on average, not individual—are genetically inferior for a kind of intelligence that pays dividends in the computer age, we would do well to know about it."

The next question is, of course, whether he believes such differences exist. In a flash, a sea change: "No! I do not!" Flynn nearly roars.

In addition to his ongoing work on IQ, Flynn has been busy promulgating his ideals on other fronts. Disappointed with New Zealand's slouch toward pure capitalism, he has sought to stem the slide by running for Parliament. He has campaigned, and lost, three times. The most recent and, he adds, final attempt was in 1996 for the Alliance Party: "The only party in New Zealand that still believes in using taxation as a means of redistributing wealth and that still believes in single-payer health and education."

Flynn has also just finished a fifth book, entitled *How to Defend Humane Ideals,* that he has been working on intermittently for many years. "Probably no one will be interested in it because people are much less interested in fundamental contributions than spectacular ones," Flynn rues. It would seem, however, that even merely spectacular results can fundamentally change things. *—Marguerite Holloway*

FIELD NOTES

HERE'S LOOKING AT YOU

A disarming robot starts to act up

Parties have a way of generating outrageous ideas. Most don't survive the night, but a scheme that bubbled to the surface at a 1992 event held by Rodney A. Brooks of the Massachusetts Institute of Technology is changing the way researchers think about thinking. Brooks, the head of M.I.T.'s artificial intelligence laboratory, was celebrating the switch-on date of the fictitious Hal 9000 computer, which appeared in the movie *2001: A Space Odyssey.* As he reflected that no silicon brain could yet rival Hal's slick mendacity, he was seized by the notion of building a humanoid robot based on biological principles, rather than on conventional approaches to robot design.

The robot, known as Cog, started to take shape in the summer of 1993. The project, which was initially to last five years, is intended to reveal problems that emerge in trying to design a humanoid machine and thereby elucidate principles of human cognition. Instead of being programmed with detailed information about its environment and then calculating how to achieve a set goal—the modus operandi of industrial robots—Cog learns about itself and its environment by trial and error. Brooks says that although there are no near-term practical goals for Cog technology, it has stimulated "a bunch" of papers.

Central to the plan was that the robot should (unlike Hal) look and move something like a human being, to encourage people to interact with it. Tufts University philosopher Daniel C. Dennett, an informal adviser to the fluid group of M.I.T. researchers who have worked on Cog, has stated that the machine "will be conscious if we get done all the things we've got written down." Another principle guiding the project was that it should

SAM OGDEN

COG, A HUMANOID ROBOT, ***can turn to stare at moving objects and reach out to touch them. Cog's biologically inspired control systems produce strangely lifelike movements.***

not include a preplanned, or explicit, internal "model" of the world. Rather the changes in Cog as it learns are, in the team's words, "meaningless without interaction with the outside world."

A little after the five-year mark, not even the most enthusiastic fan could argue that Cog is conscious. Yet it is also clear that the exercise of building it has highlighted some intriguing observations.

One day last fall Brian Scassellati and Cynthia Breazeal of Brooks's department exhibited some of Cog's tricks. The machine's upper-torso humanoid form is irresistibly reminiscent of C3PO of *Star Wars* fame. It has learned how to turn to fixate on a moving object, first switching its swiveling eyes, then moving its whole head to catch up. Cog will imitate a nod and reach out to touch things with strikingly lifelike arm movements. The movements have a fluidity not usually associated with machines, because they are driven by a system that has learned to exploit the limbs' natural dynamics.

Cog's mechanical facility is revealed in the way it quickly picks up the timing needed to play with a slinky toy attached to its hands or spontaneously rotates a crank. According to Brooks, a major milestone in Cog's development—that of having multiple systems working together simultaneously—was set to be achieved within the next few months.

Plans are under way to provide the robot with more tactile sensors, a better controlled posture and the ability to distinguish different sound sources. Cog should then be able to associate a voice with a human being in its visual field. There are no plans to add a premade speech-recognition capability, because that would violate the guiding philosophy that Cog should learn on its own.

An expandable stack of high-speed processors gives Cog enough computing power to build on its current skills, Brooks explains. Yet even in its present, simple incarnation, Cog can elicit unexpected behavior from humans. Breazeal once found herself taking turns with Cog passing an eraser between them, a game she had not planned but which the situation seemed to invite.

Breazeal is now studying emotional interactions with a disembodied Cog-type head equipped with expressive mobile eyelids, ears and a jaw. This robot, called Kismet, might yield insights that will expand Cog's mental horizons. Kismet, unlike Cog, has built-in drives for social activity, stimulation and fatigue and can create expressions of happiness, sadness, anger, fear or disgust. Like a baby, it can manipulate a soft-hearted human into providing it with a companionable level of interaction.

It is clear that Cog is still some years from mastering more sophisticated behaviors. Integrating its subbehaviors so they do not compete is a difficulty that has hardly yet been faced. And Cog has no sense of time. Finding a good way to provide one is a "real challenge," Brooks's team writes in a forthcoming publication. Because the design philosophy requires that Cog function like a human, a digital clock is not acceptable.

Cog's development, it seems, will prove slower than that of a human infant. Perhaps just as well: the team has started to consider the complications that might follow from giving Cog a sense of sexual identity. But the effort to make a machine that acts like a human could yet tell researchers a good deal about how a human acts that way.

—Tim Beardsley in Cambridge, Mass.

PUBLIC HEALTH

UNEQUAL HEALTH

The federal government targets disparities in health that result from ethnic background and economic status

The subject's condition is improving, but lingering complications rule out a clean bill of health—such were the findings of the comprehensive study *Health, United States, 1998,* released by the Department of Health and Human Services. The latest data show that in general Americans are healthier than ever: the average life expectancy in the U.S. has reached an all-time high of 76.1 years.

But some people—namely, the poor and certain minority ethnic groups—are still being left behind. For instance, life expectancy for white Americans is 76.8 years, but for black Americans it stands at just 70.2 years. In an effort to remedy the problem, the federal budget for 1999 includes just over $220 million to eliminate inequities in health by 2010. But even as the initiative proceeds, scientists are still wrestling with the reasons behind such disparities—making the prospect of devising solutions that much more difficult.

Inequalities in health are widespread: AIDS fatalities are disproportionately high among Latino and African-American men, for example. Infant mortality is twice as high for blacks as it is for whites. Hepatitis B is much more prominent among Asian-Americans than it is in the rest of the population. And on average, adults with less education have higher death rates from chronic diseases, communicable diseases and injuries than more educated adults do.

The new federal program targets six areas where disparities are particularly pronounced—infant mortality, diabetes, cancer screening and management, heart disease, HIV/AIDS, and immunizations for both children and adults. A total of $156 million will go toward improving HIV/AIDS prevention and treatment programs (particularly toward increasing access to the latest, more expensive drugs) among minority populations. Another $65 million has been set aside to address education, prevention and treatment of all the ailments.

John W. Lynch, a researcher at the University of Michigan School of Public Health, has been investigating how broad societal issues might contribute to such inequalities. Lynch and his colleagues compared mortality rates and income in 282 metropolitan areas across the U.S. "We've known for a long time that absolute income relates to health," Lynch says, referring to the well-documented observation that people with low incomes often have more health problems. But Lynch's team wondered if the connection was that simple—poor people are in poor health—or whether another determining factor was relative income, that is, how a person's financial standing compares with that of others in the community.

Using data supplied by the Federal Office of Management and Budget, the 1990 U.S. Census and the National Center for Health Statistics, the Michigan group discovered that relative income did indeed correlate with the overall health of a particular urban area. "We found that in places with a big gap between rich and poor, the poor are in much worse health than when there are smaller disparities in income," Lynch explains. He speculates that as the income gap grows larger, "there is no incentive for the rich to invest in public health care" or related programs such as public education and housing. And, he adds, "where there is a large income inequality, [there tends to be] a high level of violent crime." Not surprisingly, in many places the lower-income populations consist largely of minority groups. "We can't disentangle racial and economic inequality," Lynch says.

The study also showed that in some of the most economically divided regions—such as Pine Bluff, Ark., and Mobile, Ala.—death rates were much higher than the national annual average of 850 deaths per 100,000 people. The increase in mortality—an extra 140 deaths per 100,000 people—is equivalent to the combined rate of loss of life from lung cancer, diabetes, motor vehicle accidents, HIV, infection, suicide and homicide during 1995.

So in the face of such findings, will the federal government's $220 million really amount to all that much? Gary C. Dennis, president of the National Medical Association (NMA), an organization in Washington, D.C., representing 22,000 African-American physicians and their patients, says the NMA applauds the federal initiative but points to problems—ranging from unhealthy lifestyles common among members of poor and minority groups to their lack of health insurance—that may not receive adequate attention under the current program. Dennis also describes an emerging trend the NMA is following closely: some physicians who treat low-income or minority patients are being cut from the rosters of certain insurance companies. "Their patients tend to be sicker"—and therefore require more aggressive (read expensive) treatments—"so the doctors don't look as cost-effective," Dennis explains. That clean bill of health for the country may be a while in coming. *—Sasha Nemecek*

CYNTHIA JOHNSON *Gamma Liaison*

***HEALTH EDUCATION PROGRAMS TARGETED AT MINORITIES* will be part of a new federal initiative to address inequalities in health. The program focuses on infant mortality, diabetes, HIV/AIDS and other areas.**

SOCIOLOGY

CONSUMING FEARS

In Britain, doubts about science allow food scares to flourish

In recent months Britons have been told they might get the brain-destroying Creutzfeldt-Jakob disease from eating sheep, a bowel disorder called Crohn's disease from drinking pasteurized milk and a damaged immune system from dining on genetically modified foods. Consumer groups, newspapers and broadcasters have trumpeted the scares as though lives were at stake. Yet in the first two cases, the Department of Health described the risk as negligible, and the genetic crop worry last August was later admitted to be bogus—a scientist had muddled the results of a colleague's research, confusing rats from two different experiments.

These incidents were only the latest in about 15 years of food scares in Britain—including salmonella in eggs; listeria in cheese; *Escherichia coli,* antibiotics and hormones in meat; and pesticide residues and phthalates (benzene-related compounds) in just about everything. And of course, most infamous was the scare about beef from cows infected with bovine spongiform encephalopathy (BSE). Besides creating panic, food scares can wreak havoc with the agricultural economy—sales of beef have only recently returned to their pre-BSE levels.

Whereas genuine outbreaks of food poisoning are not uncommon, the reactions in Britain seem particularly out of line with the threat. A large part of that, notes food-safety expert Derek Burke, stems from the handling of the BSE outbreak. The ongoing inquiry has caused the complete collapse of public faith in food-regulating authorities, such as the Ministry of Agriculture Fisheries and Food (MAFF) and the Department of Health, as well as in politicians and scientists.

For instance, MAFF admitted that it knew in 1986 that prions, unusual proteins that are thought to cause BSE, might be able to infect humans and cause Creutzfeldt-Jakob. Not until 1989, however, did it introduce legislation to ban specifically high-risk material—brains and spleens—and only last year did it ban the material from use in pharmaceuticals and cosmetics. More recently, press reports last September indicated that MAFF turned a blind eye to abattoirs that flouted BSE safety requirements. "It is going to take years to get rid of that problem" of public mistrust, says Burke, who served as chairman of the government's Advisory Committee on Novel Foods and Food Processing.

Lynn Frewer agrees. She is head of the risk perception and communications group at the Institute of Food Research, which works for, among others, MAFF and the European Union in multilateral research programs. "Fifty years ago science was equated with progress. It was trusted and seen as properly regulated. But in the past 50 years there have been many symbols of it getting out of control, such as DDT, thalidomide and, more recently, BSE," she concludes.

Frewer adds another reason for the escalating concerns about foods. Many once-feared illnesses, such as polio, smallpox and scarlet fever, are preventable or curable now. That has prompted people to magnify other worries instead. Burke quips that there would be fewer food scares if war broke out.

Although questions of food safety occur in the U.S., they do not cause as much panic. Americans hold a less equivocal attitude toward science than Britons and other Europeans do. "The U.S. has never admitted it has any problems, so the FDA is still widely trusted. But I do not think its processes are intrinsically any better than [those of] the British," says Burke, who has lived and worked in the U.S.

That might explain why most Americans are not too bothered by genetically modified foods. The crops—mostly corn, potatoes and soybeans—are designed to produce their own insecticide or to withstand herbicides and can turn up anonymously in such prepared products as french fries. Because no evidence has been found that genetically modified foods are dangerous, the FDA does not require any special labeling for them.

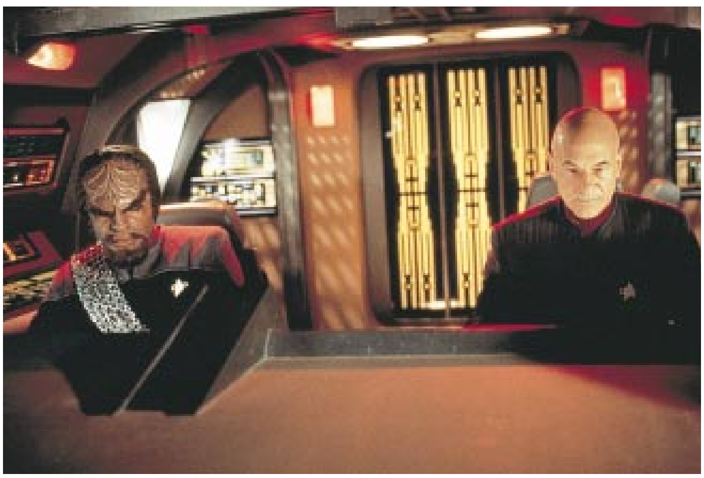
Britain and most of Europe, however, feel differently—after all, many argue, there is no evidence they are safe over the long term, either. Moreover, transgenic crops can lead to unpredictable environmental consequences: a maize trial, for instance, ended up killing off lacewings, which are beneficial crop insects. Prince Charles of Wales summed up the public mood last June by saying in the *Daily Telegraph* that the work done by genetic engineers was best left to "God and God alone."

BOLDLY GOING

Brace for (Educational) Impact

That some real science lingo can be picked up from *Star Trek* comes as no surprise. For more than 30 years on television and through nine feature films—the latest of which, *Star Trek: Insurrection,* opened December 11—the franchise has bandied about such abstract concepts as space warps and quantum singularities so often that they are taken for granted. And with closed captioning, a viewer could even learn some Klingon (*ghargh* is food best served wriggling, and *Qapla'*—success!—is a friendly sign-off). A 1994 report from Purdue University found that students overwhelmingly considered *Star Trek* to be the most influential promoter of their interest in science. But is the quality of the information presented something to worry about?

Apparently not. Although several books have dissected *Star Trek* science (teleportation as energetically impossible, for instance), overall "a lot of the science has been pretty good science," opines Terence W. Cavanaugh, who explores teaching methods at the University of South Florida. As such, it can make for an effective instructional tool.



ELLIOTT MARKS *Paramount Pictures*

PROFESSORS WORF AND PICARD
***help to teach science—in this case, from* Star Trek: Insurrection.**

In a study last year of seventh-grade schoolchildren, Cavanaugh found that *Star Trek* videos proved superior to traditional educational films as a way to teach science, largely because, he says, students "had a better attitude" while watching, for example, the *Enterprise* crew explain in one case the chemistry of life while fending off attacks from soil-dwelling, non-carbon-based organisms. The key was active watching—stopping the videotape and discussing the concepts; in this way, a single episode might take up to three days to view. The study shows, Cavanaugh concludes, that teachers have an alternative to standard educational films, which generally cost several times more than a videotape does and are usually harder to obtain.

For scientific accuracy, the two current *Star Trek* series (and the last two feature films) rely on consultant Andre Bormanis. "The writers are pretty knowledgeable about the basics," says Bormanis, who studied physics and computer science. (He landed his consultancy in 1993, after writing a screenplay that his agent took to a story-pitch meeting.) "If they put something in that's wrong, they will fix it if I give them an alternative that's viable."

Of course, Bormanis gets his share of "gotcha!" letters. He and the writers try not to repeat errors, although problematic concepts are sometimes maintained for the sake of story consistency. "*Star Trek* isn't noble. It's not our principal mission to teach science per se," Bormanis observes. "We certainly want to promote science and represent it in a credible fashion."

Paramount Pictures, *Star Trek*'s corporate parent, has taken science promotion more seriously of late. Aside from special media events, it is backing, with the Planetary Society, a nonprofit organization based in Pasadena, Calif., a project called SETI@home. This venture invites the public to join in the search for extraterrestrial intelligence. Starting in April, interested parties will be able to download a special screensaver for their personal computers (from either www.planetary.org or www.startrek.com). It contains data collected by the radio antenna at Arecibo Observatory in Puerto Rico. During idle periods of the user's machine, the screensaver would comb through the data, looking for signals that might be artificial.

Although genuine concepts undergird much of the *Star Trek* universe, it's probably best that the series serve as a conduit for teaching science, not as a source of it. For young minds learning about the natural world, Cavanaugh notes, "fun and interesting can't hurt in the long run." *Qapla'!* —*Philip Yam*

DAVID JONES AP Photo

ALTERNATIVES TO BEEF AFTER THE OUTBREAK OF MAD COW DISEASE were offered to Avon Valley School in Rugby, U.K., in 1996. Fears about the safety of meat, dairy products, vegetables and other foods are sweeping Britain.

Perhaps not surprisingly, then, the actions of ecoterrorists, who have destroyed at least 30 of more than 300 crop trials in the past few years, take place in a blaze of admiring publicity. Prosecutions are rare for fear of copycat action and adverse press. The chemical giant Monsanto, which has trials all over Britain, has requested injunctions against activist groups such as Earth First! and Genetix Snowball.

The reactions to the possible hazards of food, real or imagined, have raised questions about exactly what the public should be told and when. Both the scientific community and consumer groups agree that the current ad hoc system of reporting food concerns is inadequate. Real problems could be overlooked. For instance, a report from the Food Commission, a British lobbying group, states that some nut imports are contaminated with deadly aflatoxins, a potent liver carcinogen. MAFF has admitted the problem, but the finding has gone almost unnoticed and unreported. Tim Lobstein, co-director of the commission, pins the blame on the news media.

To streamline food regulation and the reporting of threats, the government wants to establish a food standards agency. Exactly who pays for this agency and whether the bill authorizing its creation is passed in the next legislative session, however, are still up in the air. It may be a while before Britons look at their dinner plates without apprehension once more. —*Peta Firth*

PETA FIRTH, who was an award-winning journalist for the Hong Kong daily newspaper the HK Standard, *is a freelance writer based in London.*

Out of Site

Information security is the new catchphrase at the Pentagon. Every day the U.S. military grows more reliant on information technology for everything from bookkeeping to battlefield communications, and every year it spends more to guard against hackers, terrorists and other enemies. Information security is among the few areas of the defense budget that is guaranteed to grow in coming years; like missile defense, it shares congressional, Pentagon and defense industry support, and the media has begun to pay a great deal of attention to hackers who penetrate military World Wide Web sites and other threats.

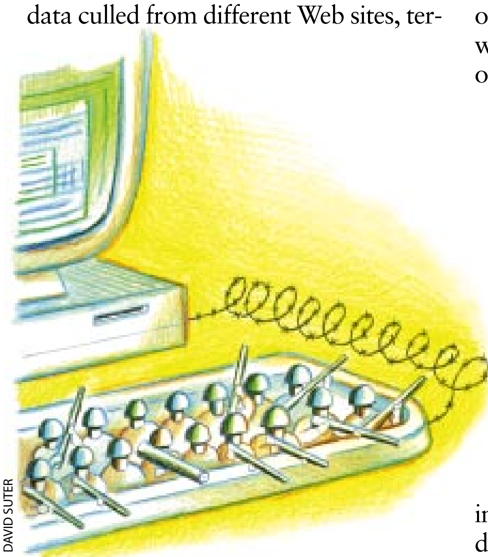
But George Smith, who edits the *Crypt Newsletter,* an on-line publication covering information warfare and security issues, suggests that the government may be overstating the threat. "It is far from proved that the country is at the mercy of possible devastating computerized attacks," he wrote in last fall's *Issues in Science & Technology*. Although threats exist, he contends that the "time and effort expended in dreaming up potentially catastrophic information warfare scenarios could be better spent implementing consistent and widespread policies and practices in basic computer security."

On the surface, it seems that the Pentagon is doing just that. Earlier this year a group of senior Defense Department officials gathered together in the recesses of the Pentagon to review what lower-level officers said was a grave and growing threat—one the military had helped create. The leaders were shown how, with a modicum of surfing, one could gather from the Internet unlisted telephone numbers and other personal information about top officials, in addition to other "sensitive" data that might interest potential saboteurs. It was apparently an effective presentation. On September 24, Deputy Defense Secretary John Hamre issued a directive instructing all military services and agencies to "ensure national security is not compromised or personnel placed at risk" by information on Pentagon sites. If that sounds perfectly reasonable, it is. For years the military has had policies in place that demand no less. So why the new emphasis on security?

It's hard to fathom. In six years as a Pentagon reporter I've seen dozens of classified documents, but none of them were found on the Internet. In fact, many of the scores of military Web sites I frequent are months out-of-date and utterly lacking in anything that could compromise national security.

The Internet, Hamre points out, provides an excellent way to inform the public about what the Pentagon is doing, which he says is "fundamental to the American democratic process." The trick, he adds, is to guarantee that military sites are "carefully balanced against the potential security and privacy risks created by having aggregated DOD data more readily accessible to a worldwide audience."

The key words here are "aggregated" and "more readily accessible." Pentagon leaders lately have come to believe that by adding together bits of unclassified data culled from different Web sites, terrorists could discover vulnerabilities that would otherwise remain unexposed.



DAVID SUTER

But because the Pentagon has always demanded that only unclassified information be made available on its Web sites, the new security guidelines have essentially fostered a novel kind of military secrecy. It has decreed that, in some cases, unclassified information available to the public on paper cannot be made public electronically. "It's as though they are making a new category of information that is sufficiently boring to be passed out in hard copy but is too sensitive to be available in soft copy," says John Pike of the Federation of American Scientists.

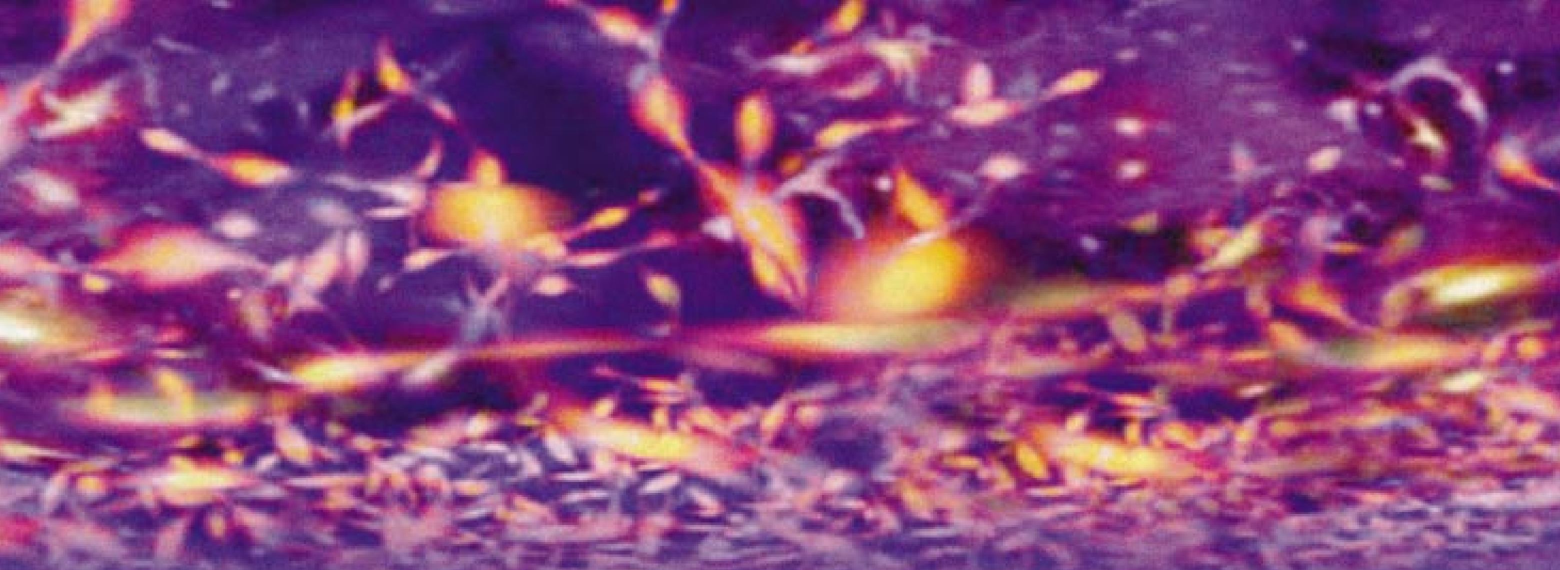
William M. Arkin, an army veteran, defense analyst and editor of *U.S. Military Online,* thinks the Pentagon has grossly misjudged both the power of the Internet and the military's ability to control it. He contends that there is "too much information about all of us floating in the public domain" but adds that much of it can be found on commercial, not military, sites.

Arkin and Pike probably spend more time on military Web sites than anyone else; both say they rarely come across much worth worrying about. Both believe personal information doesn't belong on military Web sites, but they are concerned that the new security guidelines are so broad that information with "public policy significance" will be kept off the Web. "It's easy to be risk-averse to the point of being uncommunicative," Arkin says. Pike concurs; he declares the Pentagon issued its new policy not because of any new threats but out of "a desire to show vigilance, coupled with a profound lack of understanding of information and computer security."

Pike and Arkin are especially critical of the army for its reaction to the new policy. Hours after the directive was issued, the army—alone among Pentagon components—shut down all of its nearly 1,000 Web sites without warning or explanation. Many were back within a few days, but six weeks later a significant number of sites still weren't available. Of those that were, some were stripped of data that in no way seemed to meet the Pentagon's definition of sensitive information.

Of course, people like Pike and Arkin—and defense reporters—have always disagreed with the Pentagon about how much information should be made public. And the Internet has become a crucial tool for journalists everywhere—we're never pleased when access is restricted. But we'll survive. A few years ago I needed to get a comment on a hot story from a four-star member of the Joint Chiefs of Staff, and I needed it fast. Because I was on deadline and it was after normal business hours, I was forced to call him at home. So to get his telephone number, I turned to the most sophisticated, publicly accessible source I had. I dialed information. —*Daniel G. Dupont*

DANIEL G. DUPONT edits Inside the Pentagon, *an independent newsletter, and* Inside Defense, *an on-line news service.*



SPECIAL REPORT

REVOLUTION IN COSMOLOGY

At a conference last May entitled "The Missing Energy in the Universe," cosmologists took a vote. Did they believe recent observations of distant exploding stars, which implied—in defiance of all expectations—that the universe is growing at an ever faster rate? Although astronomers have known since the 1920s that the universe is enlarging in size, pushing galaxies ever farther apart, they have always assumed that this expansion is mellowing out as gravity exerts its force of restraint. If in fact the growth is accelerating, the universe must be filled with some unknown form of matter or energy whose gravity repels rather than attracts. Hitherto unseen energy is, well, a repulsive thought for physicists. And yet of the 60 researchers present for the vote, 40 said they accepted the new findings.

Astronomers had suspected for more than a decade that all was not well in the halls of modern cosmology. When observers totted up the ordinary matter in the universe, it fell short of the amount needed to slow the cosmic expansion as predicted by the theory of inflation, an elegant explanation of the earliest stages of the big bang. Until now, the evidence against the theory has never been strong enough to overcome its advantages. But today even the theorists accept that something is amiss. At the very least, the expansion is not decelerating as rapidly as once thought. Either scientists must reconcile themselves to kooky energy, or they must modify or abandon inflation.

In this issue, SCIENTIFIC AMERICAN presents three sides of the story. First, three observers relate how and why their work on supernovae has caused such commotion. Then a theorist explains why these results attest to an ethereal energy that threads empty space. Finally, a pair of cosmologists offer another interpretation that extends the theory of inflation to times "before" the big bang. *—The Editors*

Surveying Space-time with

Exploding stars seen across immense distances show that the cosmic expansion may be accelerating—a sign that the universe may be driven apart by an exotic new form of energy

by Craig J. Hogan, Robert P. Kirshner and Nicholas B. Suntzeff

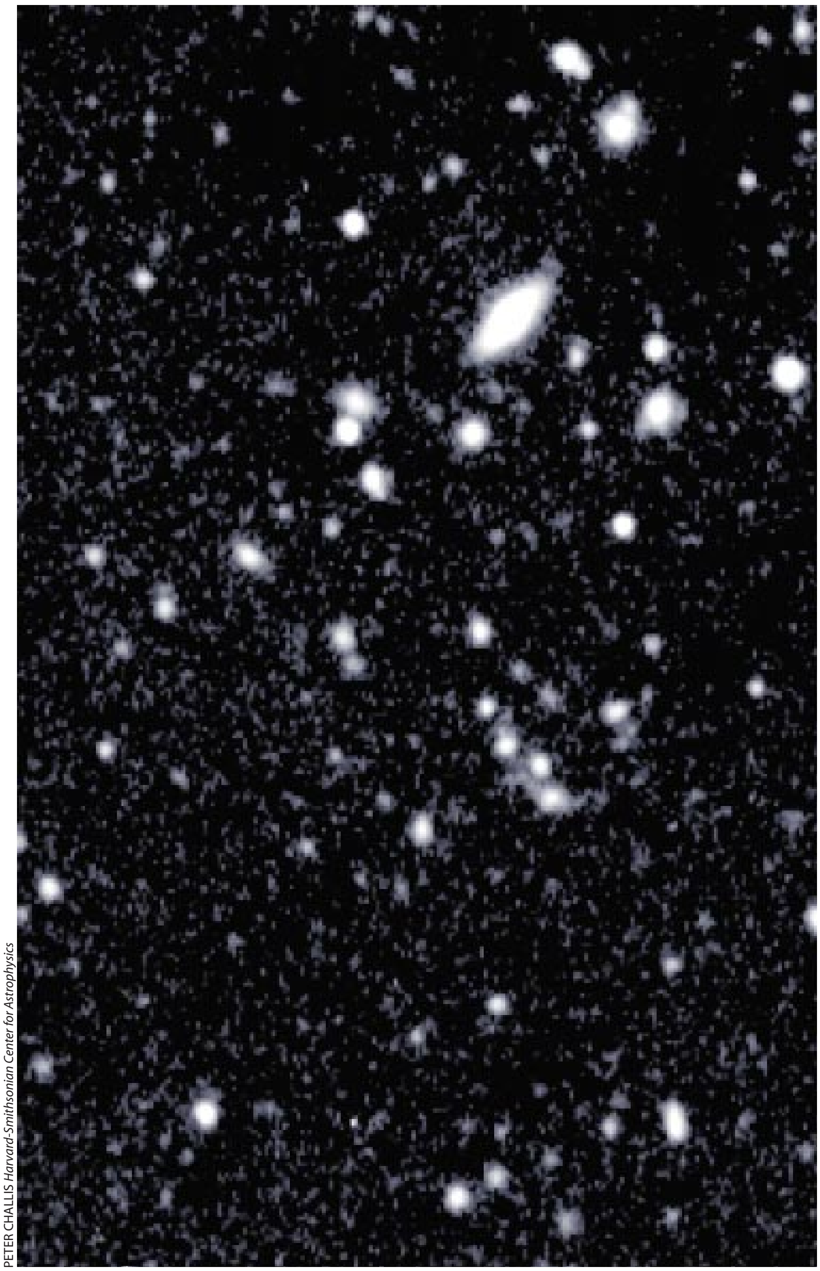
A long time ago (some five billion years), in a galaxy far, far away (about 2,000 megaparsecs), a long-dead star exploded with a flash brighter than a billion suns. Its light spread out across space, fading and stretching with the expanding cosmos, before some of it finally reached the earth. Within 10 minutes during one dark night in 1997, a few hundred photons from this supernova landed on the mirror of a telescope in Chile. A computer at the observatory then created a digital image that showed the arrival of this tiny blip of light. Though not very impressive to look at, for us this faint spot was a thrilling sight—a new beacon for surveying space and time.

We and our colleagues around the world have tracked the arrival of light from several dozen such supernovae and used these observations to map the overall shape of the universe and to chronicle its expansion. What we and another team of astronomers have recently discerned challenges decades of conventional wisdom: it seems the universe is bigger and emptier than suspected. Moreover, its ongoing expansion is not slowing down as much as many cosmologists had anticipated; in fact, it may be speeding up.

Star Warps

The history of cosmic expansion has been of keen interest for most of this century, because it reflects on both the geometry of the universe and the nature of its constituents—matter, light and possibly other, more subtle forms of energy. Albert Einstein's general theory of relativity knits together these fundamental properties of the universe and describes how they affect the motion of matter and the propagation of light, thereby offering predictions for concrete things that astronomers can actually measure.

Before the publication of Einstein's theory in 1916 and the first observations of cosmic expansion during the following decade, most scientists

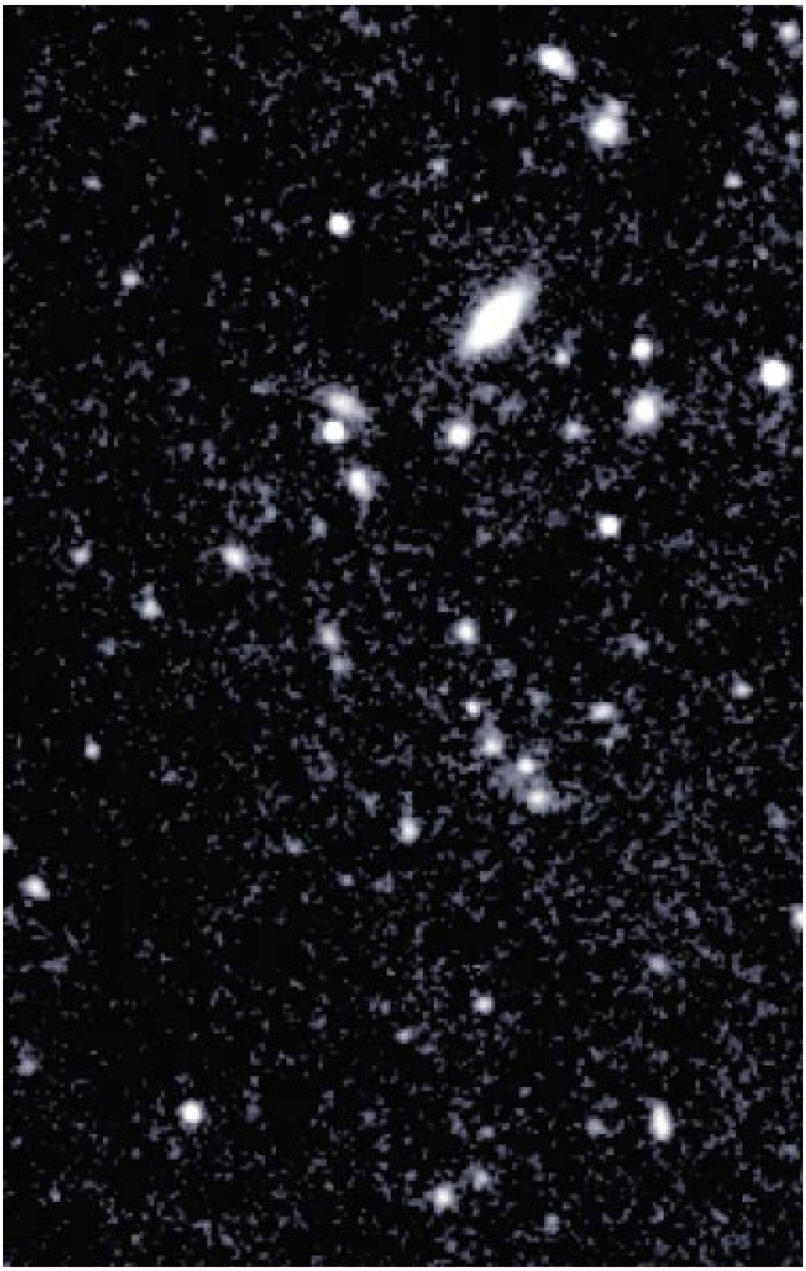


PETER CHALLIS *Harvard-Smithsonian Center for Astrophysics*

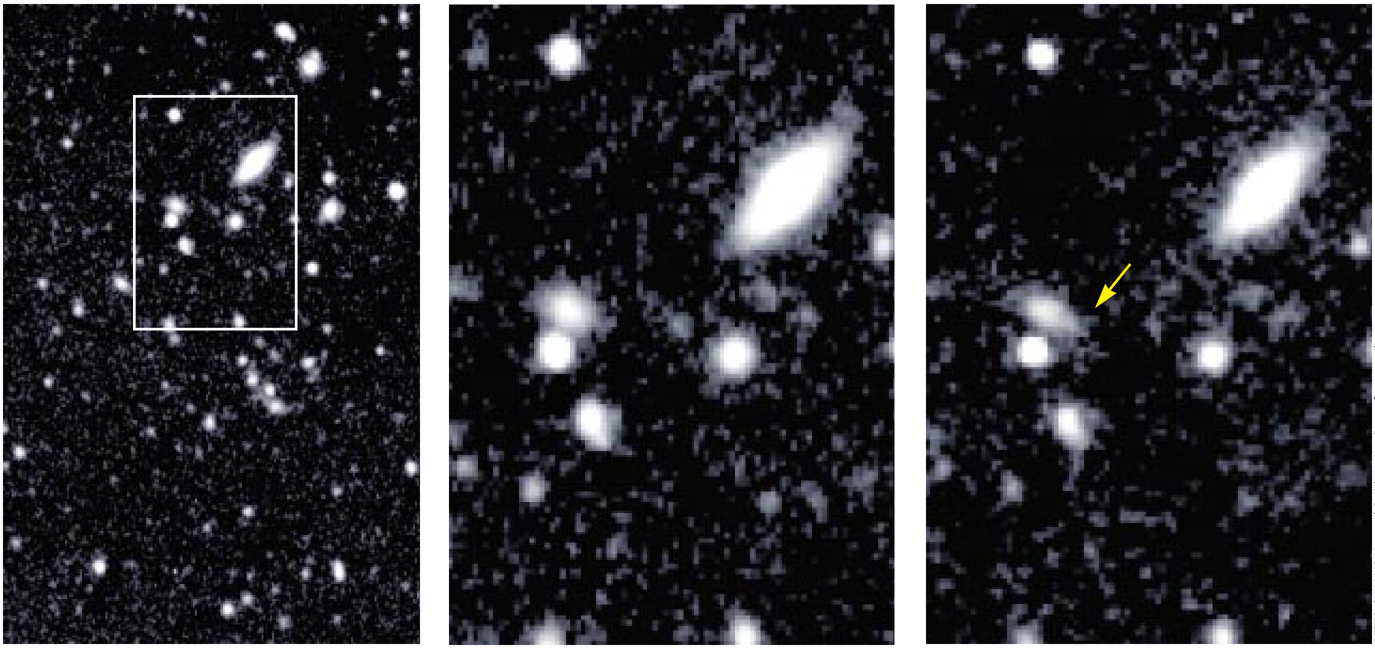
Supernovae

thought the universe stayed the same size. Indeed, Einstein himself distrusted his equations when he realized they implied a dynamic universe. But new measurements of galactic motions by Edwin P. Hubble and others left no doubt: faint, distant galaxies were flying away from the earth faster than bright, nearby ones, matching the predictions of general relativity for a universe that grows and carries galaxies farther apart. These researchers determined the outward velocities of galaxies from the shift of visible spectral lines to longer wavelengths (so-called redshifts). Though often ascribed to the Doppler effect—the phenomenon responsible for changing the pitch of a passing train whistle or car horn—the cosmological redshift is more correctly thought of as a result of the ongoing expansion of the universe, which stretches the wavelength of light passing between galaxies. Emissions from more distant objects, having traveled for a greater time, become more redshifted than radiation from nearer sources.

The technology of Hubble's day limited the initial probing of cosmic expansion to galaxies that were comparatively close. In the time it took light from these nearby galaxies to reach the earth, the universe had expanded by only a small fraction of its overall size. For such modest changes, redshift is directly proportional to distance; the fixed ratio of the two is called Hubble's constant and denotes the current rate of cosmic expansion. But astronomers have long expected that galaxies farther away would depart from this simple relation between redshift and distance, either because the pace of expansion has changed over time or because the intervening space is warped. Measuring this effect thus constitutes an important goal for cosmologists—but it is a difficult one, for it requires the means to determine the distances to galaxies situated tremendously far away.



WHERE'S THE SUPERNOVA? This pair of images, made by the authors' team using the four-meter-diameter Blanco Telescope at Cerro Tololo Inter-American Observatory in Chile, provided first evidence of one supernova. In the image at the right, obtained three weeks after the one at the left, the supernova visibly (but subtly) alters the appearance of one of the galaxies. Can you find it? Some differences are caused by varying atmospheric conditions. To check your identification, consult the key on the next page.



PETER CHALLIS *Harvard-Smithsonian Center for Astrophysics*

DISTANT SUPERNOVA, with a redshift of $z = 0.66$, appears by the arrow. The explosion of this star affects just a few picture elements in the image taken after the event.

Hubble and other pioneers estimated distances to various galaxies by assuming that they all had the same intrinsic brightness. According to their logic, ones that appeared bright were comparatively close and the ones that appeared dim were far away. But this methodology works only crudely, because galaxies differ in their properties. And it fails entirely for distant sources whose light takes so long to reach the earth that it reveals the faraway galaxies as they were billions of years ago (that is, in their youth), because their intrinsic brightness could have been quite different from that of more mature galaxies seen closer to home. It is difficult to disentangle these evolutionary changes from the effects of the expansion, so astronomers have long sought other "standard candles" whose intrinsic brightness is better known.

To be visible billions of light-years away, these beacons must be very bright. During the early 1970s, some cosmic surveyors tried using quasars, which are immensely energetic sources (probably powered by black holes swallowing stars and gas). But the quasars they studied proved even more diverse than galaxies and thus were of little use.

About the same time, other astronomers began exploring the idea of using supernovae—exploding stars—as standard candles for cosmological studies. That approach was controversial because supernovae, too, show wide variation in their properties. But in the past decade research by members of our team has enabled scientists to determine the intrinsic brightness of one kind of supernova—type Ia—quite precisely.

Death Star

What is a type Ia supernova? Essentially, it is the blast that occurs when a dead star becomes a natural thermonuclear bomb. Spectacular as this final transformation is, the progenitor begins its life as an ordinary star, a stable ball of gas whose outer layers are held up by heat from steady nuclear reactions in its core, which convert hydrogen to helium, carbon, oxygen, neon and other elements. When the star dies, the nuclear ashes coalesce into a glowing ember, compressed by gravity to the size of the earth and a million times the density of ordinary matter.

Most such white dwarf stars simply cool and fade away, dying with a whimper. But if one is orbiting near another star, it can slurp up material from its companion and become denser and denser until a runaway thermonuclear firestorm ignites. The nuclear cataclysm blows the dwarf star completely apart, spewing out material at about 10,000 kilometers per second. The glow of this expanding fireball takes about three weeks to reach its maximum brightness and then declines over a period of months.

These supernovae vary slightly in their brilliance, but there is a pattern: bigger, brighter explosions last somewhat longer than fainter ones. So by monitoring how long they last, astronomers can correct for the differences and deduce their inherent brightness to within 12 percent. Over the past decade studies of nearby type Ia supernovae with modern detectors have made these flashes the best calibrated standard candles known to astronomers.

One of these candles lights up somewhere in a typical galaxy about once every 300 years. Although such stellar explosions in our own Milky Way are rare celestial events, if you monitor a few thousand other galaxies, you can expect that about one type Ia supernova will appear every month. Indeed, there are so many galaxies in the universe that, somewhere in the sky, supernovae bright enough to study erupt every few seconds. All astronomers have to do is find them and study them carefully. For the past few years, that effort has occupied both our research group, dubbed the "High-Z Team" (for the letter that astronomers use to denote redshift), a loose affiliation organized in 1995 by Brian P. Schmidt of Mount Stromlo and Siding Spring Observatories in Australia, and a competing collaboration called the Supernova Cosmology Project, which began in 1988 and is led by Saul Perlmutter of Lawrence Berkeley National Laboratory.

Although the two teams have independent programs, they are exploiting the same fundamental advance: the deployment of large electronic light detectors on giant telescopes, a combination that produces digital images of faint objects over sizable swaths of the sky. A prime example of this new technology (one that has served both teams) is the Big Throughput Camera, which was developed by Gary M. Bernstein of the University of Michigan and J. Anthony Tyson of Lucent Technologies. When this camera is placed at the focus of the four-meter Blanco Telescope at Cerro Tololo Inter-American Observatory in Chile, a single exposure covers an area about as big as the full moon and creates a picture of about 5,000 galaxies in 10 minutes.

Finding distant supernovae is just a matter of taking images of the same part of the sky a few weeks apart and searching for changes that might be exploding stars. Because the digital light detectors can count the number of photons in each picture element precisely, we simply subtract the first image from the second and look for significant differences from zero. Because we are checking thousands of galaxies in each image pair, we can be confident that the search of multiple pairs will find many supernovae—as long as the weather is good. Fortunately, the location of the observatory, in the foothills of the Andes on the southern fringe of Chile's Atacama Desert (one of the driest places in the world), usually provides clear skies. Betting that we will make some good discoveries, we schedule observing time in advance on a battery of other telescopes around the world so that follow-up measurements can start before the supernovae fade away.

In practice, the search for exploding stars in the heavens whips up its own burst of activity on the ground, because we must acquire and compare hundreds of large, digital images at a breakneck pace. We commandeer computers scattered throughout the Cerro Tololo observatory for the tasks of aligning the images, correcting for differences in atmospheric transparency and image size, and subtracting the two scans. If all goes well, most of the galaxies disappear, leaving just a little visual "noise" in the difference of the two images. Larger signals indicate some new or changing object, such as variable stars, quasars, asteroids—and in a few cases, supernovae.

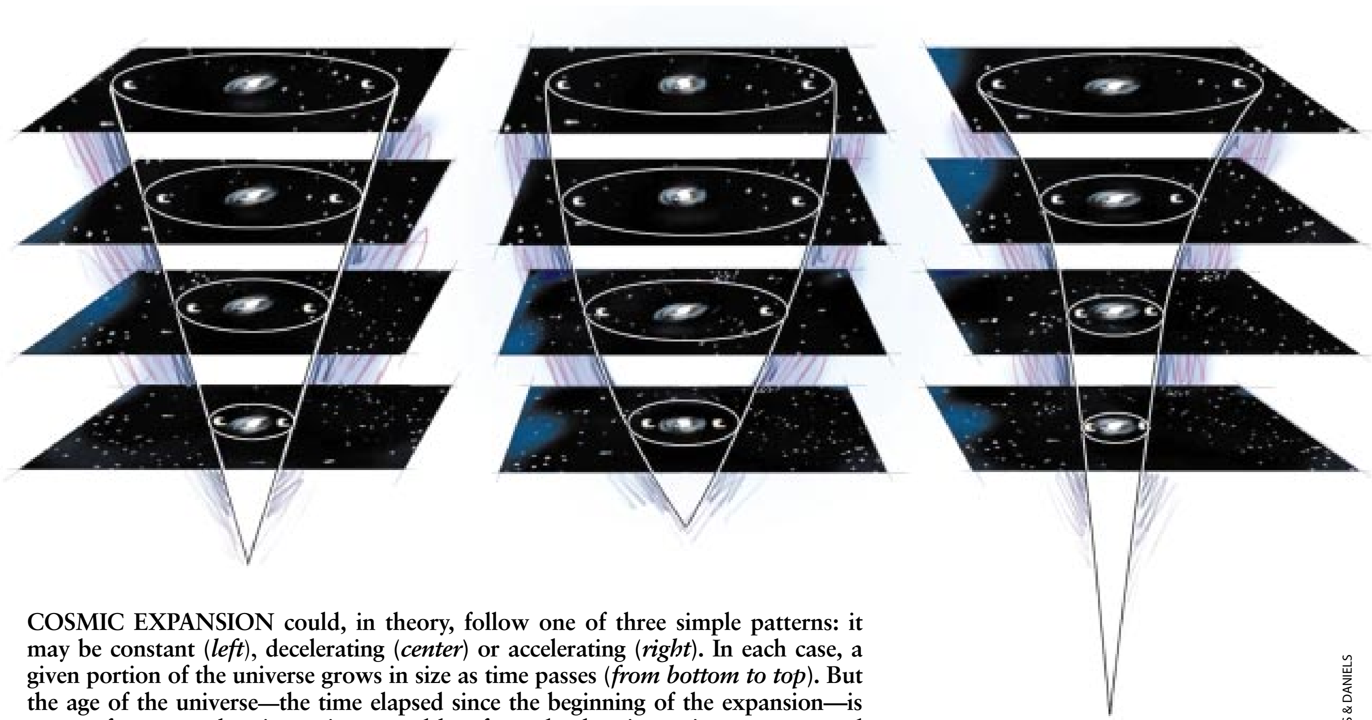
Our software records the position of new objects and attempts to identify which are truly supernovae. But the automated tests are imperfect, and we must scrutinize the images by eye to decide whether a putative supernova is real. Because we must immediately pursue our discoveries with other telescopes, the analysis must be done quickly. During these exhausting times, the observatory becomes a sweatshop of astronomers and visiting students, who work around the clock for days at a stretch, sustained by enthusiasm and Chilean pizza.

We next target the best supernova candidates with the largest optical instruments in the world, the recently constructed Keck telescopes in Hawaii. These critical observations establish whether or not the objects discovered are in fact type Ia supernovae, gauge their intrinsic brightness more exactly and determine their redshifts.

On the Dark Side

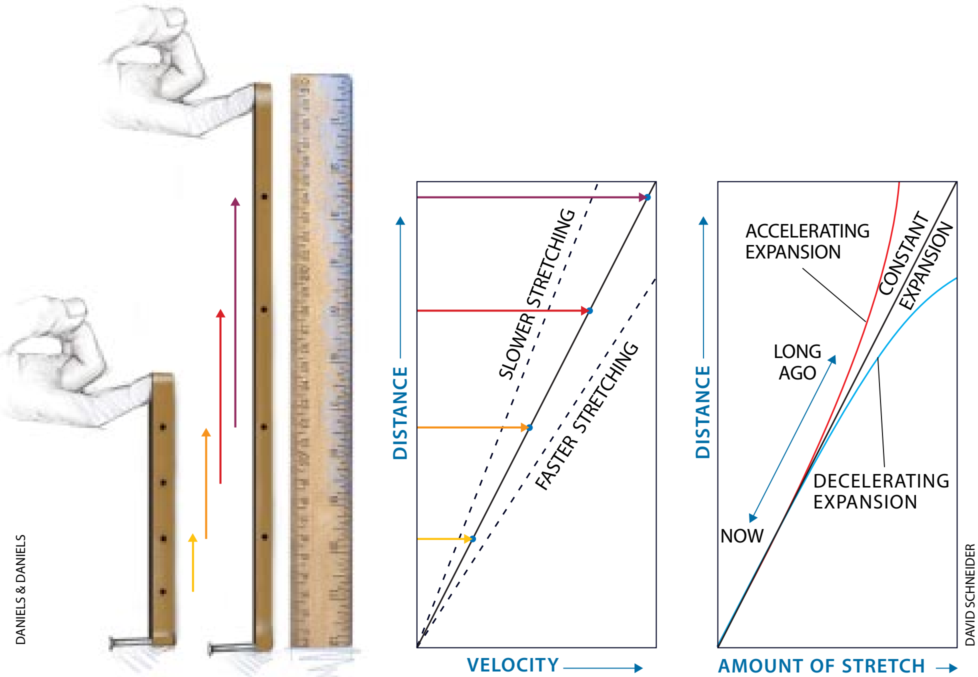
Others in our group, working with telescopes in Australia, Chile and the U.S., also follow these supernovae to track how their brilliance peaks and then slowly fades. The observing campaign for a single supernova spans months, and the final analysis often has to wait a year or more, when the light of the exploded star has all but disappeared, so we can obtain a good image of its host galaxy. We use this final view to subtract the constant glow of the galaxy from the images of the supernova. Our best measurements come from the Hubble Space Telescope, which captures such fine details that the exploding star stands out distinctly from its host galaxy.

The two teams have now studied a total of a few score high-



COSMIC EXPANSION could, in theory, follow one of three simple patterns: it may be constant (*left*), decelerating (*center*) or accelerating (*right*). In each case, a given portion of the universe grows in size as time passes (*from bottom to top*). But the age of the universe—the time elapsed since the beginning of the expansion—is greater for an accelerating universe and less for a decelerating universe, compared with the constant expansion case.

DANIELS & DANIELS



RUBBER BAND EXPERIMENT shows the linear relation between recession velocity and distance. Here two snapshots are shown of a rubber band pulled upward at a certain rate. The velocity of different points marked on the band is given by the length of the colored arrows. For example, the point closest to the origin moves the least during the interval between snapshots, so its velocity is the smallest (*yellow arrow*). In contrast, the farthest point moves the most, so its velocity is the highest (*violet arrow*). The slope of the resulting line is the rate of expansion (*left graph*). If the rate changes over time, the slope, too, will change (*right graph*). Earlier times plot toward the upper right, because light from more distant objects takes longer to reach the earth, the origin of the plot. If the rate was slower in the past—indicating that the expansion has been accelerating—the line will curve upward (*red line*). If the rate was faster—as in a decelerating expansion—it will curve downward (*blue line*).

DANIELS & DANIELS

DAVID SCHNEIDER

redshift supernovae, ones that erupted between four and seven billion years ago, when the universe was between one half and two thirds of its present age. Both groups were hit with a major surprise: the supernovae are fainter than expected. The difference is slight, the distant supernovae being, on average, only 25 percent dimmer than forecast. But this result is enough to call long-standing cosmological theories into question.

Before drawing any sweeping conclusions, astronomers on both teams have been asking themselves whether there is a prosaic explanation for the relative dimness of these distant supernovae. One culprit could be murkiness caused by cosmic dust, which might screen out some of the light. We think we can discount this possibility, however, because dust grains would tend to filter out blue light more than red, causing the supernovae to appear redder than they really are (in the same way that atmospheric dust colors the setting sun). We observe no such alteration. Also, we would expect that cosmic dust, unless it is spread very smoothly throughout space, would introduce a large amount of variation in the measurements, which we do not see either.

Another possible disturbance is gravitational lensing, the bending of light rays as they skirt galaxies en route. Such lensing occasionally causes brightening, but most often it causes demagnification and thus can contribute to the dimness of distant supernovae. Yet calculations show that this effect becomes important only for sources located even farther away than the supernovae we are studying, so we can dismiss this complication as well.

Finally, we worried that the distant supernovae are somehow different from the nearby ones, perhaps forming from younger stars that contain fewer heavy elements than is typical in more mature galaxies. Although we cannot rule out this possibility, our analysis already tries to take such differences into account. These adjustments appear to work well when we apply them to nearby galaxies, which range widely in age, makeup and the kinds of supernovae seen.

Because none of these mundane effects fits the new observations, we and many other scientists are now led to think that the unexpected faintness of distant supernovae is indeed caused by the structure of the cosmos. Two different properties of space and of time might be contributing.

First, space might have negative curvature. Such warping is easier to comprehend with a two-dimensional analogy. Creatures living in a perfectly flat, two-dimensional world (like the characters in Edwin A. Abbott's classic novel *Flatland*) would find that a circle of radius r has a circumference of exactly $2\pi r$. But if their world were subtly bent into a saddle shape, it would have a slight negative curvature [see "Inflation in a Low-Density Universe," by Martin A. Bucher and David N. Spergel, on page 62]. The two-dimensional residents of Saddleland might be oblivious to this curvature until they measured a large circle of some set radius and discovered that its circumference was greater than $2\pi r$.

Most cosmologists have assumed, for various theoretical reasons, that our three-dimensional space, like Flatland, is not curved. But if it had negative curvature, the large sphere of radiation given off by an ancient supernova would have a greater area than it does in geometrically flat space, making the source appear strangely faint.

A second explanation for the unexpected dimness of distant supernovae is that they are farther away than their redshifts suggest. Viewed another way, supernovae located at these enormous distances seem to have less redshift than anticipated. To account for the smaller redshift, cosmologists postulate that the universe must have expanded more slowly in the past than they had expected, giving less of an overall stretch to the universe and to the light traveling within it.

The Force

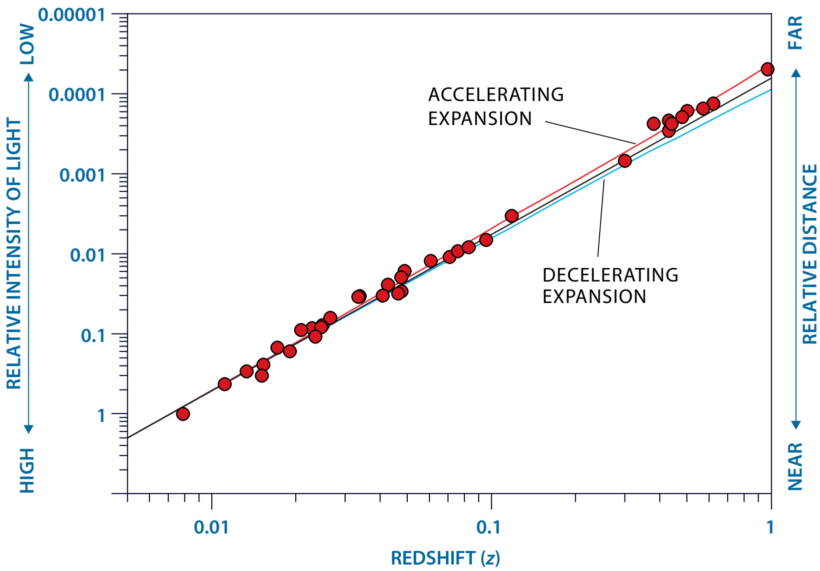
What is the significance of the cosmic expansion slowing less quickly than previously thought? If the universe is made of normal matter, gravity must steadily slow the expansion. Little slowing, as indicated by the supernovae measurements, thus implies that the overall density of matter in the universe is low.

Although this conclusion undermines theoretical preconceptions, it agrees with several other lines of evidence. For example, astronomers have noted that certain stars appear to be older than the accepted age of the universe—a clear impossibility. But if the cosmos expanded more slowly in the past, as the supernovae now indicate, the age of the universe must be revised upward, which may resolve the conundrum. The new results also accord with other recent attempts to ascertain the total amount of matter, such as studies of galaxy clusters [see "The Evolution of Galaxy Clusters," by J. Patrick Henry, Ulrich G. Briel and Hans Böhringer; Scientific American, December 1998].

What does the new understanding of the density of matter in the universe say about its curvature? According to the principles of general relativity, curvature and deceleration are connected. To paraphrase John A. Wheeler, formerly at Princeton University: matter tells space-time how to curve, and space-time tells matter how to move. A small density of matter implies negative curvature as well as little slowing. If the universe is nearly empty, these two dimming effects are both near their theoretical maximum.

The big surprise is that the supernovae we see are fainter than predicted even for a nearly empty universe (which has maximum negative curvature). Taken at face value, our observations appear to require that expansion is actually accelerating with time. A universe composed only of normal matter cannot grow in this fashion, because its gravity is always attractive. Yet according to Einstein's theory, the expansion can speed up if an exotic form of energy fills empty space everywhere. This strange "vacuum energy" is embodied in Einstein's equations as the so-called cosmological constant. Unlike ordinary forms of mass and energy, the vacuum energy adds gravity that is repulsive and can drive the universe apart at ever increasing speeds [see "Cosmological Antigravity," by Lawrence M. Krauss, on page 52]. Once we admit this extraordinary possibility, we can explain our observations perfectly, even assuming the flat geometry beloved of theorists.

Evidence for a strange form of energy imparting a repulsive gravitational force is the most interesting result we could have hoped for, yet it is so astonishing that we and others remain suitably skeptical. Fortunately, advances in the technology available to astronomers, such as new infrared detectors and the Next Generation Space Telescope, will soon permit us to test our conclusions by offering greater precision and reliability. These marvelous instruments will also allow us to perceive even fainter beacons that flared still longer ago in galaxies that are much, much farther away. SA



DAVID SCHNEIDER; SOURCE: CRAIG J. HOGAN, ROBERT P. KIRSHNER AND NICHOLAS B. SUNTZEFF

SUPERNOVA OBSERVATIONS by the authors' team (*red dots*) deviate slightly but significantly from the pattern that many theoreticians expected—namely, a fairly rapid deceleration (*blue line*) that should occur if the universe is "flat" and has no cosmological constant. These observations indicate that the universe has only 20 percent of the matter necessary to make it flat, because it is decelerating more slowly than predicted (*black line*). The measurements even suggest that expansion is accelerating, perhaps because of a nonzero cosmological constant (*red line*).

The Authors

CRAIG J. HOGAN, ROBERT P. KIRSHNER and NICHOLAS B. SUNTZEFF share a longstanding interest in big things that go bang. Hogan earned his doctorate at the University of Cambridge and is now a professor and chair of the astronomy department at the University of Washington. Kirshner attained his Ph.D. at the California Institute of Technology, studying a type Ia supernova observed in 1972 (the brightest one seen since 1937). He is a professor of astronomy at Harvard University and also serves as an associate director of the Harvard-Smithsonian Center for Astrophysics. Suntzeff received his Ph.D. at the University of California, Santa Cruz. He works at Cerro Tololo Inter-American Observatory in La Serena, Chile, and is made of elements formed in supernovae over five billion years ago.

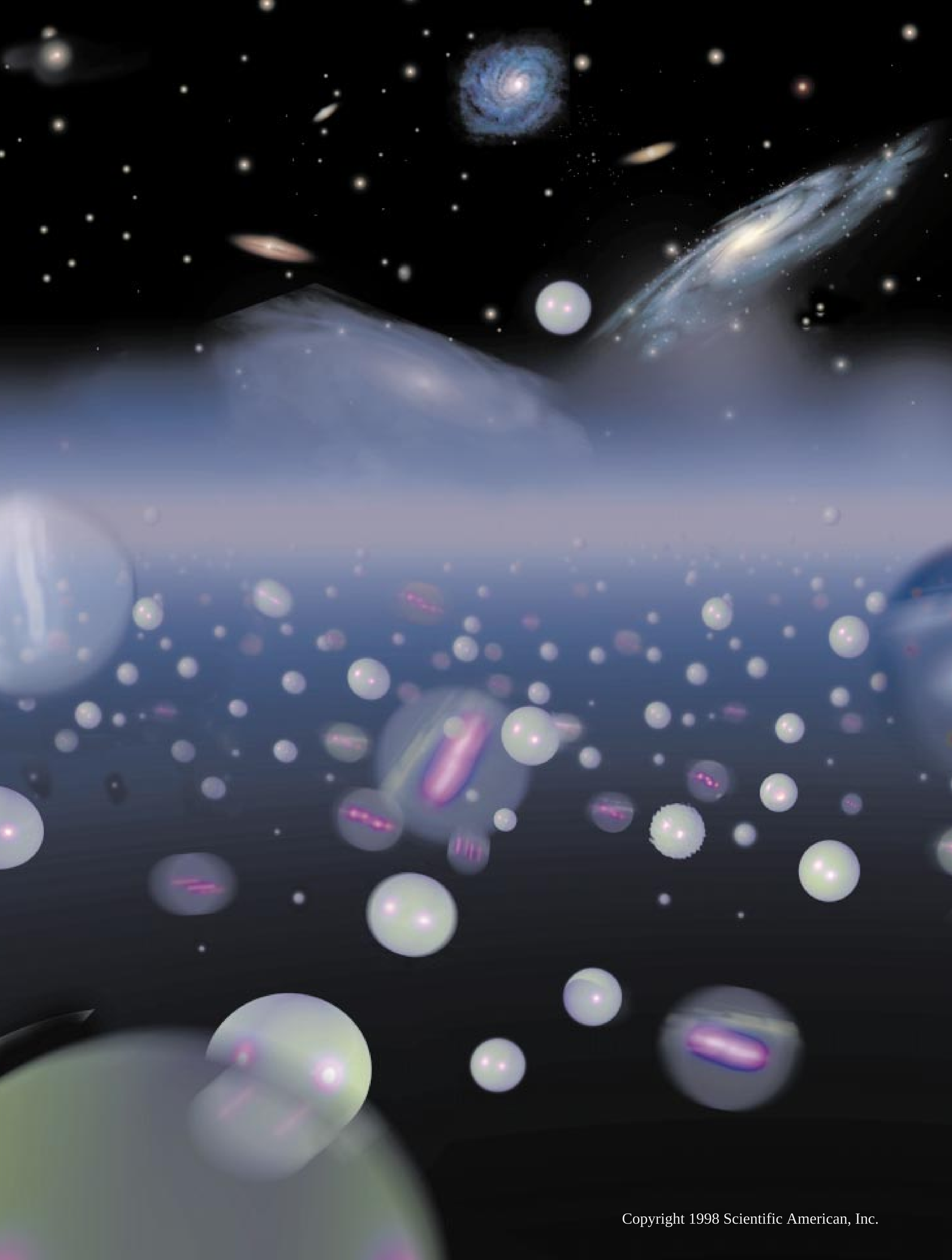
Further Reading

The Little Book of the Big Bang. Craig J. Hogan. Springer-Verlag, 1998.

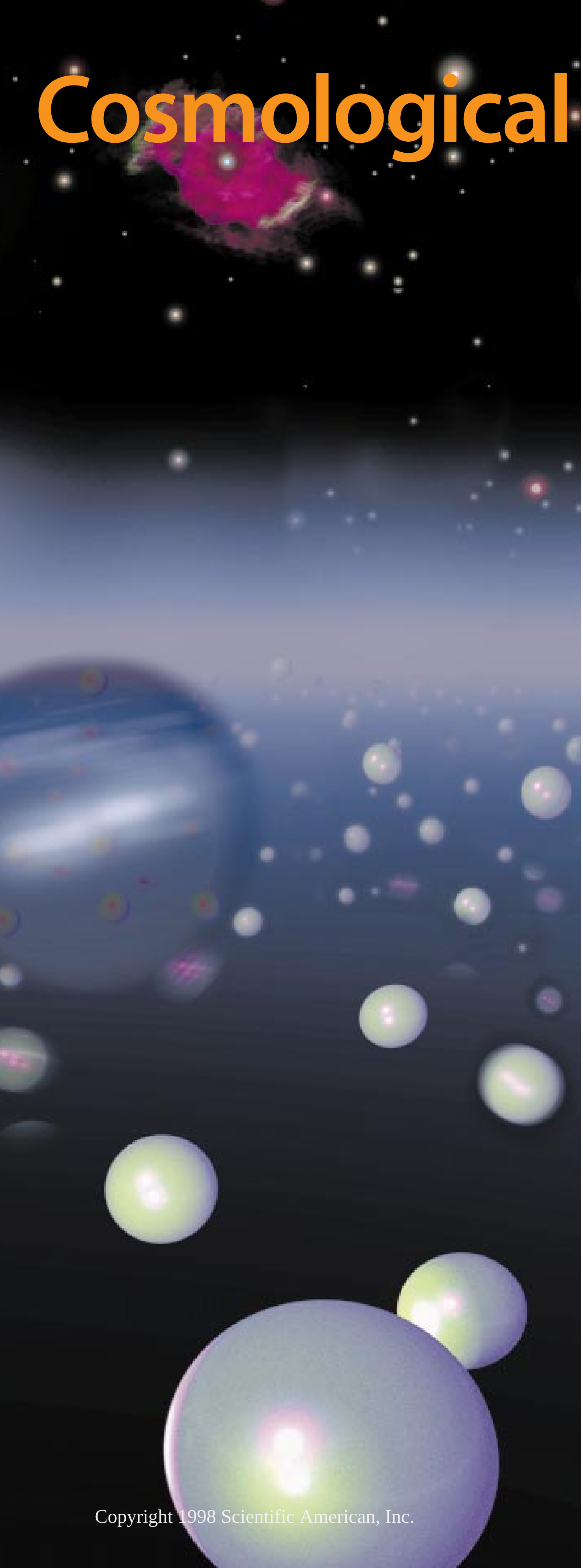
Discovery of a Supernova Explosion at Half the Age of the Universe. S. Perlmutter, G. Aldering, M. Della Valle, S. Deustua, R. S. Ellis, S. Fabbro, A. Fruchter, G. Goldhaber, D. E. Groom, I. M. Hook, A. G. Kim, M. Y. Kim, R. A. Knop, C. Lidman, R. G. McMahon, Peter Nugent, R. Pain, N. Panagia, C. R. Pennypacker, P. Ruiz-Lapuente, B. Schaefer and N. Walton (The Supernova Cosmology Project) in *Nature*, Vol. 391, pages 51–54; January 1, 1998. Preprint available at xxx.lanl.gov/abs/astro-ph/9712212 on the World Wide Web.

Observational Evidence from Supernovae for an Accelerating Universe and a Cosmological Constant. Adam G. Riess, Alexei V. Filippenko, Peter Challis, Alejandro Clocchiattia, Alan Diercks, Peter M. Garnavich, Ron L. Gilliland, Craig J. Hogan, Saurabh Jha, Robert P. Kirshner, B. Leibundgut, M. M. Phillips, David Reiss, Brian P. Schmidt, Robert A. Schommer, R. Chris Smith, J. Spyromilio, Christopher Stubbs, Nicholas B. Suntzeff and John Tonry in *Astronomical Journal*, Vol. 116, No. 3, pages 1009–1038; September 1998. Preprint at xxx.lanl.gov/abs/astro-ph/9805201 on the World Wide Web.

Additional information on supernova searches is available at cfa-www.harvard.edu/cfa/oir/Research/supernova/HighZ.html and www-supernova.lbl.gov/ on the World Wide Web.



Cosmological Antigravity



The long-derided cosmological constant—a contrivance of Albert Einstein's that represents a bizarre form of energy inherent in space itself—is one of two contenders for explaining changes in the expansion rate of the universe

by Lawrence M. Krauss

Novelist and social critic George Orwell wrote in 1946, "To see what is in front of one's nose requires a constant struggle." These words aptly describe the workings of modern cosmology. The universe is all around us—we are part of it—yet scientists must sometimes look halfway across it to understand the processes that led to our existence on the earth. And although researchers believe that the underlying principles of nature are simple, unraveling them is another matter. The clues in the sky can be subtle. Orwell's adage is doubly true for cosmologists grappling with the recent observations of exploding stars hundreds of millions of light-years away. Contrary to most expectations, they are finding that the expansion of the universe may not be slowing down but rather speeding up.

Astronomers have known that the visible universe is expanding since at least 1929, when Edwin P. Hubble demonstrated that distant galaxies are moving apart as they would if the entire cosmos were uniformly swelling in size. These outward motions are counteracted by the collective gravity of galaxy clusters and all the planets, stars, gas and dust they contain. Even the minuscule gravitational pull of, say, a paper clip retards cosmic expansion by a slight amount. A decade ago a congruence of theory and observations suggested that there were enough paper clips and other matter in the universe to almost, but never quite, halt the expansion. In the geometric terms that Albert Einstein encour-

SO-CALLED EMPTY SPACE is actually filled with elementary particles that pop in and out of existence too quickly to be detected directly. Their presence is the consequence of a basic principle of quantum mechanics combined with special relativity: nothing is exact, not even nothingness. The aggregate energy represented by these "virtual" particles, like other forms of energy, could exert a gravitational force, which could be either attractive or repulsive depending on physical principles that are not yet understood. On macroscopic scales the energy could act as the cosmological constant proposed by Albert Einstein.

ALFRED T. KAMAJIAN

Types of Matter

Type	Likely Composition	Main Evidence	Approximate Contribution to Ω
Visible matter	Ordinary matter (composed mainly of protons and neutrons) that forms stars, dust and gas	Telescope observations	0.01
Baryonic dark matter	Ordinary matter that is too dim to see, perhaps brown or black dwarfs (massive compact halo objects, or MACHOs)	Big bang nucleosynthesis calculations and observed deuterium abundance	0.05
Nonbaryonic dark matter	Exotic particles such as "axions," neutrinos with mass or weakly interacting massive particles (WIMPs)	Gravity of visible matter is insufficient to account for orbital speeds of stars within galaxies and of galaxies within clusters	0.3
Cosmological "dark matter"	Cosmological constant (energy of empty space)	Microwave background suggests cosmos is flat, but there is not enough baryonic or nonbaryonic matter to make it so	0.6

CONTENTS OF THE UNIVERSE include billions and billions of galaxies, each one containing an equally mind-boggling number of stars. Yet the bulk seems to consist of "dark matter" whose identity is still uncertain. The cosmological constant, if its existence is confirmed, would act like a yet more exotic form of dark matter on cosmological scales. The quantity omega, Ω, is the ratio of the density of matter or energy to the density required for flatness.

aged cosmologists to adopt, the universe seemed to be "flat."

The flat universe is an intermediate between two other plausible geometries, called "open" and "closed." In a cosmos where matter does battle with the outward impulse from the big bang, the open case represents the victory of expansion: the universe would go on expanding forever. In the closed case, gravity would have the upper hand, and the universe would eventually collapse again, ending in a fiery "big crunch." The open, closed and flat scenarios are analogous to launching a rocket faster than, slower than or exactly at the earth's escape velocity—the speed necessary to overcome the planet's gravitational attraction.

EINSTEIN ARCHIVES, UNIVERSITY OF ZURICH

LETTER FROM EINSTEIN, then at the Prussian Academy of Sciences in Berlin, to German mathematician Hermann Weyl concedes that a universe of unchanging size would be prone to expansion or collapse: "In the De Sitter universe two fluid and unstable distinct points separate at an accelerated pace. If there is no quasi-static world, then away with the cosmological term!"

That we live in a flat universe, the perfect balance of power, is one of the hallmark predictions of standard inflationary theory, which postulates a very early period of rapid expansion to reconcile several paradoxes in the conventional formulation of the big bang. Although the visible contents of the cosmos are clearly not enough to make the universe flat, celestial dynamics indicate that there is far more matter than meets the eye. Most of the material in galaxies and assemblages of galaxies must be invisible to telescopes. Over a decade ago I applied the term "quintessence" to this so-called dark matter, borrowing a term Aristotle used for the ether—the invisible material supposed to permeate all of space [see "Dark Matter in the Universe," by Lawrence M. Krauss; Scientific American, December 1986].

Yet an overwhelming body of evidence now implies that even the unseen matter is not enough to produce a flat universe. Perhaps the universe is not flat but rather open, in which case scientists must modify—or discard—inflationary theory [see "Inflation in a Low-Density Universe," by Martin A. Bucher and David N. Spergel, on page 62]. Or maybe the universe really is flat. If that is so, its main constituents cannot be visible matter, dark matter or radiation. Instead the universe must be composed largely of an even more ethereal form of energy that inhabits empty space, including that which is in front of our noses.

Fatal Attraction

The idea of such energy has a long and checkered history, which began when Einstein completed his general theory of relativity, more than a decade before Hubble's convincing demonstration that the universe is expanding. By tying together space, time and matter, relativity promised what had previously been impossible: a scientific understanding not merely of the dynamics of objects within the universe but of the universe itself. There was only one problem. Unlike other fundamental forces felt by matter, gravity is universally attractive—it only pulls; it cannot push. The unrelenting gravitational attraction of matter could cause the universe to collapse eventually. So Einstein, who presumed the universe to be static and stable, added an extra term to his equations, a "cosmological term," which could stabilize the universe by producing a new long-range force throughout space. If its value were positive, the term would represent a repulsive force—a kind of antigravity that could hold the universe up under its own weight.

Alas, within five years Einstein abandoned this kludge, which he associated with his "biggest blunder." The stability offered by the term turned out to be illusory, and, more important, evidence had begun to mount that the universe is expanding. As early as 1923, Einstein wrote in a letter to mathematician Hermann Weyl that "if there is no quasi-static world, then away with the cosmological term!" Like the ether before it, the term appeared to be headed for the dustbin of history.

Physicists were happy to do without such an intrusion. In the general theory of relativity, the source of gravitational forces (whether attractive or repulsive) is energy. Matter is simply one form of energy. But Einstein's cosmological term is distinct. The energy associated with it does not depend on position or time—hence the name "cosmological constant." The force caused by the constant operates even in the complete absence of matter or radiation. Therefore, its source must be a curious energy that resides in empty space. The cosmological constant, like the ether, endows the void with an almost metaphysical aura. With its demise, nature was once again reasonable.

Or was it? In the 1930s glimmers of the cosmological constant arose in a completely independent context: the effort to combine the laws of quantum mechanics with Einstein's special theory of relativity. Physicists Paul A. M. Dirac and later Richard Feynman, Julian S. Schwinger and Shinichiro Tomonaga showed that empty space was more complicated than anyone had previously imagined. Elementary particles, it turned out, can spontaneously pop out of nothingness and disappear again, if they do so for a time so short that one cannot measure them directly [see "Exploiting Zero-Point Energy," by Philip Yam; Scientific American, December 1997]. Such virtual particles, as they are called, may appear as far-fetched as angels sitting on the head of a pin. But there is a difference. The unseen particles produce measurable effects, such as alterations to the energy levels of atoms as well as forces between nearby metal plates. The theory of virtual particles agrees with observations to nine decimal places. (Angels, in contrast, normally have no discernible effect on either atoms or plates.) Like it or not, empty space is not empty after all.

Virtual Reality

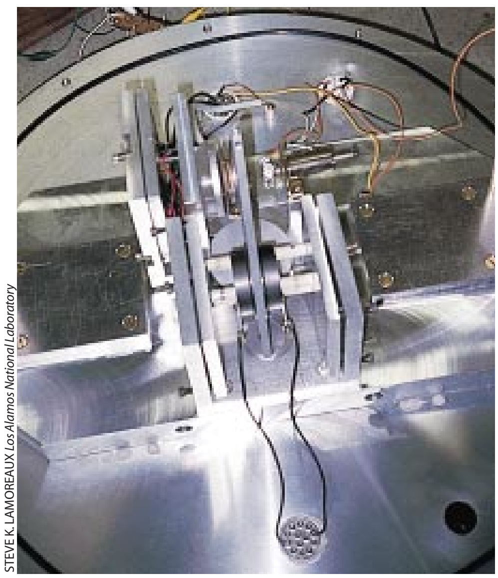
If virtual particles can change the properties of atoms, might they also affect the expansion of the universe? In 1967 Russian astrophysicist Yakov B. Zeldovich showed that the energy of virtual particles should act precisely as the energy associated with a cosmological constant. But there was a serious problem. Quantum theory predicts a whole spectrum of virtual particles, spanning every possible wavelength. When physicists add up all the effects, the total energy comes out infinite. Even if theorists ignore quantum effects smaller than a certain wavelength—for which poorly understood quantum gravitational effects presumably alter things—the calculated vacuum energy is roughly 120 orders of magnitude larger than the energy contained in all the matter in the universe.

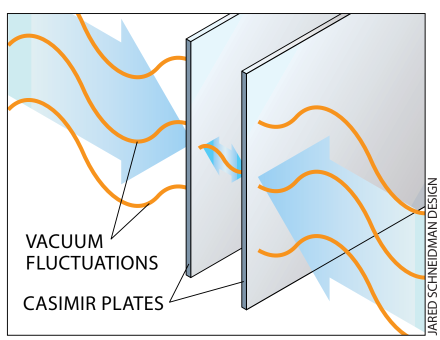
What would be the effect of such a humongous cosmological constant? Taking a cue from Orwell's maxim, you can easily put an observational limit on its value. Hold out your hand and look at your fingers. If the constant were as large as quantum theory naively suggests, the space between your eyes and your hand would expand so rapidly that the light from your hand would never reach your eyes. To see what is in front of your face would be a constant struggle (so to speak), and you would always lose. The fact that you can see anything at all means that the energy of empty space cannot be large. And the fact that we can see not only to the ends of our arms but also to the far reaches of the universe puts an even more stringent limit on the cosmological constant: almost 120 orders of magnitude smaller than the estimate mentioned above. The discrepancy between theory and observation is the most perplexing quantitative puzzle in physics today [see "The Mystery of the Cosmological Constant," by Larry Abbott; Scientific American, May 1988].

The simplest conclusion is that some as yet undiscovered physical law causes the cosmological constant to vanish. But as much as theorists might like the constant to go away, various astronomical observations—of the age of the universe, the density of matter and the nature of cosmic structures—all independently suggest that it may be here to stay.

Determining the age of the universe is one of the longstanding issues of modern cosmology. By measuring the velocities of galaxies, astronomers can calculate how long it took them to arrive at their present positions, assuming they all started out at the same place. For a first approximation, one can ignore the deceleration caused by gravity. Then the universe would expand at a constant speed and the time interval would just be the ratio of the distance between galaxies to their measured speed of separation—that is, the reciprocal of the famous Hubble constant. The higher the value of the Hubble constant, the faster the expansion rate and hence the younger the universe.

Hubble's first estimate of his eponymous constant was almost 500 kilometers per second per megaparsec—which would mean that two galaxies separated by a distance of one megaparsec (about three million light-years) are moving apart, on average, at 500 kilometers per second. This value would imply a cosmic age of about two billion years, which is in painful contradiction with the known age of the earth—about four billion years. When the gravitational attraction of matter is taken into account,





DEMONSTRATION OF CASIMIR EFFECT is one way that physicists have corroborated the theory that space is filled with fleeting "virtual particles." The Casimir effect generates forces between metal objects—for instance, an attractive force between parallel metal plates (*above*). Loosely speaking, the finite spacing of the plates prevents virtual particles larger than a certain wavelength from materializing in the gap. Therefore, there are more particles outside the plates than between them, an imbalance that pushes the plates together (*right*). The Casimir effect has a distinctive dependence on the shape of the plates, which allows physicists to tease it out from other forces of nature.

the analysis predicts that objects moved faster early on, taking even less time to get to their present positions than if their speed had been constant. This refinement reduces the age estimate by one third, unfortunately worsening the discrepancy.

Over the past seven decades, astronomers have improved their determination of the expansion rate, but the tension between the calculated age of the universe and the age of objects within it has persisted. In the past decade, with the launch of the Hubble Space Telescope and the development of new observational techniques, disparate measurements of the Hubble constant are finally beginning to converge. Wendy L. Freedman of the Carnegie Observatories and her colleagues have inferred a value of 73 kilometers per second per megaparsec (with a most likely range, depending on experimental error, of 65 to 81) [see "The Expansion Rate and Size of the Universe," by Wendy L. Freedman; Scientific American, November 1992]. These results put the upper limit on the age of a flat universe at about 10 billion years.

The Age Crisis

Is that value old enough? It depends on the age of the oldest objects that astronomers can date. Among the most ancient stars in our galaxy are those found in tight groups known as globular clusters, some of which are located in the outskirts of our galaxy and are thus thought to have formed before the rest of the Milky Way. Estimates of their age, based on calculations of how fast stars burn their nuclear fuel, traditionally ranged from 15 to 20 billion years. Such objects appeared to be older than the universe.

To determine whether this age conflict was the fault of cosmology or of stellar modeling, in 1995 my colleagues—Brian C. Chaboyer, then at the Canadian Institute of Theoretical Astrophysics, Pierre Demarque of Yale University and Peter J. Kernan of Case Western Reserve University—and I reassessed the globular cluster ages. We simulated the life cycles of three million different stars whose properties spanned the existing uncertainties, and then compared our model stars with those in globular clusters. The oldest, we concluded, could be as young as 12.5 billion years old, which was still at odds with the age of a flat, matter-dominated universe.

But two years ago the Hipparcos satellite, launched by the European Space Agency to measure the locations of over 100,000 nearby stars, revised the distances to these stars and, indirectly, to globular clusters. The new distances affected estimates of their brightness and forced us to redo our analysis, because brightness determines the rate at which stars consume fuel and hence their life spans. Now it seems that globulars could, at the limit of the observational error bars, be as young as 10 billion years old, which is just consistent with the cosmological ages.

But this marginal agreement is uncomfortable, because it requires that both sets of age estimates be near the edge of their allowed ranges. The only thing left that can give is the assumption that we live in a flat, matter-dominated universe. A lower density of matter, signifying an open universe with slower deceleration, would ease the tension somewhat. Even so, the only way to lift the age above 12.5 billion years would be to consider a universe dominated not by matter but by a cosmological constant. The resulting repulsive force would cause the Hubble expansion to accelerate over time. Galaxies would have been moving apart slower than they are today, taking longer to reach their present separation, so the universe would be older.

The current estimates of age are merely suggestive. Meanwhile other pillars of observational cosmology have recently been shaken, too. As astronomers have surveyed ever larger regions of the cosmos, their ability to tally up its contents has improved. Now the case is compelling that the total amount of matter is insufficient to yield a flat universe.

This cosmic census first involves calculations of the synthesis of elements by the big bang. The light elements in the cosmos—hydrogen and helium and their rarer isotopes, such as deuterium—were created in the early universe in relative amounts that depended on the number of available protons and neutrons, the constituents of normal matter. Thus, by comparing the abundances of the various isotopes, astronomers can deduce the total amount of ordinary matter that was produced in the big bang. (There could, of course, also be other matter not composed of protons and neutrons.)

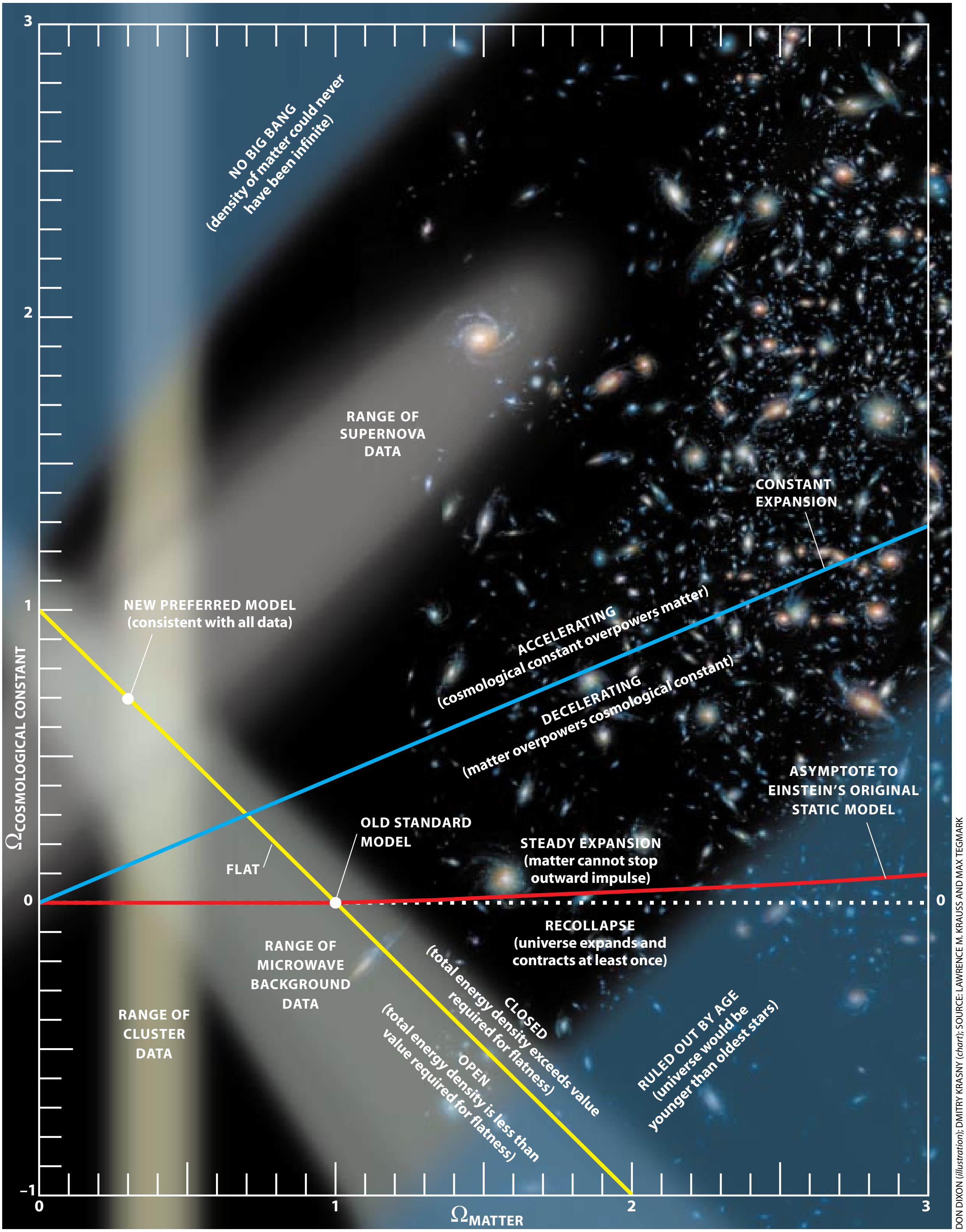
The relevant observations took a big step forward in 1996 when David R. Tytler and Scott Burles of the University of California at San Diego and their colleagues measured the primordial abundance of deuterium using absorption of quasar light by intergalactic hydrogen clouds. Because these clouds have never contained stars, their deuterium could only have been created by the big bang. Tytler and Burles's finding implies that the average density of ordinary matter is between 4 and 7 percent of the amount needed for the universe to be flat.

Astronomers have also probed the density of matter by studying the largest gravitationally bound objects in the universe: clusters of galaxies. These groupings of hundreds of galaxies account for almost all visible matter. Most of their luminous content takes the form of hot intergalactic gas, which emits x-rays. The temperature of this gas, inferred from the spectrum of the x-rays, depends on the total mass of the cluster: in more massive clusters, the gravity is stronger and hence the pressure that supports the gas against gravity must be larger, which drives the temperature higher. In 1993 Simon D. M. White, now at the Max Planck Institute for Astrophysics in Garching, Germany, and his colleagues compiled information about several different clusters to argue that luminous matter accounted for between 10 and 20 percent of the total mass of the objects. When combined with the measurements of deuterium, these results imply that the total density of clustered matter—including protons and neutrons as well as

Summary of Inferred Values of Cosmic Matter Density

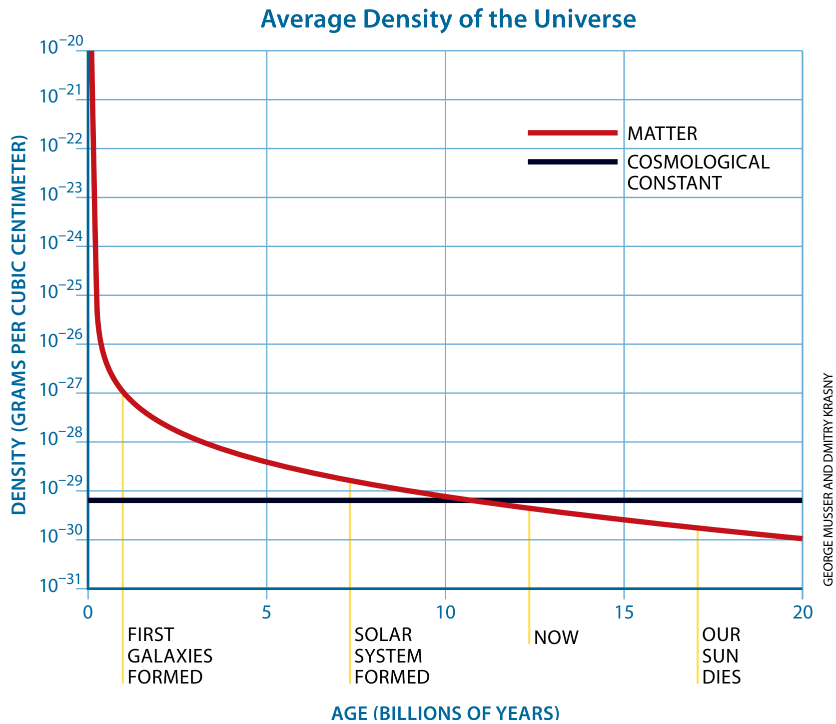
Observation	Ω_{matter}
Age of universe	<1
Density of protons and neutrons	0.3–0.6
Galaxy clustering	0.3–0.5
Galaxy evolution	0.3–0.5
Cosmic microwave background radiation	$\lesssim 1$
Supernovae type Ia	0.2–0.5

MEASUREMENTS of the contribution to Ω from matter are in rough concordance. Although each measurement has its skeptics, most astronomers now accept that matter alone cannot make Ω equal to 1. But other forms of energy, such as the cosmological constant, may also pitch in.



MAP OF MODELS shows how the unfolding of the universe depends on two key cosmological quantities: the average density of matter (*horizontal axis*) and the density of energy in the cosmological constant (*vertical axis*). Their values, given here in standard cosmological units, have three distinct effects. First, their sum (which represents the total cosmic energy content) determines the geometry of space-time (*yellow line*). Second, their difference (which represents the relative strength of expansion and gravity) determines how the expansion rate changes over time (*blue line*). These two effects have been probed by recent observations (*shaded regions*). The third, a balance of the two densities, determines the fate of the universe (*red line*). The three effects have many permutations—unlike the view of cosmology in which the cosmological constant is assumed to be zero and there are only two possible outcomes.

COSMIC COINCIDENCE is one of many mysteries swirling about the cosmological constant. The average density of ordinary matter decreases as the universe expands (*red*). The equivalent density represented by the cosmological constant is fixed (*black*). So why, despite these opposite behaviors, do the two have nearly the same value today? The consonance is either happenstance, a precondition for human existence (an appeal to the weak anthropic principle) or an indication of a mechanism not currently envisaged.



The Fate of the Universe

The cosmological constant changes the usual simple picture of the future of the universe. Traditionally, cosmology has predicted two possible outcomes that depend on the geometry of the universe or, equivalently, on the average density of matter. If the density of a matter-filled universe exceeds a certain critical value, it is "closed," in which case it will eventually stop expanding, start contracting and ultimately vanish in a fiery apocalypse. If the density is less than the critical value, the universe is "open" and will expand forever. A "flat" universe, for which the density equals the critical value, also will expand forever but at an ever slower rate.

Yet these scenarios assume that the cosmological constant equals zero. If not, it—rather than matter—may control the ultimate fate of the universe. The reason is that the constant, by definition, represents a fixed density of energy in space. Matter cannot compete: a doubling in radius dilutes its density eightfold. In an expanding universe the energy density associated with a cosmological constant must win out. If the constant has a positive value, it generates a long-range repulsive force in space, and the universe will continue to expand even if the total energy density in matter and in space exceeds the critical value. (Large negative values of the constant are ruled out because the resulting attractive force would already have brought the universe to an end.)

Even this new prediction for eternal expansion assumes that the constant is indeed constant, as general relativity suggests that it should be. If in fact the energy density of empty space does vary with time, the fate of the universe will depend on how it does so. And there may be a precedent for such changes—namely, the inflationary expansion in the primordial universe. Perhaps the universe is just now entering a new era of inflation, one that may eventually come to an end. *—L.M.K.*

more exotic particles such as certain dark-matter candidates—is at most 60 percent of that required to flatten the universe.

A third set of observations, ones that also bear on the distribution of matter at the largest scales, supports the view that the universe has too little mass to make it flat. Perhaps no other subfield of cosmology has advanced so much in the past 20 years as the understanding of the origin and nature of cosmic structures. Astronomers had long assumed that galaxies coalesced from slight concentrations of matter in the early universe, but no one knew what would have produced such undulations. The development of the inflationary theory in the 1980s provided the first plausible mechanism—namely, the enlargement of quantum fluctuations to macroscopic size.

Numerical simulations of the growth of structures following inflation have shown that if dark matter was not made from protons and neutrons but from some other type of particle (such as so-called WIMPs), tiny ripples in the cosmic microwave background radiation could grow into the structures now seen. Moreover, concentrations of matter should still be evolving into clusters of galaxies if the overall density of matter is high. The relatively slow growth of the number of rich clusters over the recent history of the universe suggests that the density of matter is less than 50 percent of that required for a flat universe [see "The Evolution of Galaxy Clusters," by J. Patrick Henry, Ulrich G. Briel and Hans Böhringer; Scientific American, December 1998].

Nothing Matters

These many findings that the universe has too little matter to make it flat have become convincing enough to overcome the strong theoretical prejudice against this possibility. Two interpretations are viable: either the universe is open, or it is made flat by some additional form of energy that is not associated with ordinary matter. To distinguish between these alternatives, astronomers have been pushing to measure the microwave background at high resolution. Initial indications now favor a flat universe. Meanwhile researchers studying distant supernovae have provided the first direct, if tentative, evidence that the expansion of the universe is accelerating, a telltale sign of a cosmological constant with the same value implied by the other data [see "Surveying Space-time with Supernovae," by Craig J. Hogan, Robert P. Kirshner and Nicholas B. Suntzeff, on page 46]. Observations of the microwave background and of supernovae illuminate two different aspects of cosmology. The microwave background reveals the geometry of the universe, which is sensitive to the total density of energy, in whatever form, whereas the supernovae directly probe the expansion rate of the universe, which depends on the difference between the density of matter (which slows the expansion) and the cosmological constant (which can speed it up).

Together all these results suggest that the constant contributes between 40 and 70 percent of the energy needed to make the universe flat [*see illustration on page 57*]. Despite the preponderance of evidence, it is worth remembering the old saw that an astronomical theory whose predictions agree with all observations is probably wrong, if only because some of the measurements or some of the predictions are likely to be erroneous. Nevertheless, theorists are already scrambling to understand what 20 years ago would have been unthinkable: a cosmological constant greater than zero yet much smaller than current quantum theories predict. Some feat of fine-tuning must subtract virtual-particle energies to 123 decimal places but leave the 124th untouched—a precision seen nowhere else in nature.

One direction, explored recently by Steven Weinberg of the University of Texas at Austin and his colleagues, invokes the last resort of cosmologists, the anthropic principle. If the observed universe is merely one of an infinity of disconnected universes—each of which might have slightly different constants of nature, as suggested by some incarnations of inflationary theory combined with emerging ideas of quantum gravity—then physicists can hope to estimate the magnitude of the cosmological constant by asking in which universes intelligent life is likely to evolve. Weinberg and others have arrived at a result that is compatible with the apparent magnitude of the cosmological constant today.

Most theorists, however, do not find these notions convincing, as they imply that there is no reason for the constant to take on a particular value; it just does. Although that argument may turn out to be true, physicists have not yet exhausted the other possibilities, which might allow the constant to be constrained by fundamental theory rather than by accidents of history [see "The Anthropic Principle," by George Gale; Scientific American, December 1981].

Another direction of research follows in a tradition established by Dirac. He argued that there is one measured large number in the universe—its age (or, equivalently, its size). If certain physical quantities were changing over time, they might naturally be either very large or very small today [see "P. A. M. Dirac and the Beauty of Physics," by R. Corby Hovis and Helge Kragh; Scientific American, May 1993]. The cosmological constant could be one example. It might not, in fact, be constant. After all, if the cosmological constant is fixed and nonzero, we are living at the first and only time in the cosmic history when the density of matter, which decreases as the universe expands, is comparable to the energy stored in empty space. Why the coincidence? Several groups have instead imagined that some form of cosmic energy mimics a cosmological constant but varies with time.

This concept was explored by P. James E. Peebles and Bharat V. Ratra of Princeton University a decade ago. Motivated by the new supernova findings, other groups have resurrected the idea. Some have drawn on emerging concepts from string theory. Robert Caldwell and Paul J. Steinhardt of the University of Pennsylvania have reproposed the term "quintessence" to describe this variable energy. It is one measure of the theoretical conundrum that the dark matter that originally deserved this term now seems almost mundane by comparison. As much as I like the word, none of the theoretical ideas for this quintessence seems compelling. Each is ad hoc. The enormity of the cosmological-constant problem remains.

How will cosmologists know for certain whether they have to reconcile themselves to this theoretically perplexing universe? New measurements of the microwave background, the continued analysis of distant supernovae and measurements of gravitational lensing of distant quasars should be able to pin down the cosmological constant over the next few years. One thing is already certain. The standard cosmology of the 1980s, postulating a flat universe dominated by matter, is dead. The universe is either open or filled with an energy of unknown origin. Although I believe the evidence points in favor of the latter, either scenario will require a dramatic new understanding of physics. Put another way, "nothing" could not possibly be more interesting. SA

The Author

LAWRENCE M. KRAUSS works at the interface of physics and astronomy. He studies the workings of stars, black holes, gravitational lenses and the early universe in order to shed light on particle physics beyond the current Standard Model, including the unification of forces, quantum gravity and explanations for dark matter. Krauss is currently chair of the physics department at Case Western Reserve University. He is the author of four popular books, most recently *Beyond Star Trek,* which looks at the science depicted in movies and on television.

Further Reading

Dreams of a Final Theory. Steven Weinberg. Pantheon Books, 1992.

Principles of Physical Cosmology. P. James E. Peebles. Princeton University Press, 1993.

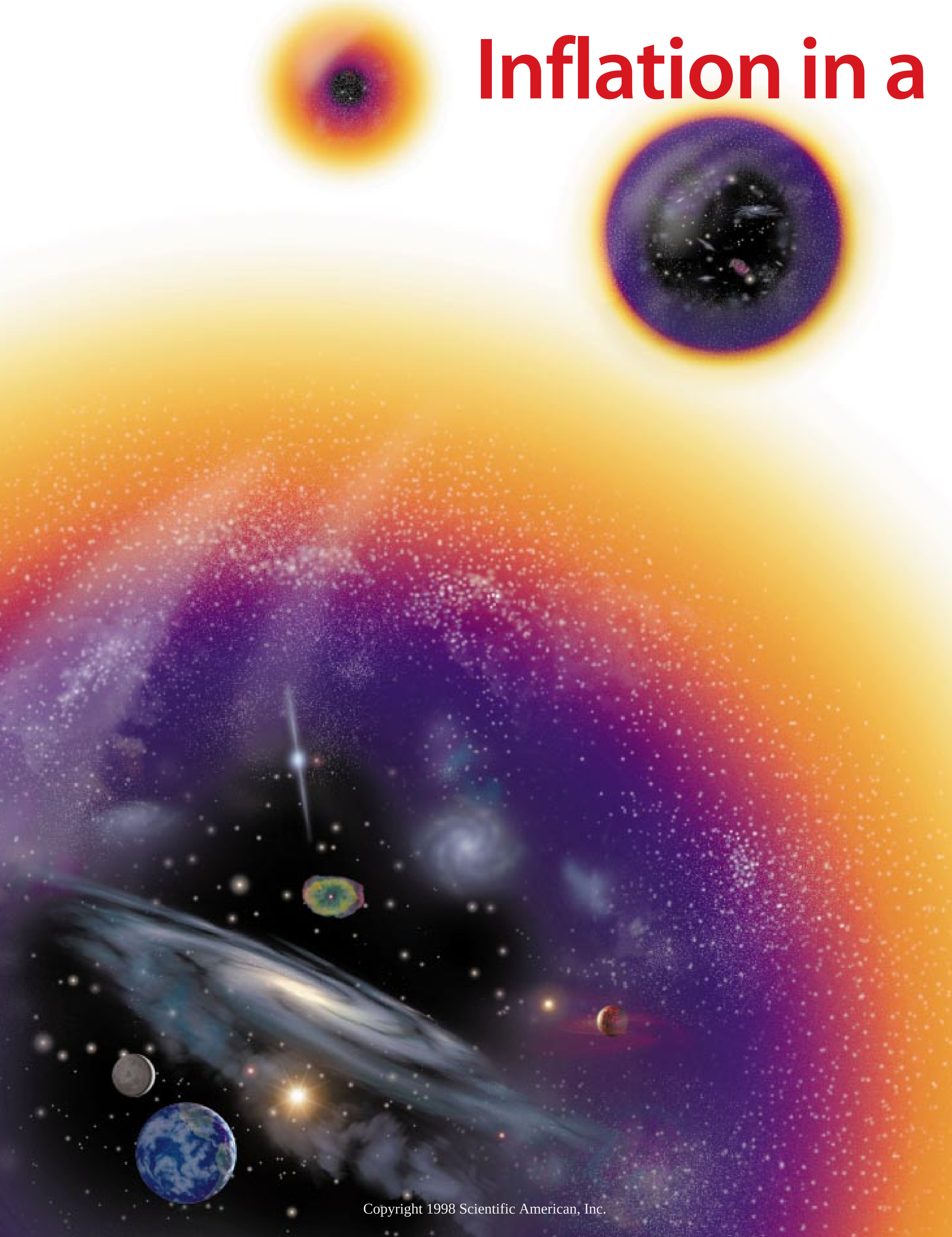
Before the Beginning: Our Universe and Others. Martin Rees. Addison-Wesley, 1997.

The Age of Globular Clusters in Light of Hipparcos: Resolving the Age Problem? Brian Chaboyer, Pierre Demarque, Peter J. Kernan and Lawrence M. Krauss in *Astrophysical Journal,* Vol. 494, No. 1, pages 96–110; February 10, 1998. Preprint available at http://xxx.lanl.gov/abs/astro-ph/9706128 on the World Wide Web.

The End of the Age Problem, and the Case for a Cosmological Constant Revisited. Lawrence M. Krauss in *Astrophysical Journal,* Vol. 501, No. 2, pages 461–466; July 10, 1998. Preprint available at xxx.lanl.gov/abs/astro-ph/9706227 on the World Wide Web.

Living with Lambda. J. D. Cohn. Preprint available at xxx.lanl.gov/abs/astro-ph/9807128 on the World Wide Web.

Inflation in a



Low-Density Universe

Evidence has gradually accumulated that the universe has less matter, and therefore is expanding faster, than the theory of inflation traditionally predicts. But a more sophisticated version of the theory readily explains the observations

by Martin A. Bucher and David N. Spergel

Cosmology has a reputation as a difficult science, but in many ways explaining the whole universe is easier than understanding a single-celled animal. On the largest cosmic scales, where stars, galaxies and even galaxy clusters are mere flecks, matter is spread out evenly. And it is governed by only one force, gravity. These two basic observations—large-scale uniformity and the dominance of gravity—are the basis of the big bang theory, according to which our universe has been expanding for the past 12 billion years or so. Despite its simple underpinnings, the theory is remarkably successful in explaining the velocity of galaxies away from one another, the relative amounts of light elements, the dim microwave glow in the sky and the general evolution of astronomical structures. The unfolding of the cosmos, it seems, is almost completely insensitive to the details of its contents. Unfortunately for biologists, the same principle does not apply to even the simplest organism.

Yet there are paradoxes inherent in the big bang theory. Two decades ago cosmologists resolved these troubling inconsistencies by incorporating ideas from particle physics—giving rise to the theory of "inflation." But now this elaboration is itself facing a crisis, brought on by recent observations that contradict its prediction for the average density of matter in the cosmos. Cosmologists are realizing that the universe may not be quite so simple as they had thought. Either they must posit the existence of an exotic form of matter or energy, or they must add a layer of complexity to the theory of inflation. In this article we will focus on the second option.

Strictly speaking, the big bang theory does not describe the birth of the universe, but rather its growth and maturation. According to the theory, the infant universe was an extremely hot, dense cauldron of radiation. A part of it, a chunk smaller than a turnip, eventually enlarged into the universe observable today. (There are other parts of the universe, perhaps infinite in extent, that we cannot see, because their light has not yet had time to reach the earth.) The idea of an expanding universe can be confusing; even Albert Einstein initially regarded it with suspicion. When the cosmos expands, the distance between any two independent objects increases. Distant galaxies move apart because the space between them is getting larger of its own accord, just as raisins move apart in a rising loaf of bread.

A natural consequence of the expansion of a uniform universe is Hubble's law, whereby galaxies are moving away from the earth (or from any other point in the universe) at speeds proportional to their distance. Not all objects in the universe obey this law, because mutual gravitational attraction fights against the swelling of space. For example, the sun and the earth are not moving apart. But it holds on the largest scales. In the simplest version of the big bang, the expansion has always proceeded at much the same rate.

In the Beginning, Paradox

As the youthful universe expanded, it cooled, thinned out and became increasingly complex. Some of the radiation condensed into the familiar elementary particles and simple atomic nuclei. Within roughly 300,000 years, the temperature had dropped to 3,000 degrees Celsius, cool enough for the electrons and protons to combine and form hydrogen atoms. At this moment the universe became transparent, setting loose the famous cosmic microwave background radiation. The radiation is very smooth, indicating that the density of matter in different regions of the early universe varied by only one part in 100,000. Tiny though these differences were, the slight concentrations eventually grew into galaxies and galaxy clusters

BUBBLE UNIVERSES are self-contained universes that grow within a larger and otherwise empty "multiverse." True to the weirdness of relativity, time and space have different meanings inside and outside each bubble; time, as perceived inside, increases toward the center of the bubble; the wall of the bubble represents the big bang for that universe. Of course, this painting depicts an impossible perspective. Even if an observer could exist outside the bubble, he or she or it could not peer inside, because the bubble expands at the speed of light. Such ideas may sound like science fiction, but so does any other cutting-edge science.

ALFRED T. KAMAJIAN

[see "The Evolution of the Universe," by P. James E. Peebles, David N. Schramm, Edwin L. Turner and Richard G. Kron; Scientific American, October 1994].

Despite its successes, the standard big bang theory cannot answer several profound questions. First, why is the universe so uniform? Two regions on opposite sides of the sky look broadly the same, yet they are separated by more than 24 billion light-years. Light has been traveling for only about 12 billion years, so the regions have yet to see each other. There has never been enough time for matter, heat or light to flow between them and homogenize their density and temperature [*see illustration on page 69*]. Somehow the uniformity of the universe must have predated the expansion, but the theory does not explain how.

Conversely, why did the early universe have any density variations at all? Fortunately, it did: without these tiny undulations, the universe today would still be of uniform density—a few atoms per cubic meter—and neither the Milky Way nor the earth would exist.

Finally, why is the rate of cosmic expansion just enough to counteract the collective gravity of all the matter in the universe? Any significant deviation from perfect balance would have magnified itself over time. If the expansion rate had been too large, the universe today would seem nearly devoid of matter. If gravity had been too strong, the universe would have already collapsed in a big crunch, and you would not be reading this article.

Cosmologists express this question in terms of the variable omega, Ω, the ratio of gravitational energy to kinetic energy (that is, the energy contained in the motion of matter as space expands). The variable is proportional to the density of matter in the universe—a higher density means stronger gravity, hence a larger Ω. If Ω equals one, its value never changes; otherwise it rapidly decreases or increases in a self-reinforcing process, as either kinetic or gravitational energy comes to dominate. After billions of years, Ω should effectively be either zero or infinity. Because the current density of the universe is (thankfully) neither zero nor infinity, the original value of Ω must have been exactly one or extraordinarily close to it (within one part in 10^{18}). Why? The big bang theory offers no explanation apart from dumb luck.

These shortcomings do not invalidate the theory—which neatly explains billions of years of cosmic history—but they do indicate that it is incomplete. To fill in the gap, in the early 1980s cosmologists Alan H. Guth, Katsuhiko Sato, Andrei D. Linde, Andreas Albrecht and Paul J. Steinhardt developed the theory of inflation [see "The Inflationary Universe," by Alan H. Guth and Paul J. Steinhardt; Scientific American, May 1984].

The price paid for resolving the paradoxes is to make big bang theory more complicated. The inflationary theory postulates that the baby universe went through a period of very rapid expansion (hence the name). Unlike conventional big bang expansion, which decelerates over time, the inflationary expansion accelerated. It pushed any two independent objects apart at an ever increasing clip—eventually faster than light. This motion did not violate relativity, which prohibits bodies of finite mass from moving through space faster than light. The objects, in fact, stood still relative to the space around them. It was space itself that came to expand faster than light.

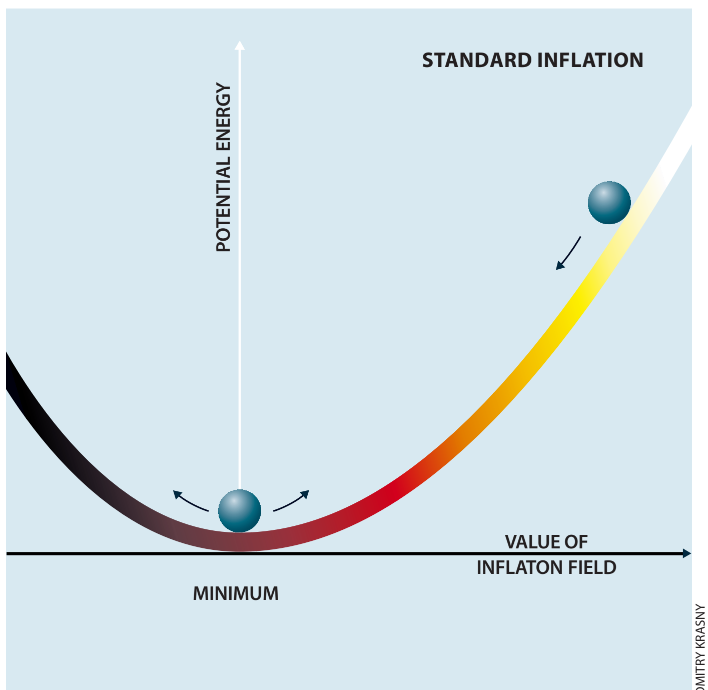
Such rapid expansion early on explains the uniformity of the universe seen today. All parts of the visible universe were once so close together that they were able to attain a common density and temperature. During inflation, different parts of this uniform universe fell out of touch; only later, after inflation ended, did light have time to catch up with the slower, big bang expansion. If there is any nonuniformity in the broader universe, it has yet to come into view.

Fieldwork

To bring about the rapid expansion, inflationary theory adds a new element to cosmology, drawn from particle physics: the "inflaton" field. In modern physics, elementary particles, such as protons and electrons, are represented by quantum fields, which resemble the familiar electric, magnetic and gravitational fields. A field is simply a function of space and time whose oscillations are interpreted as particles. Fields are responsible for the transmission of forces.

The inflaton field imparts an "antigravity" force that stretches space. Associated with a given value of the inflaton field is a potential energy. Much like a ball rolling down a hill, the inflaton field tries to roll toward the bottom of its potential [*see illustration below*]. But the expansion of the universe introduces what may be described as a cosmological friction, impeding the descent. As long as the friction dominates, the inflaton field is almost stuck in place. Its value is nearly constant, so the antigravity force gains in strength relative to gravity—causing the distance between once nearby objects to increase at ever faster rates. Eventually the field weakens and converts its remaining energy into radiation. Afterward the expansion of the universe continues as in the standard big bang.

Cosmologists visualize this process in terms of the shape of the universe. According to Einstein's general theory of rela-



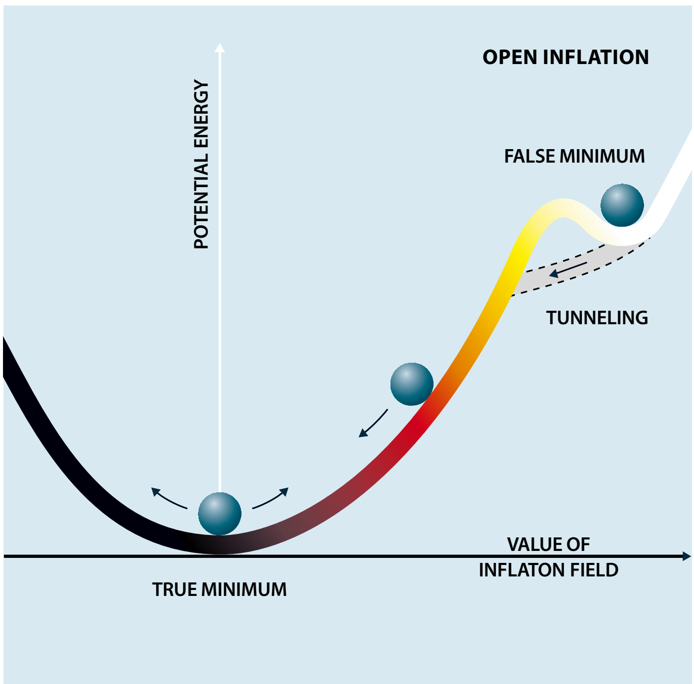
INFLATON FIELD, the origin of the force that caused space to expand, behaved like a ball rolling down a hill: it sought to minimize its potential energy (*vertical axis*) by changing its value (*horizontal axis*). The field began high up the hill because of quantum processes at the dawn of time. In standard inflation (*left*), the field then rolled straight to its lowest value. But in open inflation (*right*), it got caught in a

tivity, gravity is a geometric effect: matter and energy warp the fabric of space and time, distorting the paths that objects follow. The overall expansion of the universe, which itself is a kind of bending of space and time, is controlled by the value of Ω [*see box on page 67*]. If Ω is greater than one, the universe has a positive curvature, like the surface of an orange but in three spatial dimensions (the spherical, or "closed," geometry). If Ω is less than one, the universe has a negative curvature, like a potato chip (the hyperbolic, or "open," geometry). If it equals one, the universe is flat, like a pancake (the usual Euclidean geometry).

Inflation flattens the observable universe. Whatever the initial shape of the universe, the rapid expansion bloats it to colossal size and pushes most of it out of sight. The small visible fraction might seem flat, just as a small part of the earth's surface seems flat. Inflation thus pushes the observed value of Ω toward one. At the same time, any initial irregularities in the density of matter and radiation get evened out.

So in standard inflationary theory, cosmic flatness and uniformity are linked. For the universe to be as homogeneous as it is, the theory says the universe should be very, very flat, with Ω equal to one within one part in 100,000. Any deviation from exact flatness should be utterly impossible for astronomers to detect. Thus, for most of the past two decades observational flatness has been viewed as a firm prediction of the theory.

And that is the problem. A wide variety of astronomical observations, involving galaxy clusters and distant supernovae, now suggest that gravity is too weak to combat the expansion. If so, the density of matter must be less than predicted—with Ω equal to about 0.3. That is, the universe might be curved and open. There are three ways to interpret this result.



valley, or "false minimum." Throughout most of the universe it stayed there, and inflation never ended. In a few lucky regions, the field "tunneled" out of its valley and completed its descent. One such region became the bubble in which we live. In both styles of inflation, once the field approached its final resting place, it sloshed back and forth, filling space with matter and radiation. The big bang had begun.

The first is that inflationary theory is completely wrong. But if cosmologists abandon inflation, the formidable paradoxes so nicely resolved by the theory would reappear, and a new theory would be required. No such alternative is known.

A second interpretation takes heart from the accelerating expansion inferred from the observations of distant supernovae [see "Surveying Space-time with Supernovae," by Craig J. Hogan, Robert P. Kirshner and Nicholas B. Suntzeff, on page 46]. Such expansion hints at additional energy in the form of a "cosmological constant." This extra energy would act as a weird kind of matter, bending space much as ordinary matter does. The combined effect would be to flatten space, in which case the inflationary theory has nothing to worry about [see "Cosmological Antigravity," by Lawrence M. Krauss, on page 52]. But the inference of the cosmological constant is plagued by uncertainties about dust and the nature of the stars that undergo supernova explosions. So cosmologists are keeping their options open (so to speak).

Bubble Universes

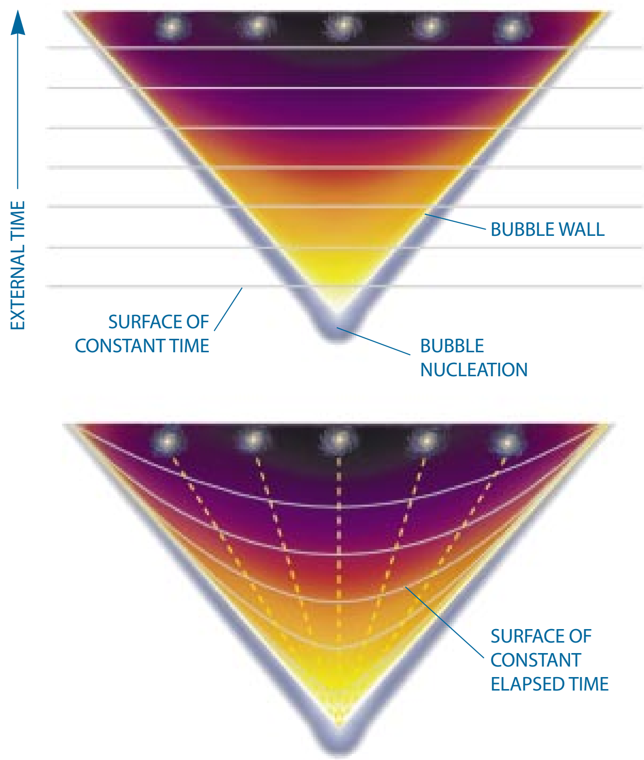
A third path is to take the observations at face value and ask whether a flat universe really is an inevitable consequence of inflation. This approach involves yet another extension of the theory to still earlier times, with some new complexity. The route was first mapped in the early 1980s by Sidney R. Coleman and Frank de Luccia of Harvard University and J. Richard Gott III of Princeton University. Ignored for over a decade, the ideas were recently developed by one of us (Bucher), along with Neil G. Turok, now at the University of Cambridge, and Alfred S. Goldhaber of the State University of New York at Stony Brook, and by Misao Sasaki and Takahiro Tanaka, now at Osaka University, and Kazuhiro Yamamoto of Kyoto University. Linde and his collaborators have also proposed some concrete models and extensions of these ideas.

If the inflaton field had a different potential-energy function, inflation would have bent space in a precise and predictable way—leaving the universe slightly curved rather than exactly flat. In particular, suppose the potential-energy function had two valleys—a false (local) minimum as well as a true (global) minimum [*see illustration at left*]. As the inflaton field rolled down, the universe expanded and became uniform. But then the field got stuck in the false minimum. Physicists call this state the "false vacuum," and any matter and radiation in the cosmos were almost entirely replaced by the energy of the inflaton field. The fluctuations inherent in quantum mechanics caused the inflaton field to jitter and ultimately enabled it to escape from the false minimum—just as shaking a pinball machine can free a trapped ball.

The escape, called false-vacuum decay, did not occur everywhere at the same time. Rather it first took place at some random location and then spread. The process was analogous to bringing water to a boil. Water heated to its boiling point does not instantaneously turn into steam everywhere. First, because of the random motion of atoms, scattered bubbles nucleate throughout the liquid—rather like the burbling of a pot of soup. Bubbles smaller than a certain minimum size collapse because of surface tension. But in larger bubbles, the energy difference between the steam and the superheated water overcomes surface tension; these bubbles expand at the speed of sound in water.

In false-vacuum decay, quantum fluctuations played the role of the random atomic motion, causing bubbles of true vacuum to nucleate. Surface tension destroyed most of the bubbles, but a few managed to grow so large that quantum effects became unimportant. With nothing to oppose them, their radius continued to increase at the speed of light. As the outside wall of a bubble passed through a point in space, the inflaton field at that point was jolted out of the false minimum and resumed its downward descent. Thereafter the space inside the bubble inflated much as in standard inflationary theory. The interior of this bubble corresponds to our universe. The moment that the inflaton field broke out of its false minimum corresponds to the big bang in older theories.

For points at different distances from the center of nucleation, the big bang occurred at different times. This disparity seems strange, to say the least. But careful examination of the inflaton field reveals what went on. The inflaton acted as a chronometer: its value at a given point represented the time elapsed since the big bang occurred at that point. Because of the time lag in the commencement of the big bang, the value of the inflaton was not the same everywhere; it was highest at the wall of the bubble and fell steadily toward the center. Mathematically, the value of the inflaton was constant on surfaces with the shape of hyperbolas [*see illustration below*].



DON DIXON; SOURCE: MARTIN A. BUCHER AND DAVID N. SPERGEL

INFINITE UNIVERSE IN FINITE SPACE? This seemingly paradoxical arrangement is possible because space and time are perceived differently outside (*top*) and inside (*bottom*) the bubble universe. Here, time—as seen by exterior observers—marches upward. Space, by definition, is any line or surface that connects points at a certain time (*horizontal lines*). The bubble looks finite. Interior observers, however, are aware only of elapsed time, the amount that has passed since the bubble first arrived at a given position. As elapsed time increases, temperature decreases—which impels physical change (*hot is yellow; cool is black*). Surfaces of constant elapsed time are hyperbolas, which bend upward and never touch the bubble wall. Points inside move apart because of cosmic expansion (*dotted lines*). Thus we count ourselves kings of infinite space.

The value of the inflaton is no mere abstraction. It determined the basic properties of the universe inside the bubble—namely, its average density and the temperature of the cosmic background radiation (today 2.7 degrees C above absolute zero). Along a hyperbolic surface, the density, temperature and elapsed time were constant. These surfaces are what observers inside the bubble perceive as constant "time." It is not the same as time experienced outside the bubble.

How is it possible for something so fundamental as time to be different on the inside and on the outside? Based on the understanding of space and time before Einstein's theories of relativity, such a feat would indeed have seemed impossible. But in relativity, the distinction between space and time blurs. What any observer calls "space" and "time" is largely a matter of convenience. Loosely speaking, time represents the direction in which things change, and change inside the bubble is driven by the inflaton.

Bounded in a Nutshell

According to relativity, the universe has four dimensions—three for space, one for time. Once the direction of time is determined, the three remaining directions must be spatial; they are the directions in which time is constant. Therefore, a bubble universe seems hyperbolic from the inside. For us, to travel out in space is, in effect, to move along a hyperbola. To look backward in time is to look toward the wall of the bubble. In principle, we could look outside the bubble and before the big bang, but in practice, the dense, opaque early universe blocks the view.

This melding of space and time allows an entire hyperbolic universe (whose volume is infinite) to fit inside an expanding bubble (whose volume, though increasing without limit, is always finite). The space inside the bubble is actually a blend of both space and time as perceived outside the bubble. Because external time is infinite, so is internal space.

The seemingly bizarre concept of bubble universes frees inflationary theory from its insistence that Ω equal one. Although the formation of the bubble created hyperbolas, it said nothing about their precise scale. The scale is instead determined by the details of the inflaton potential, and it varies over time in accordance with the value of Ω. Initially, Ω inside the bubble equals zero. During inflation, its value increases, approaching one. Thus, hyperbolas start off with an abrupt bend and gradually flatten out. The inflaton potential sets the rate and duration of flattening. Eventually inflation in the bubble comes to an end, at which point Ω is poised extremely near but very slightly below one. Then Ω starts to decrease. If the duration of inflation inside the bubble is just right (to within a few percent), the current value of Ω will match the observed value.

At first glance the process may seem baroque, but the main conclusion is simple: the uniformity and geometry of the universe need not be linked. Instead they could result from different stages of inflation: uniformity, from inflation before the nucleation of the bubble; geometry, from inflation within the bubble. Because the two properties are not intertwined, the need for uniformity does not determine the duration of inflation, which lasts just long enough to give the hyperbolas the desired degree of flatness.

In fact, this formulation is a straightforward extension of the big bang theory. The standard view of inflation describes what happened just before the conventional big bang expansion. The new conception, known as open inflationary theory,

The Geometry of the Universe

If the universe had an "outside" and people could view it from that perspective, cosmology would be much easier. Lacking these gifts, astronomers must infer the basic shape of the universe from its geometric properties. Everyday experience indicates that space is Euclidean, or "flat," on small scales. Parallel lines never meet, triangles span 180 degrees, the circumference of a circle is $2\pi r$, and so on. But it would be wrong to assume that the universe is Euclidean on large scales, just as it would be wrong to conclude that the earth is flat just because a small patch of it looks flat.

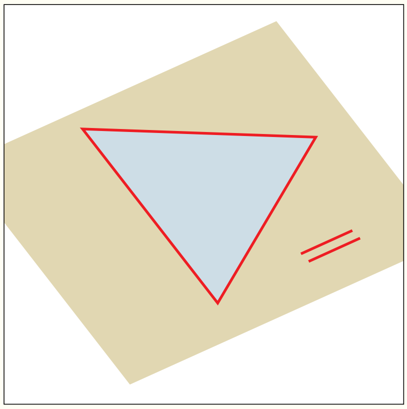
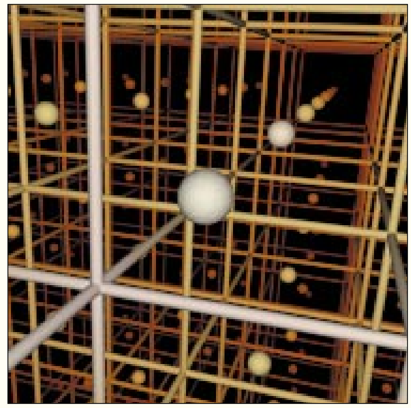
There are two other possible three-dimensional geometries consistent with the observations of cosmic homogeneity (the equivalence of all points in space) and isotropy (the equivalence of all directions). They are the spherical, or "closed," geometry and the hyperbolic, or "open," geometry. Both are characterized by a curvature length analogous to the earth's radius. If the curvature is positive, the geometry is spherical; if negative, hyperbolic. For distances much smaller than this length, all geometries look Euclidean.

In a spherical universe, as on the earth's surface, parallel lines eventually meet, triangles can span up to 540 degrees, and the circumference of a circle is smaller than $2\pi r$. Because the space curves back on itself, the spherical universe is finite. In a hyperbolic universe, parallel lines diverge, triangles have less than 180 degrees, and the circumference of a circle is larger than $2\pi r$. Such a universe, like Euclidean space, is infinite in size. (There are ways to make hyperbolic and flat universes finite, but they do not affect the conclusions of inflationary theory.)

These three geometries have quite different effects on perspective (*right*), which distort the appearance of features in the cosmic microwave background radiation. The largest ripples in the background have the same absolute size regardless of the specific process of inflation. If the universe is flat, the largest undulations would appear to be about one degree across. But if the universe is hyperbolic, the same features should appear to be only half that size, simply because of geometric distortion of light rays.

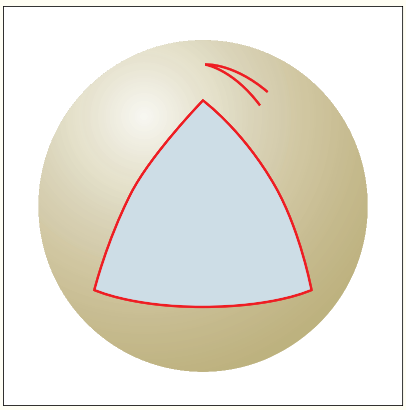
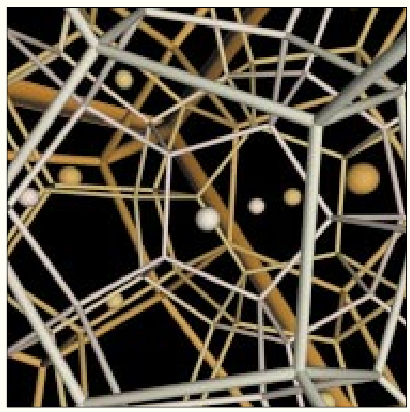
Preliminary observations hint that the ripples are indeed one degree across [see News and Analysis, "The Flip Side of the Universe," by George Musser; SCIENTIFIC AMERICAN, September 1998]. If confirmed, these results imply that the open inflationary theory is wrong. But tentative findings are often proved wrong, so astronomers await upcoming satellite observations for a definitive answer. —*M.A.B. and D.N.S.*

THREE GEOMETRIES are shown here from two different perspectives: a hypothetical outside view that ignores, for the sake of illustration, one of the spatial dimensions (*left column*) and an inside view that shows all three dimensions as well as a reference framework (*right column*). The outside view is useful for seeing the basic geometric rules. The inside view reveals the apparent sizes of objects (which, in these diagrams, are the same actual size) at different distances. Here objects and framework redden with distance.

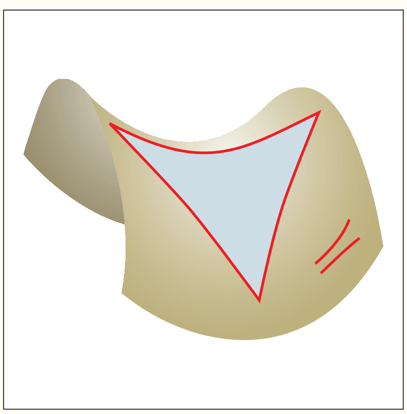
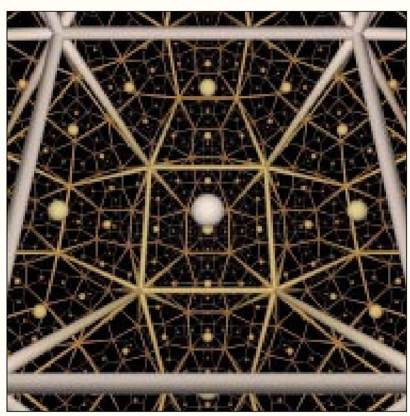



STUART LEVY *National Center for Supercomputing Applications* AND TAMARA MUNZNER *Stanford University*

Flat space obeys the familiar rules of Euclidean geometry. The angular size of identical spheres is inversely proportional to distance—the usual vanishing-point perspective taught in art class.

Spherical space has the geometric properties of a globe. With increasing distance, the spheres at first seem smaller. They reach a minimum apparent size and subsequently look larger. (Similarly, lines of longitude emanating from a pole separate, reach a maximum separation at the equator and then refocus onto the opposite pole.) This framework consists of dodecahedra.

Hyperbolic space has the geometry of a saddle. Angular size shrinks much more rapidly with distance than in Euclidean space. Because angles are more acute, five cubelike objects fit around each edge, rather than only four.

How Did the Universe Begin?

The laws of physics generally describe how a physical system develops from some initial state. But any theory that explains how the universe began must involve a radically different kind of law, one that explains the initial state itself. If normal laws are road maps telling you how to get from A to B, the new laws must justify why you started at A to begin with. Many creative possibilities have been proposed.

In 1983 James B. Hartle of the University of California at Santa Barbara and Stephen W. Hawking of the University of Cambridge applied quantum mechanics to the universe as a whole, producing a cosmic wave function analogous to the wave function for atoms and elementary particles. The wave function determines the initial conditions of the universe. According to this approach, the usual distinction between future and past breaks down in the very early universe; the time direction takes on the properties of a spatial direction. Just as there is no edge to space, there is no identifiable beginning to time. In an alternative hypothesis, Alexander Vilenkin of Tufts University proposed a "tunneling" wave function determined by the relative probabilities for a universe of zero size to become a universe of finite size of its own accord.

Last year Hawking and Neil G. Turok, also at Cambridge, suggested the spontaneous creation of an open inflationary bubble from nothingness. This new version of open inflation bypasses the need for false-vacuum decay, but Vilenkin and Andrei D. Linde of Stanford University have challenged the assumptions in the calculation.

Linde has tried to skirt the problem of initial conditions by speculating that inflation is a process without beginning [see "The Self-Reproducing Inflationary Universe," by Andrei Linde; Scientific American, November 1994]. In the classical picture, inflation comes to an end as the inflaton field rolls down its potential. But because of quantum fluctuations, the field can jump up the potential as well as down. Thus, there are always regions of the universe—in fact, constituting a majority of its volume—that are inflating. They surround pockets of space where inflation has ended and a stable universe has unfolded. Each pocket has a different set of physical constants; we live in the one whose constants are suited for our existence. The rest of the universe carries on inflating and always has. But Vilenkin and Arvind Borde, also at Tufts, have argued that even this extension of inflation does not describe the origin of the universe completely. Although inflation can be eternal in the forward time direction, it requires an ultimate beginning.

J. Richard Gott III and Li-Xin Li of Princeton University recently proposed that the universe is trapped in a cyclic state, rather like a time traveler who goes back in time and becomes her own mother. Such a person has no family tree; no explanation of her provenance is possible. In Gott and Li's hypothesis, our bubble broke off from the cyclic proto-universe; it is no longer cyclic but instead is always expanding and cooling.

Unfortunately, it may be very difficult (though perhaps not impossible) for astronomers to test any of these ideas. Inflation erases almost all observational signatures of what preceded it. Many physicists suspect that a fuller explanation of the preinflationary universe—and of the origin of the physical laws themselves—will have to await a truly fundamental theory of physics, perhaps string theory. *—M.A.B. and D.N.S.*

adds another stage preceding standard inflation. Another theory describing even earlier times will be needed to explain the original creation of the universe [*see box at left*].

Life in a bubble universe has a number of interesting consequences (not to mention possibilities for science-fiction plots). For instance, an alien observer could safely pass from the outside to the inside of the bubble. But once inside, the observer (like us) could never leave, for doing so would require traveling faster than light. Another implication is that our universe is only one of an infinity of bubbles immersed in a vast, frothy sea of eternally expanding false vacuum. What if two bubbles collided? Their meeting would unleash an explosion of cosmic proportions, destroying everything inside the bubbles near the point of impact. Fortunately, because the nucleation of bubbles is an extremely rare process, such cataclysms are improbable. Even if one occurred, a substantial portion of the bubbles would not be affected. To observers inside the bubbles but at a safe distance, the event would look like a broiling-hot region in the sky.

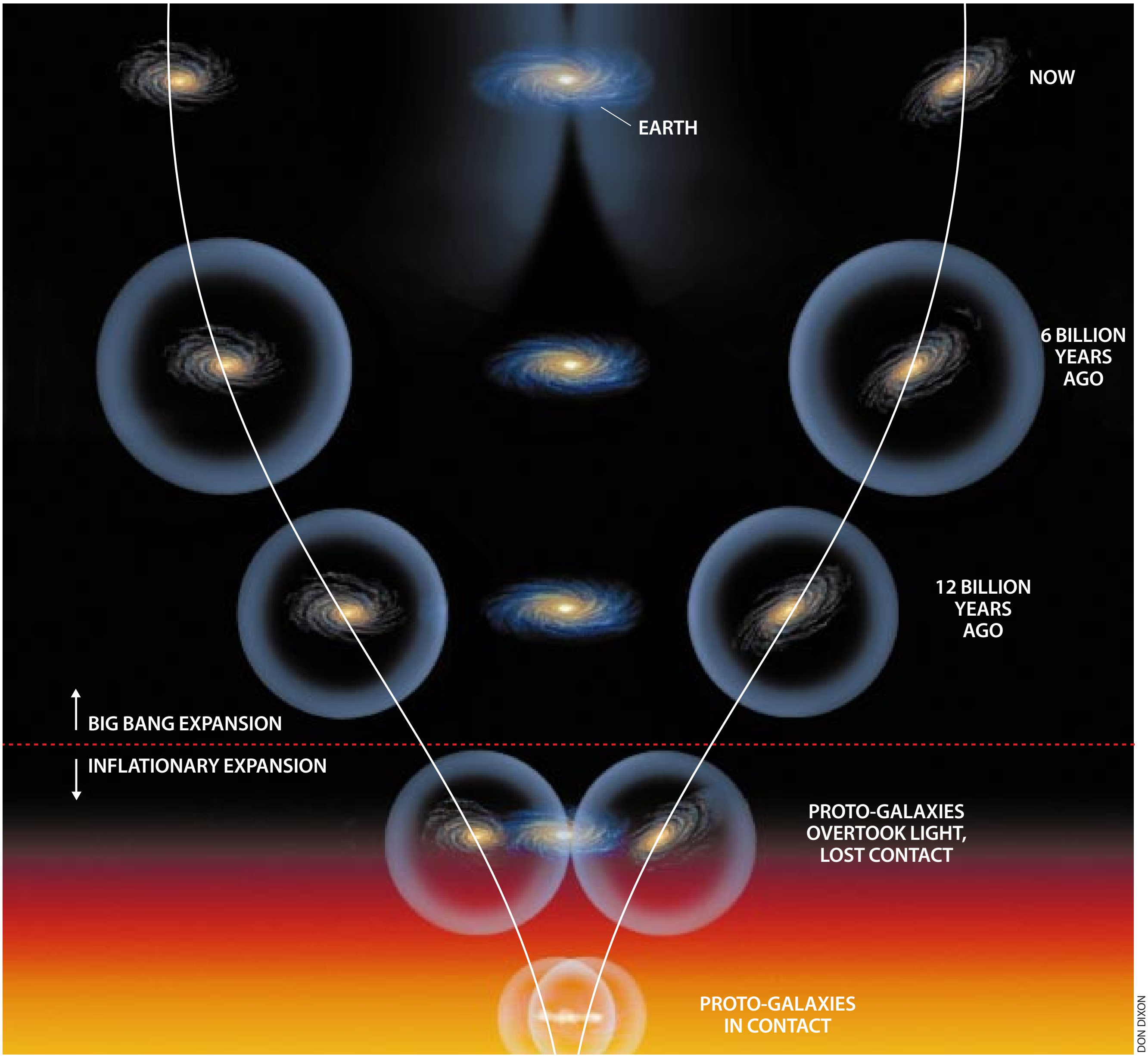
Corroborating Evidence

How does one test this theory? To explain why the universe is uniform is certainly a good thing. But validating a theory requires that some quantitative predictions be compared with observations. The specific effects of open inflation were calculated in 1994 with contributions by the two groups that refined the theory, as well as Bharat V. Ratra and P. James E. Peebles of Princeton.

Both the old and the new concepts of inflation make definite forecasts based on quantum effects, which caused different points in space to undergo slightly different amounts of inflation. When inflation ended, some energy was left over in the inflaton field and became ordinary radiation—the fuel of the subsequent big bang expansion. Because the duration of inflation varied from place to place, so did the amount of residual energy and therefore the density of the radiation.

The cosmic background radiation provides a snapshot of these undulations. In open inflation, it is affected not only by fluctuations that develop within the universe but also by ones that arise outside the bubble and propagate inside. Other ripples are set in motion by imperfections in the nucleation of the bubble. These patterns ought to be most notable on the largest scales. In effect, they allow us to look outside our bubble universe. In addition, one of us (Spergel), working with Marc Kamionkowski, now at Columbia University, and Naoshi Sugiyama of the University of Tokyo, realized that open inflation should have other, purely geometric effects [*see box on preceding page*].

At the current level of precision, the observations cannot distinguish between the predictions of the two inflationary theories. The moment of truth will come with the planned deployment late next year of the Microwave Anisotropy Probe (MAP) by the National Aeronautics and Space Administration. A more advanced European counterpart, Planck, is due for launch in 2007. These satellites will perform observations similar to those of the Cosmic Microwave Background Explorer (COBE) satellite nearly a decade ago, but at much higher resolution. They should be able to pick out which theory—either the cosmological constant or open inflation—is correct. Or it could well turn out that neither fits, in which case researchers will have to start over and find some new ideas for what happened in the very early universe. SA



MAJOR PARADOX in cosmology is the near uniformity of the universe. In the normal big bang expansion, such regularity is impossible (*upper part of diagram*). Billions of years ago two galaxies on opposite sides of the sky began to shine. Although the universe was expanding, the light was able to overtake other galaxies and finally reach us in the Milky Way. Humans, viewing the galaxies through telescopes, remarked that they looked much the same. Yet light from either galaxy had not yet arrived at the other. How, without seeing each other, could the two have harmonized their appearance? Inflation (*lower part*) provides an answer. In the first split second of cosmic history, the predecessors of the galaxies were touching. Then the universe expanded at an accelerating rate, pulling them apart at faster than the speed of light. Ever since, the galaxies have been unable to see each other. When inflation ended, light began to overtake them again; after billions of years, the galaxies will come back into contact.

The Authors

MARTIN A. BUCHER and DAVID N. SPERGEL study the physics of the very early universe. Bucher, currently in the department of applied mathematics and theoretical physics at the University of Cambridge, is one of the pioneers of open inflationary theory. Spergel is professor in the department of astrophysical sciences at Princeton University. The father of two young children, Julian and Sarah, he is a member of the science team for the Microwave Anisotropy Probe, a satellite designed to map the cosmic microwave background with high resolution. Both authors also work on such matters as galaxy formation and structure and defects in the fabric of space and time.

Further Reading

THE SELF-REPRODUCING UNIVERSE. Eugene F. Mallove in *Sky & Telescope,* Vol. 76, No. 3, pages 253–256; September 1988.

OPEN UNIVERSE FROM INFLATION. Martin Bucher, Alfred S. Goldhaber and Neil Turok in *Physical Review D,* Vol. 52, No. 6, pages 3314–3337; September 15, 1995. Preprint at xxx.lanl.gov/abs/hep-ph/9411206 on the World Wide Web.

THE VERY FIRST LIGHT: THE TRUE INSIDE STORY OF THE SCIENTIFIC JOURNEY BACK TO THE DAWN OF THE UNIVERSE. J. Mather, with J. Boslough. BasicBooks, 1996.

THE INFLATIONARY UNIVERSE: THE QUEST FOR A NEW THEORY OF COSMIC ORIGINS. Alan H. Guth. Perseus Press, 1997.

Additional information on the Microwave Anisotropy Probe and Planck satellite can be found at map.gsfc.nasa.gov/ and astro.estec.esa.nl/SA-general/Projects/Planck/ on the World Wide Web.

Child Care among the Insects

Why do some insect parents risk their lives to care for their young?

by Douglas W. Tallamy

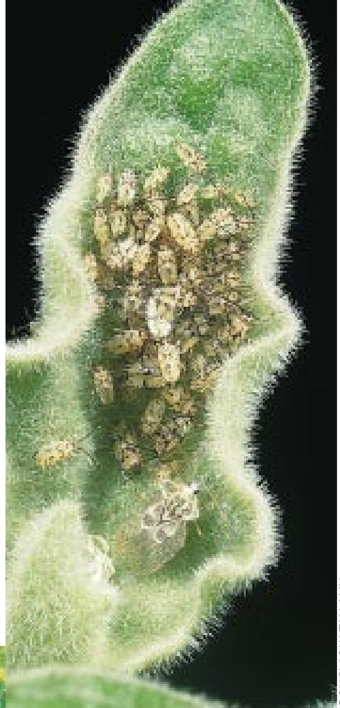
Photographs by Ken Preston-Mafham

RAIN FOREST NURSERY in Brazil features ants attending to a brood of young treehoppers (*Aetalion reticulatum*), while their mothers stand guard over freshly laid batches of eggs. The ants feed on a sweet secretion called honeydew produced by the nymphs and so defend them from predators. As a result, the adult treehoppers look after only their eggs, abandoning the young when they hatch to the ants' capable care.

Throughout the southeastern U.S., lace bugs of the genus *Gargaphia* live on horse-nettle plants. The female usually guards her eggs and, once they hatch, the nymphs that emerge. One fearsome enemy is the damsel bug: it brandishes a sharp, hard beak and, given a chance, will devour every last nymph. The lace bug has no such weapons: she diverts the damsel bug by fanning her wings and climbing on its back.

Meanwhile the nymphs rush to the midrib of the leaf and, using it as a highway, flee up the stem into a young, curled leaf, where they hide. If the mother can get away, she follows them and guards the stem of the leaf. There she can intercept the predator, which is likely to follow. Sometimes the mother may be able to fend off the attacker momentarily; in that case, she scurries to guide the nymphs to an ideal leaf by blocking a branch they might mistakenly follow. All too often, though, she dies in the attack, her sacrifice giving the nymphs time to escape with their lives.

HIDING IN FOLDS of a young leaf (*right*), nymphal lace bugs of the genus *Gargaphia* seek shelter from predators. The species is common in the southeastern U.S. In another episode (*below*), the mother faces off with the lethal larva of a lacewing, which, despite her efforts, is eating the nymphs.



DOUGLAS W. TALLAMY

DOUGLAS W. TALLAMY

POISED TO DEFEND her offspring is a Ugandan assassin bug (*Pisilus tipuliformis*), which watches over her emerging nymphs (*right*). Shield bugs *Cocoteris* (*center*) from New Guinea and *Antiteuchus* (*far right*) from Brazil are quite likely to lose to predators those young they cannot fit under their sheltering shields.

Swedish naturalist Adolph Modeer first described parental care in insects as early as 1764. He noticed that the female European shield bug, *Elasmucha grisea,* remained steadfast over her eggs and tilted her body toward attacking predators rather than taking flight. But as late as 1971 many scientists hotly contested the idea that some insects actively care for their young. Even those who accepted the observations assumed that parental care was an innovation that only the most sophisticated bugs had managed to achieve.

Such behavior is indeed analogous to that of "advanced" life-forms, such as birds and mammals. But caring for offspring is hardly a recent innovation. It is common in invertebrates, including mollusks, worms, rotifers and even jellyfish. Among arthropods, it is the rule for centipedes, spiders, scorpions, sea spiders and the likely closest relatives of insects, the crustaceans. In fact, the relative rarity of "parental" insects—they are scattered, seemingly at random, throughout 13 insect orders—seems to reflect its widespread loss from early lineages.

Still, the ecological penalties for parental care can be so severe for insects that some entomologists wonder why it has persisted at all. The far easier strategy, followed by most insects, is simply to produce an

JOSTLING FOR SAFETY underneath the body of their mother, larvae of the Brazilian tortoise beetle (*Acromis sparsa*) arrange themselves into a symmetrical ring (*left*). At the ends of their bodies are anal hooks on which they wave their feces, repelling incautious predators with a mouthful. The mother guards the offspring from the time they are eggs (*above*) and shepherds the hatchlings to food sources, taking care to round up stragglers.

FORMIDABLE MOTHER, the praying mantis *Oxyophthalmellus somalicus* (*left*) has positioned herself at the base of a twig in a Kenyan desert. There she can intercept predators interested in her nymphs. The *Galepsus* praying mantis (*below*) is also from Kenya; she has hidden her egg sac so that it blends in with the bark surfaces. But she stands guard in case the ruse fails.

abundance of eggs. In his widely acclaimed synthesis, *Sociobiology,* Edward O. Wilson described parental care as a response to unusually favorable or unusually harsh environments. He argued that it should be most prevalent when resources are rich, in which case competition is intense, or when food is difficult to obtain or process, when physical conditions are particularly harsh or when predation is severe.

A Rich, Rough World

Burying beetles and dung beetles around the world have responded to competition for unusually nutritious but ephemeral resources such as carrion and dung by evolving a specific form of child care. Either the female alone or both parents secure the resource in an underground chamber as quickly as possible to protect it from competitors and from drying out. A pair of *Nicrophorus* carrion beetles, for instance, might bury a small dead rodent and then mold it into a cup that will hold and nourish the young. When the larvae hatch, the female—and occasionally the male—supplements their diet with regurgitated liquids. Michelle P. Scott of the University of New Hampshire and Gonzalo Halffter of the Institute of Ecology in Veracruz, Mexico, have shown that the males of such species prevent other males from usurping their prize and from killing their offspring.

Parents can also process food for the young. For example, *Sehirus* burrowing bugs provide their delicate nymphs, hidden within a soil depression, with seeds. *Umbonia* treehoppers expose plant phloem tubes, those that carry nutrients, to tiny nymphs by cutting a series of spiral slits in the bark. Wood eaters face the challenge of converting a tough, indigestible food source that is unusually low in nitrogen into a form that their young can use. *Cryptocercus* wood roaches and passalid bess beetles solve this problem by feeding the offspring directly from the anus with macerated wood fibers or with protozoans (which colonize the intestines and break down cellulose), feces and gut fluids that may be high in nitrogen. Bark beetles, on the other hand, chew tunnels within which they lay eggs and inoculate the excavated wood chips with symbiotic fungi that convert the cellulose to digestible forms for the larvae.

Insect caregivers typically protect only the eggs, but in some species one or both parents will defend the young as well. In that case, the parent and offspring must communicate extensively and coordinate their movement. *Gargaphia* lace bugs, sawflies, tortoise beetles and fungus beetles protect their larvae as they forage for food. A mother can guard only offspring that remain in a single group, so she herds them together by blocking the paths of the wayward ones.

As these examples suggest, mothers are most likely to care for their young. On rare occasions, however, the fathers take over, permitting the species to use habitats that would otherwise be too inhospitable. Water bugs, for instance, have large eggs that are in danger of drying out if laid above water or of drowning if laid within. Somehow, the eggs have to be moistened and aired.

In a primitive group of giant water bugs called *Lethocerus,* the female lays eggs on a stick above the water. The male repeatedly dives into the water and climbs out to drip onto the eggs to keep them moist; he also

HARLEQUIN STINK BUG (*Tectocoris diophthalmus*) from Australia defends her eggs aggressively. Because she lays only one batch, they are her sole chance for reproductive success.

drives off predators. But male *Belostoma* giant water bugs (often seen in swimming pools) instead carry the eggs, which the females glue onto the males' backs. A male has to keep floating to the surface and exposing these to air. He moves his hind legs back and forth or holds on to a twig and does push-ups for hours to keep aerated water flowing over the eggs. Similarly, *Bledius* rove beetles, *Bembidion* ground beetles and *Heterocerus* marsh-loving beetles all prevent their eggs from drowning within tidal mudflats by plugging their narrow-necked brood chambers when the tide is in and removing the plugs when waters recede.

The Cost of Care

Wilson has undoubtedly identified conditions that promote parental behavior in insects. Still, one wonders why some insects meet these challenges by caring for their young, whereas other species—even close relatives—reproduce under the same conditions using other strategies. One approach to this question is a simple cost-benefit analysis.

Both males and females can pay severe penalties for confronting, rather than fleeing from, predators. Such risks are difficult to quantify, and data are scarce. But when I measured the chances of *Gargaphia* lace bugs surviving the predation of jumping spiders, the mothers guarding nymphs were three times less likely to survive than females without such responsibilities.

Care is costly also because—with rare exceptions—it restricts parents to the site of the nest. Eggs are outstandingly expensive to produce, and mothers standing guard over their first clutch cannot forage for the nutrients that a new batch of eggs would require. This trade-off in fecundity can be substantial: *Gargaphia* females that are experimentally restricted from caring for eggs lay more than twice as many eggs as females that guard their young.

Such high costs have on occasion prompted alternative behaviors even within the same species. Some *Gargaphia* lace bugs and *Polyglypta* treehoppers dodge the risks and losses of guarding their young by laying eggs in the egg masses of other females of their species whenever possible. If they succeed, these "egg dumpers" are free to lay a second clutch almost immediately, whereas the recipients cannot resume laying until their first eggs hatch (in *Polyglypta*) or until their nymphs reach adulthood (in *Gargaphia*). If a *Gargaphia* female has no opportunity to dump her eggs, she reduces the risk by defending her young aggressively only when she is old and has little to lose or when her nymphs are in the final stages of development and have a good chance of reaching maturity.

Catherine M. Bristow of Michigan State University has found that *Pubilia* treehoppers limit maternal costs in a different way. A mother remains with her young until ants discover the group and begin to eat the sugary secretion, called honeydew, produced by the nymphs. Then the mother abandons them, transferring care of her young to the very capable ants, which defend the nymphs from predators.

As substantial as parental costs are for females, they are typically prohibitive for males. Physiologically, sperm are cheap. So although baby-sitting means less time for foraging, the reduction in nutrients should not hinder a male's ability to manufacture sperm. The trade-off is instead in the loss of promiscuity: when committed to guarding one batch of offspring, a male is no longer free to roam for additional females and to father many more young.

Exacerbating this loss is the inability of most insect males to guarantee their paternity. Females usually can store sperm and can even choose that of one male over another within their bodies. Such uncertainty about who fathered the eggs makes paternal care a dubious investment for most males.

Not surprisingly, exclusive paternal care is extremely rare in insects, occurring only in three families of true bugs. In a few species of assassin bugs, in even fewer leaf-footed bugs and in all giant water bugs, males manage to avoid the costs of care.

MALE GIANT WATER BUG (*Abedus herberti*) from Sycamore Canyon in Arizona carries around the eggs he has fertilized, which are glued onto his back by the female. The male—one of very few paternal insects—goes to great lengths to keep the eggs moist and aerated.

C. ALLAN MORGAN *Peter Arnold, Inc.*

Rhinocoris assassin bugs, for instance, make a display of their attention to an egg mass. Neighboring females seem to assume that a male that is already attending to eggs has a commitment to such behavior and seek him out for matings. Because females refuse to mate with males that are not guarding (except early in the season, when few eggs have been laid), males will fight over egg masses to protect. Such behavior pays off because the females lay eggs right after they mate, or even while mating, so that the male is accepting care for ones he more certainly fathered.

Moreover, in both water bugs and paternal assassin bugs, the density of females seeking males is high. Lisa Thomas, then at the University of Cambridge, has found that *Rhinocoris tristis,* a paternal assassin bug from Kenya, lives only on *Stylosanthes* plants; it drinks nectar from the flowers and hunts insect prey around them. Because the bugs concentrate on a particular host, and females can easily find males for guard duty, the latter do not suffer reduced promiscuity. Robert L. Smith of the University of Arizona offers similar arguments concerning the relatively dense populations of giant water bugs confined to ponds: the male does not need to roam, for the females come to him.

The Last Resort

Notwithstanding all these instances, the vast majority of insects avoid the costs of parental care by resorting to a variety of mechanisms by which eggs can survive. Piercing ovipositors (swordlike appendages used for laying eggs) or hard, impenetrable egg coatings allow many insects to hide their eggs in plant tissue or seal them out of harm's way in natural cracks and crevices. At the core of all such innovations is the development of iteroparous reproduction. Rather than laying all their eggs at one time—called semelparous reproduction—and then guarding them well, most insects have acquired the ability to lay them in many small clutches, thus spreading their eggs over time and space.

This strategy alone is a very effective way of cutting losses: if a predator discovers one clutch, it has access to only a small fraction of the total number of eggs laid by the mother. If *Gargaphia* lace bugs are, for instance, prevented from guarding their large clutches—typically well over 100 eggs—56 percent of those eggs are destroyed by predators before they hatch. In contrast, *Corythucha ciliata,* the sycamore lace bug, lays 33 small clutches rather than one large clutch and distributes them on many different leaves throughout its host. By this means, these bugs lose only 16 percent of their eggs to predators.

So then why have all insects not abandoned caregiving? Let us reconsider the cost analysis. A parent pays a substantial cost for caring only if it implies a loss in fecundity. Thus, if advancing winter or resources that are otherwise limited somehow prohibit future production of eggs, such costs do not enter the equation. Care may consequently become an effective option.

For example, females of the Japanese burrowing bug, *Parastrachia japonensis,* rear young only on fallen fruits of *Schoepfia* trees and must confine reproduction to the brief period when fruits are abundant. A female has food enough to produce only one large clutch, which she guards and provisions for weeks without sacrificing subsequent opportunities for reproduction.

Almost all parental species are thus constrained to no more than one clutch by seasonal change, ephemeral or scarce resources, or some other ecological limitation. If parental care in insects is viewed in terms of iteroparity and semelparity, the lack of pattern that has puzzled scientists for so long becomes easier to explain. For most insects, the opportunity to spread reproduction over time and space has made child care both prohibitively expensive and unnecessary. But for those with fewer chances to breed, it can be the only way to ensure that their offspring live on after them. SA

The Author

DOUGLAS W. TALLAMY studies the evolution of maternal and paternal care in arthropods, as well as the chemical interactions between insects and plants. He obtained his Ph.D. in entomology from the University of Maryland in 1980, served as a postdoctoral associate at the University of Iowa and is now a professor at the University of Delaware. Many of Tallamy's studies feature the *Gargaphia* lace bug, whose behavior he is able to observe in his backyard.

Further Reading

INSECT PARENTAL CARE. Douglas W. Tallamy in *BioScience,* Vol. 34, No. 1, pages 20–24; January 1984.

COSTS AND BENEFITS OF THE EGG-DUMPING ALTERNATIVE IN *GARGAPHIA* LACE BUGS (*HEMIPTERA: TINGIDAE*). D. W. Tallamy and L. A. Horton in *Animal Behavior,* Vol. 39, pages 352–360; February 1990.

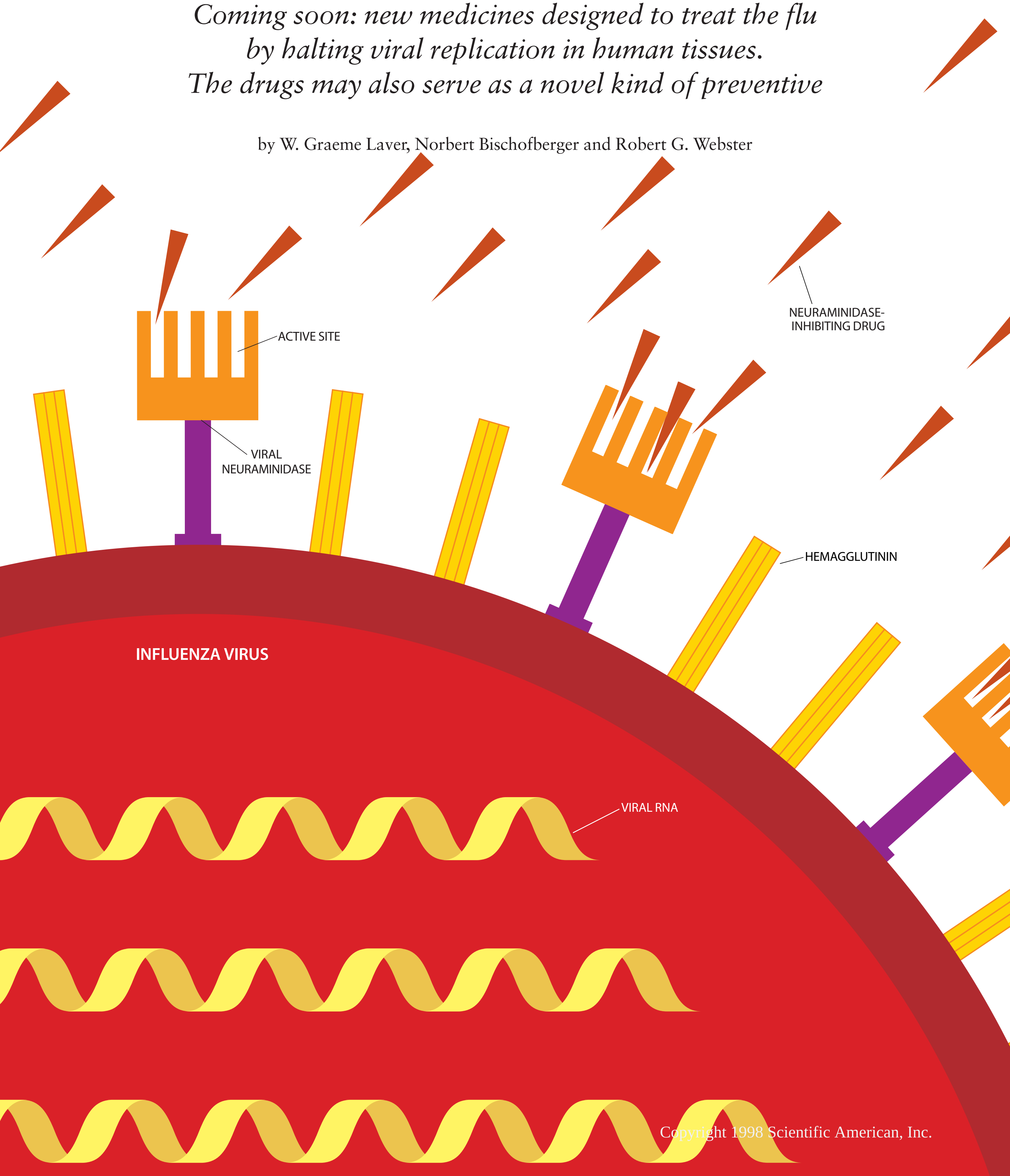
THE EVOLUTION OF PARENTAL CARE. T. H. Clutton-Brock. Princeton University Press, 1991.

THE EVOLUTION OF SOCIAL BEHAVIOR IN INSECTS AND ARACHNIDS. Edited by Jae C. Choe and Bernard J. Crespi. Cambridge University Press, 1997.

Disarming Flu Viruses

Coming soon: new medicines designed to treat the flu by halting viral replication in human tissues. The drugs may also serve as a novel kind of preventive

by W. Graeme Laver, Norbert Bischofberger and Robert G. Webster



Every so often, a strain of influenza unfamiliar in humans suddenly begins passing from person to person. Because the virus is so unusual, few if any people have built-in immunity from past exposures. Even the vaccinated have no defense; flu shots shield against influenza variants that health experts have anticipated will be active in a given flu season, not against other, unforeseen kinds. Finding no deterrent, the new strain spreads unabated, causing illness—and death—on a global scale.

The worst worldwide epidemic, or pandemic, on record struck in 1918 and killed more than 20 million people, sometimes within hours after the first symptoms appeared. This disaster, traced to the so-called Spanish influenza virus, was followed by epidemics of Asian flu in 1957, Hong Kong flu in 1968 and Russian flu in 1977. (The names reflect popular impressions of where the pandemics began, although all four episodes, and perhaps most others, are now thought to have originated in China.)

Public health experts warn that another pandemic can strike any time now and that it could well be as vicious as the 1918 episode. In 1997, when a lethal influenza variant afflicted 18 people in Hong Kong, contributing to the death of six, officials feared the next wave had begun. Authorities in the region managed to contain the problem quickly, however, by finding the source—infected chickens, ducks and geese—and then destroying all the poultry in Hong Kong.

Next time, humankind may not be so fortunate. If a virus as deadly as that Hong Kong strain tore through the world's crowded communities today, 30 percent of the earth's population could conceivably be dead (from the virus itself or from secondary bacterial infections) before a vaccine became available to protect those who initially managed to escape infection. Vaccines against any given influenza variant take about six months to produce, test for safety and distribute—too long to do much good in the face of a fast-moving pandemic.

If the feared pandemic does not materialize until next year or beyond, though, new methods for limiting sickness and death could be available. Later this year two drugs being tested in large clinical trials could be approved for sale as new missiles in the fight against the flu. The agents—called zanamivir (Relenza) and GS 4104—show great promise for preventing influenza infections and for reducing the duration and severity of symptoms in people who begin treatment after they start to feel sick.

Unlike vaccines (which prime the immune system to prevent viruses from gaining a foothold in the body) and unlike standard home remedies (which ease symptoms but have no effect on the infection itself), these drugs have been designed to attack the influenza virus directly. They hobble a critical viral enzyme, called neuraminidase, and in so doing markedly reduce proliferation of the virus in the body. Additional neuraminidase inhibitors, not yet evaluated in humans, are under study as well.

As many people know, two anti-flu drugs, amantadine and rimantadine, are already on the market. But those agents, which work by a different mechanism, have serious flaws. They can cause confusion and other neurological side effects, and they are ineffective against one of the two major influenza classes (type B) that afflict people. Moreover, influenza viruses seem to become resistant to the drugs fairly easily. Therefore, individuals treated in the first phases of an epidemic can spread a drug-resistant version of the virus to other people, who will then prove unresponsive to the medicines. This last problem is particularly acute in "closed" communities, such as nursing homes.

The story of how the newer drugs were developed involves a wonderful combination of luck and logic. The breakthrough that led most immediately to their design was the deciphering, in 1983, of neuraminidase's three-dimensional structure. Yet it was a series of earlier discoveries that enabled scientists to realize that a specific part of the neuraminidase molecule was probably an Achilles' heel for all influenza variants—a weakness that thoughtfully constructed drugs might exploit.

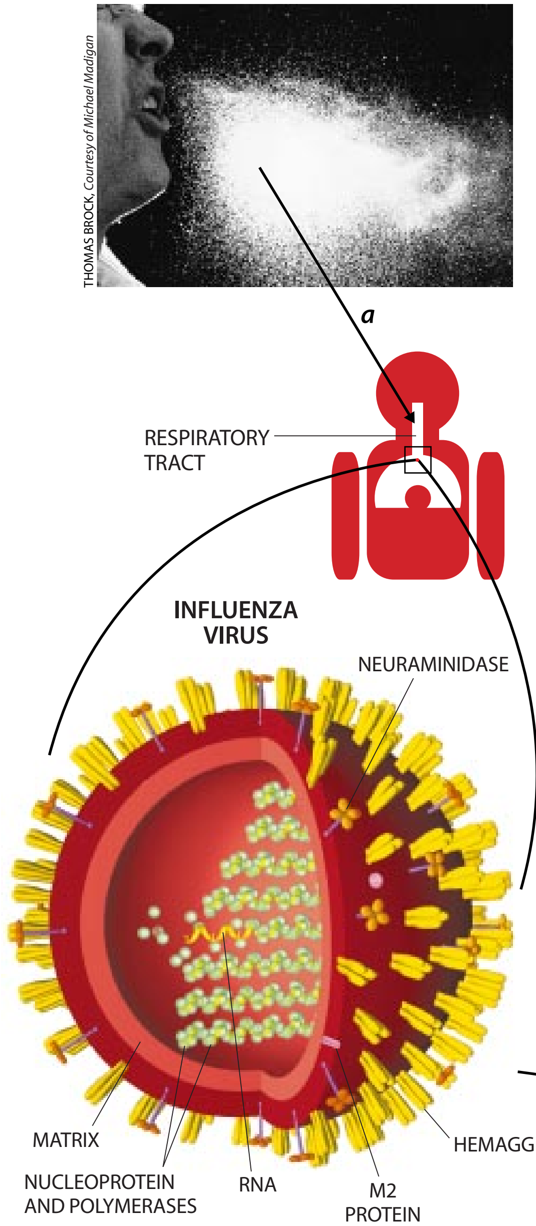
BRYAN CHRISTIE

"PLUG DRUGS" (*red wedges*) are showing great promise for preventing and treating the flu. They work by plugging the active, catalytic site of an enzyme called neuraminidase that protrudes from the surface of influenza viruses. This enzyme enables newly formed viral particles to travel from one cell to another in the body. With the enzyme inactivated, the virus is stopped in its tracks.

Understanding a Scourge

One line of that early research uncovered some essential properties of influenza viruses and of their strategy for survival. Biologists have long known that viruses are basically genes wrapped in proteins that either protect the genes or help the viruses to reproduce in the body. Sometimes, as is true for influenza, these various constituents are further enveloped in a fatty (lipid) membrane. When any virus causes disease, it does so by invading selected cell types, replicating within those cells and then pouring out of the cells to infect others. Symptoms arise both because viral proliferation disrupts the colonized cells and because the immune system attempts to contain the infection, in the process causing local inflammation and systemic aches and fever.

Influenza strains that colonize humans have a particular affinity for the



epithelial cells that form the lining of the respiratory tract. Successful infection typically leads after a day or two to such classic symptoms as runny or stuffy nose, dry cough, chills, fever, aches, deep tiredness and loss of appetite. Historical descriptions based on symptoms indicate that flu epidemics have probably plagued human populations since well before the 5th century B.C.

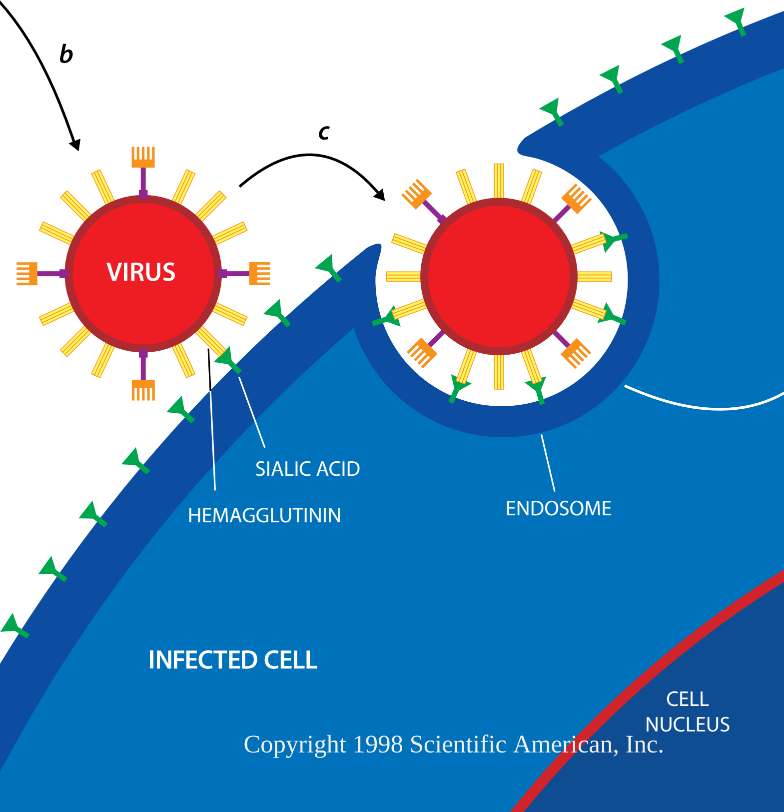
Scientists isolated an influenza strain from a human for the first time in 1933. Since then, they have learned that influenza viruses come in two main "flavors"—types A and B—that differ in certain of their internal proteins. A third type (C) does not seem to cause serious disease.

Virologists further group type A forms according to variations in two proteins that protrude from the viral surface like spikes—hemagglutinin and neuraminidase (the enzyme that is the target of the new drugs). As is true of other proteins, these consist of folded strings of amino acids. All hemagglutinin variants adopt essentially the same three-dimensional conformation, and all neuraminidase variants take on a characteristic shape. But within each group, the individual proteins can differ markedly in the sequence of their constituent amino acids. So far 15 hemagglutinin and nine neuraminidase subtypes have been identified on type A influenzas, which are named according to the hemagglutinin and neuraminidase molecules they display: H1N1, H1N2, H2N2 and so on.

Type B viruses are a more uniform lot. They carry one form of hemagglutinin and one of neuraminidase, although the amino acid sequences can differ slightly from one B strain to another. Similarly, each influenza A subtype can also come in slightly varying strains.

Aside from their chemistry, type A and type B influenzas differ in their range of activity. Type B viruses infect only humans, and they cause regional epidemics rather than pandemics. Type A influenzas, in contrast, affect pigs, horses, seals, whales and birds as well as humans, although not all strains infect all species. (Indeed, only four subtypes have been found in humans.) They are also responsible for all of this century's pandemics.

In spite of their differences, both influenza types have the same basic life cycle. For a single copy of an influenza virus, or particle, to enter a human cell, hemagglutinin on the virus must link to a sugary molecule, sialic acid, on the surface of the cell. This binding induces the cell to engulf the virus, which is initially sequestered within a kind of bubble. Soon, though, the viral genes (consisting of strands of RNA) and internal pro-



LIFE CYCLE of the influenza virus often involves transmission from one person's airways to another's via water droplets emitted during a sneeze (*a*). An individual virus (*detail, above*) enters a cell lining the respiratory tract after a molecule called hemagglutinin on the virus binds to sialic acid on the cell (*b*). This binding induces the cell to take up the virus (*c*), which soon dispatches its genetic material, made of RNA, and its internal proteins to the nucleus (*d* and *e*). Some of those proteins then help to duplicate the RNA (*f*) and to produce messenger RNA, which the cell's protein-making machinery uses as a template for making viral proteins (*g* and *h*). Next, the viral genes and proteins assemble into new viral copies, or particles (*i*), and bud from the cell.

The particles emerge coated with sialic acid. If that substance remained on the virus and on the cell, the hemagglutinin molecules on one particle would soon attach to the sialic acid on other particles and on the cell, causing the new viruses to clump together and stick to the cell. But neuraminidase on the virus clips sialic acid from the offending surfaces (*j*), leaving the new particles free to travel on (*k*) and invade other cells.

teins are freed, and they work their way into the cell nucleus.

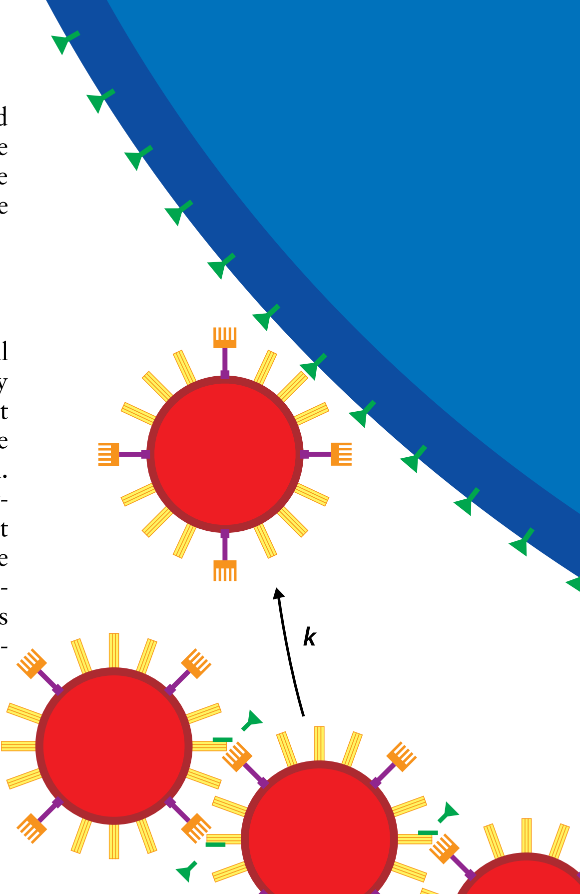
There some of the viral proteins set about replicating the viral RNA strands and also constructing a form (called messenger RNA) that can be read out and translated into proteins by the cell's protein-making machinery. Eventually the newly made genes and proteins come together and bud from the cell as new viral particles.

Inconveniently for the virus, the emerging particles are coated with sialic acid, the very substance that binds influenza viruses to the cells they attempt to invade. If the sialic acid were allowed to remain on the virus and on the surface of a virus-making cell, hemagglutinin on the newly minted particles would bind to the sialic acid, causing the particles to clump together on the cell, like insects trapped on flypaper. So trapped, they would be unable to spread to other cells.

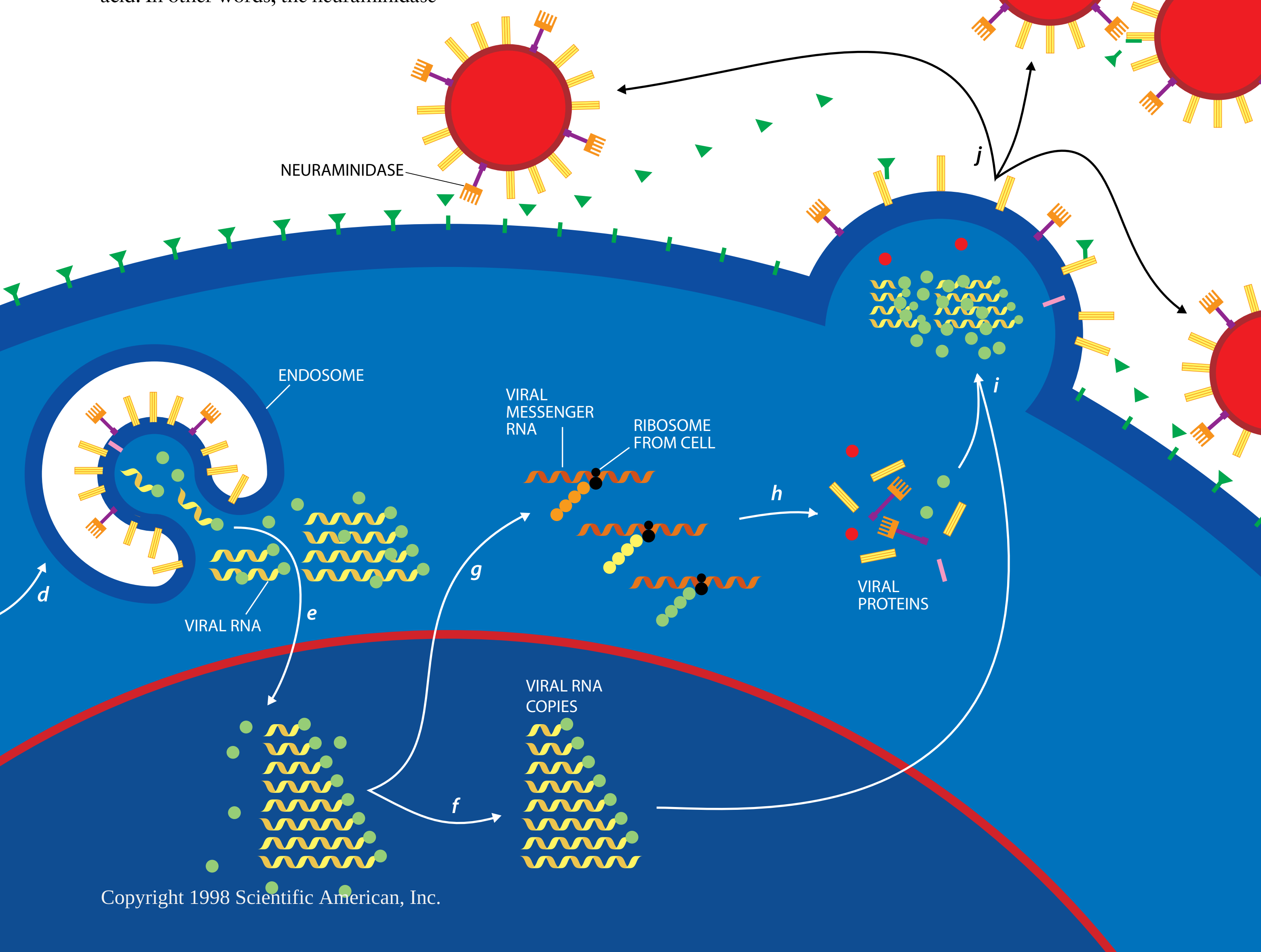
But the virus has an ace in the hole. The neuraminidase molecules on the freshly made particles can cleave sialic acid. In other words, the neuraminidase spikes essentially dissolve the unwanted sialic acid "glue," thereby enabling the viral particles to travel. The enzyme also helps the virus to plow through the mucus between cells in the airways.

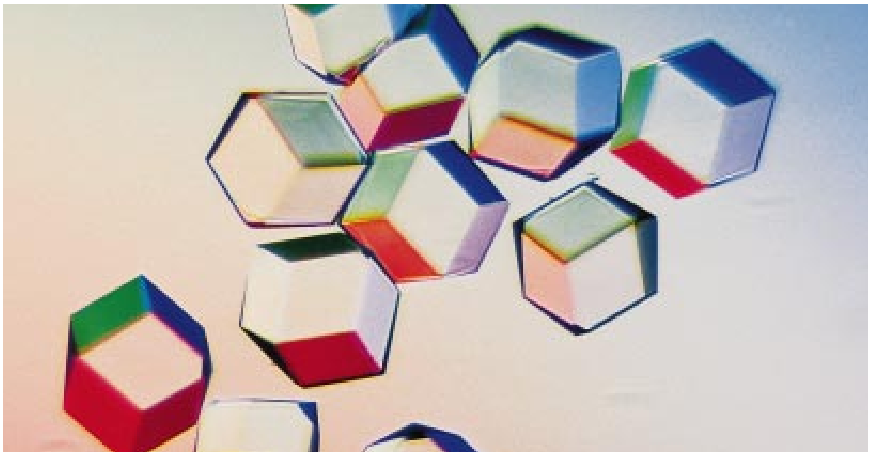
The Roots of Pandemics

By the 1960s, investigators were well aware that a drug able to block any step in the replication process might prevent the virus from causing disease or might curtail an existing infection. But no one knew precisely how to intervene. Moreover, biologists realized that because influenza viruses grow inside cells and make use of the cells' own protein-making machinery, most agents able to destroy the virus would also disturb healthy cells.



BRYAN CHRISTIE





STUART BUTTERWORTH AND W. GRAEME LAVER

CRYSTALS OF NEURAMINIDASE were obtained from an influenza virus found to infect wild birds on the Great Barrier Reef of Australia. Neuraminidase crystals have enabled scientists to determine the three-dimensional structure of the enzyme and to build drugs designed to plug its active site. The hues are reflections of colored lights.

As workers confronted this difficulty, they also continued trying to understand why some strains of influenza produce localized epidemics but others cause full-fledged pandemics. This research would eventually reveal that to be most useful, an influenza drug would have to be able to attack all influenza variants, including ones not known to cause disease in humans.

An influenza strain can produce a local or global epidemic only if the people exposed to the virus lack adequate immunity to it. For instance, when someone has the flu, the immune system produces molecules known as antibodies that recognize specific segments of hemagglutinin and neuraminidase on the viral surface. If the person is later reexposed to the same strain, these antibodies (primarily those directed at hemagglutinin) bind promptly to the virus and prevent it from causing a repeat infection.

If the virus never changed, as is the case with measles or mumps, the antibodies raised during infection or by a vaccine could provide durable immunity. But influenza viruses revise themselves all the time. In consequence, antibodies elicited in one year may well be less effective or useless if they encounter a differing form of the virus in the next flu season. The extent of change largely determines whether an epidemic becomes relatively contained or extends across the globe unchallenged.

One way influenza viruses change is by antigenic "drift," gradual revision of the amino acid sequence in a protein (antigen) able to elicit an immune response. These alterations arise through small mutations in the gene that constitutes the blueprint for that protein. Sometimes a mutation makes little difference in the protein's stability or activity. Sometimes it damages the protein and reduces the viability of the virus. Other times, though, it enhances survival, such as by reconfiguring a site on hemagglutinin that was formerly recognized by an antibody.

When the hemagglutinin or neuraminidase genes and proteins accumulate several alterations, they can become virtually unrecognizable to most of the antibodies in a population and may initiate a new epidemic. The epidemic finds boundaries, however, when it reaches groups whose immune systems have already "seen" many of the alterations before.

Influenza B viruses seem to change exclusively through such antigenic drift, evolving gradually in their human hosts as they attempt to become less recognizable to the immune repertoire of a population. Influenza A strains, in contrast, can additionally undergo a more dramatic change, known as antigenic "shift," that enables them, alone, to cause pandemics.

When antigenic shift occurs, strains crop up bearing a totally new hemagglutinin spike, and sometimes also a new neuraminidase molecule, that most people have never encountered. As a result the virus may evade the antibody repertoire carried by all populations around the globe and trigger a pandemic. In today's jet-linked world, people can spread a dangerous new virus from one part of the earth to another in a day.

Such a drastic metamorphosis cannot occur through simple genetic mutation. The best-studied process leading to antigenic shift involves the mixing of two viral strains in one host cell, so that the genes packaged in new viral particles (and their corresponding proteins) come partly from one strain and partly from the other. This reassortment can take place because the genome, or genetic complement, of the influenza virus consists of eight discrete strands of RNA (each of which codes for one or two proteins). These strands are easily mixed and matched when new influenza A particles form in a dually infected cell. For instance, some influenza viruses infect both people and pigs. If a pig were somehow invaded by a human virus and by a strain that typically infected only birds, the pig might end up producing a hybrid strain that was like the human virus in every way except for displaying, say, a hemagglutinin molecule from the bird virus.

Scientists have recently learned that antigenic shift can also occur in a second way. In this case, an animal influenza virus that has not previously been able to produce infections in people makes a direct leap into human beings.

No one knows which form of antigenic shift led to the Spanish flu pandemic of 1918, which was caused by the H1N1 subtype of influenza A. Reassortment has, however, been proved to account for the 1957 Asian flu and 1968 Hong Kong flu pandemics, which were triggered, respectively, by H2N2 and H3N2. Some work suggests that aquatic birds might have contributed the unfamiliar genes and that pigs probably served as the mixing vessels. If pigs do sometimes serve this function, their involvement might help to explain why pandemics commonly originate in China: millions of birds, pigs and people live closely there.

The virus that killed six people in Hong Kong in 1997 (H5N1), in contrast, was not a reassortment virus. It caused human disease after jumping directly from birds to people—a phenomenon that had never been seen before. H5N1 was unable to pass from human to human. Had it been given time to acquire transmissibility through mutation or reassortment, though, it might have become uncontrollable quickly.

The close call of 1997 has now convinced many public health experts that influenza cases need to be monitored not only in people (as is done now) but also in animals. Those animals should certainly include migratory birds, because they probably serve as a year-round reservoir of influenza A viruses that then spread to domestic birds and other species. Prompt identification of animal strains with potential for harming people could help avert a public health disaster.

The incident in Hong Kong has also lent new urgency to research into the nature of the so-called species barrier that prevents many influenza strains from crossing from one kind of animal to another. If the barrier were better understood, scientists might be able to seal the leaks that now allow certain animal strains to breach the barrier and cause human disease.

Eureka

Taken together, studies of influenza biology conducted before the early 1980s indicated that, in addition to blocking the activity of some molecule involved in the virus's reproductive cycle, an ideal anti-flu drug would do so by acting at a "conserved" site on the targeted molecule. That is, it would home in on an area formed from amino acids that are held constant across all strains of the virus. A drug that targeted a conserved region would presumably work against any influenza virus that turned up in people, including ones that spread abruptly from animals.

Interestingly, the structural work that enabled researchers to design neuraminidase inhibitors grew out of an accidental discovery. Back in the late 1970s, one of us (Laver) was attempting to determine whether the N2 spike on the virus that caused the 1968 Hong Kong flu pandemic (H3N2) had come from the strain responsible for the 1957 Asian flu pandemic (H2N2). As part of that effort, he wanted to compare the amino acid sequences of the molecules. To start, he had to isolate and concentrate their heads—the domains that protrude from the viruses.

When Laver freed the neuraminidase heads from purified viruses and concentrated them in a centrifuge, he found to his surprise that the resulting pellet of material was not amorphous, as proteins usually are. Instead it consisted of crystals. Crystals, which are highly ordered arrays of molecules, are essential for deciphering the three-dimensional structure of large proteins. Hence, the unexpected production of neuraminidase crystals implied that the structure of neuraminidase could perhaps be deciphered.

In 1983 Peter Colman and his colleagues at the Commonwealth Scientific and Industrial Research Organization (CSIRO) in Australia did just that. The work revealed that the neuraminidase spikes on influenza viruses consist of four identical molecules, or monomers. The resulting tetramer resembles four squarish balloons atop a single stick. The stick is embedded in the viral membrane, and the balloons protrude. Colman's group soon discovered that each neuraminidase monomer in the foursome has a deep, central dent, or cleft, on its surface.

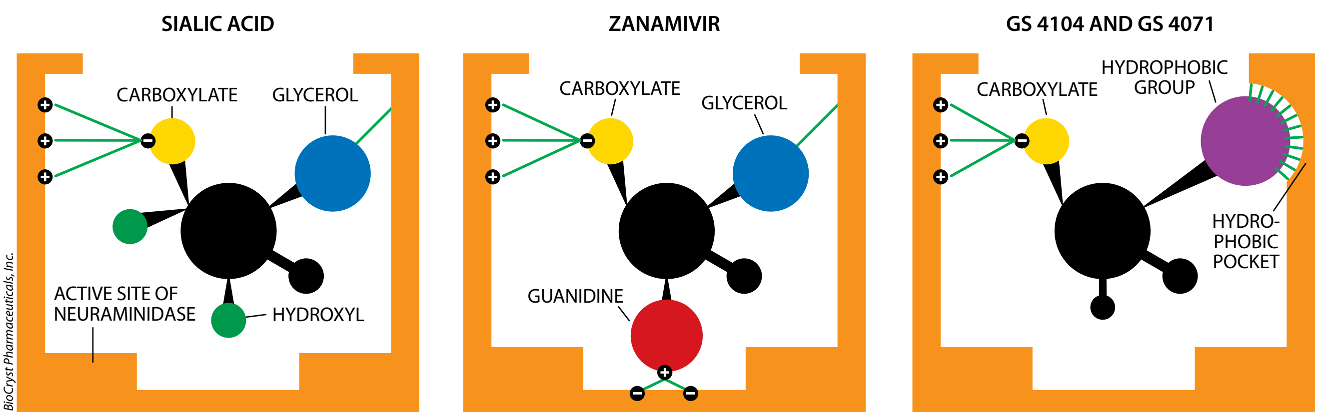
The team also found that even though influenza neuraminidase molecules could differ in the precise amino acids they contained, all known versions—including those from type A and from type B viruses—had a striking commonality. The amino acids that lined the wall of the cleft were invariant.

When parts of molecules resist change, the constancy usually implies that the unchanged components are essential to the molecules' functioning. In this case, the uniformity suggested that the cleft formed the active, sialic acid cleaving site of neuraminidase and that the unchanging, or conserved, amino acids in the cleft were critical to maintaining catalytic function. Subsequent work confirmed this suggestion.

Given that influenza viruses cannot spread readily from cell to cell without help from neuraminidase, the new discoveries implied that a drug able to occupy, and jam, the active site would inhibit neuraminidase in all versions of the influenza virus. That is, such a "plug" drug might serve as a universal cure for the flu.

The Making of a Plug Drug

Pursuing this tantalizing idea, Colman's group identified the amino acids in the active site that normally contact sialic acid. They also looked for amino acids in the cleft that did not



BRYAN CHRISTIE; SOURCE: R. SCOTT ROWLAND *BioCryst Pharmaceuticals, Inc.*

NEW DRUGS clog the active site of neuraminidase by binding to it more readily than sialic acid—the substance it normally cleaves. Sialic acid (*left*) is held in the site, a cleft, mainly through its glycerol and carboxylate groups, which form bonds (*green lines*) with amino acids in the active site. Zanamivir (*center*) adds other bonds by replacing the hydroxyl of a sialic acid derivative with a large, positively charged guanidine, which forms strong attachments to two negatively charged amino acids at the bottom of the cleft. GS 4104 is converted to GS 4071 in the body. The resulting molecule (*right*) retains the carboxylate bonds made by sialic acid but also makes use of a hydrophobic group. This group induces the binding cleft to form a similarly hydrophobic pocket, which holds the drug in place through hydrophobic attractions (*short green lines*).

Building Better Vaccines

When contemplating the prospects for flu drugs, observers might reasonably wonder whether the disease would be better controlled by a universal vaccine—one able to prevent infection by inducing the body to mount a protective immune response against any influenza strain that might appear. Regrettably, no such all-purpose vaccine has yet materialized.

Nevertheless, immunologists are honing ways to speed vaccine production, so that immunization can be carried out swiftly if a virulent epidemic starts abruptly. They are also working on injection-free vaccines, to improve acceptance and to encourage immunization of children. Although the elderly tend to become sickest when they have the flu, children account for much of its spread.

Flu vaccines have been common since the 1940s. Today the manufacturing process begins after influenza samples collected by 110 surveillance sites around the world are analyzed. In February the World Health Organization pinpoints three strains—two type A and one type B—that seem likely to account for most of the flu that will occur in the upcoming season (November to March in the Northern Hemisphere). These become the basis for the vaccine.

A simple way to make a vaccine would be to grow vast numbers of the selected strains, inactivate them so that they cannot cause infection and combine them in a single preparation. Unfortunately, the strains that are selected tend to grow slowly in the laboratory and are thus difficult to mass-produce. To overcome this obstacle, scientists begin by basically inserting immune-stimulating proteins—hemagglutinin and neuraminidase—from the surface of selected strains into a form of influenza virus that will grow quickly in the lab. For each strain, they infect chick embryos with both the fast-growing and the chosen virus. Many of the virus particles made by the embryos grow rapidly but now display the hemagglutinin and neuraminidase spikes of the strains expected to cause this year's epidemics. These high-growth reassortments are then isolated and delivered to vaccine manufacturers, who mass-produce them in more chick embryos.

At one time, inactivated forms of these viruses, including reassortments for all three of the selected strains, served as the vaccine. Now most manufacturers take the process a step further. They break the viruses apart and compose the vaccine of the viral proteins. The proteins elicit immunity but are totally unable to cause any kind of infection. In both cases, the vaccines prod the immune system to make antibodies able to bind to, and help eliminate, infectious viruses bearing those same proteins.

In an alternative approach, investigators are testing vaccines made of weakened live viruses, because live viruses evoke production not only of antibodies but also of white blood cells known as T

bind sialic acid but might be exploited to help anchor a plug drug. They noted, for instance, that the cleft included three positively charged amino acids that held tightly to a negatively charged group (carboxylate) on sialic acid.

In addition, at the bottom of the cleft they spotted a small pocket containing two negatively charged amino acids. These amino acids—glutamates—made no contact with sialic acid but were nonetheless present in all influenza neuraminidases examined. A hydroxyl (OH) group on the bound sialic acid pointed down toward that extra pocket but did not reach it.

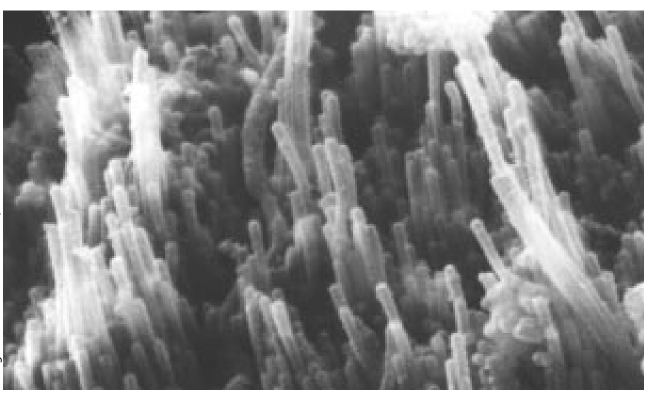
These features suggested that replacing this OH with a large, positively charged atomic grouping might yield a tight-binding derivative. The positive group would presumably nestle into the extra pocket at the bottom of the active site and would lock itself there by binding to the previously unused, negatively charged glutamates in the pocket.

After some trial and error, in 1993 Mark von Itzstein and his colleagues at Monash University in Melbourne found that substituting a guanidino group (which is both large and positively charged) for the OH group on sialic acid produced an extraordinarily potent inhibitor of influenza neuraminidases. Further, the inhibitor had little effect on related enzymes made by bacteria and mammals, a sign that the compound probably would not disrupt human cells.

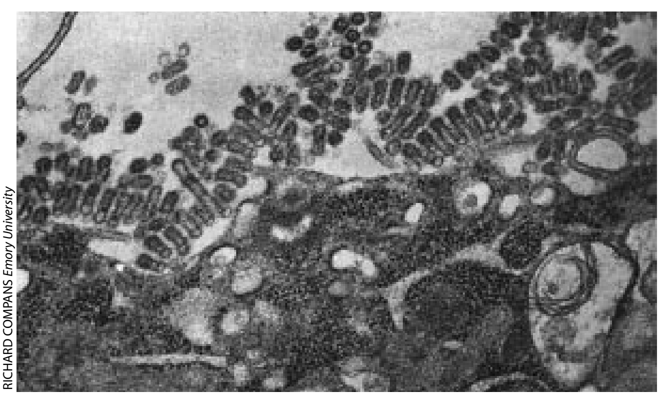
Studies in animals and preliminary trials in humans then revealed that the substance—zanamivir—prevented flu symptoms in individuals subsequently infected with influenza viruses and also reduced the severity of symptoms in those who took the drug after being infected. The compound, however, did not work if swallowed as a pill; it had to be inhaled into the respiratory system through the nose or mouth.

Ironically, the guanidino group that makes zanamivir such a good inhibitor is the reason the substance cannot be taken as a pill. An ingested drug has to cross the cells lining the intestine and migrate into the bloodstream before traveling to other parts of the body. But charged molecules have difficulty crossing cell membranes, which are fatty and permeable mostly to noncharged substances.

Because inhalation is a common way to take medicine meant to work in the respiratory tract, Glaxo Wellcome in Stevenage, England, initiated further human testing of zanamivir. Yet be-



DAVID HOCKLEY *National Institute for Biological Standards and Control, London*



RICHARD COMPANS *Emory University*

STEPHEN ROSE *Gamma Liaison*

ANNUAL VACCINATION is currently the best way to evade the flu. For some, shots may one day give way to a nasal spray vaccine.

lymphocytes. These cells recognize and eliminate virus-infected cells. T cells turn out to respond to closely related strains of influenza, not just to the single strains recognized by individual antibodies. Hence, they could potentially provide immunity for a while even after an influenza strain underwent small changes in the structure of its surface molecules.

A live-virus vaccine that is delivered as a nasal mist has been developed by Aviron in Mountain View, Calif. It has tested well in people, including children, and will probably be on the market in a year or two. Unfortunately, live-virus vaccines cannot be produced much more quickly than killed-virus types, and so they probably would not provide a rapid defense against a sudden pandemic.

To shorten production time, scientists are examining manufacturing methods that sidestep the need for acquiring large numbers of fertilized eggs. One approach inserts hemagglutinin and neuraminidase genes from selected influenza strains into another kind of virus, such as a baculovirus, that grows readily in cultured cells—something influenza viruses do only poorly. As the genetically altered viruses reproduce in the cells, they also make large quantities of the encoded influenza proteins, which can then be purified for use in vaccines. Recombinant vaccines can be prepared and distributed in just two or three months, but their effectiveness is still being evaluated.

Yet another vaccine strategy, able to yield a product even faster, relies on "naked" DNA. In this scheme, investigators fit desired hemagglutinin and neuraminidase genes into rings of DNA known as plasmids. In theory, if such plasmids were injected into skin or muscle, cells in the vicinity would take them up and use them to make influenza proteins. These proteins would then be displayed on the cells' surface, where cells of the immune system could spot them. In response, the immune system would deploy antibodies and T cells able to neutralize free virus and eradicate any infected cells. Naked-DNA influenza vaccines have worked well in lab animals but have yet to be tested in people. *—W.G.L., N.B. and R.G.W.*

cause some patients prefer to take a pill, Gilead Sciences in Foster City, Calif., joined with F. Hoffman–La Roche in Basel, Switzerland, to search for a neuraminidase inhibitor that could be taken as a tablet. A systemic drug that traveled in the blood could also conceivably help fight any highly virulent influenzas that infected cells outside the respiratory tract.

A Second Candidate Emerges

After synthesizing and testing many compounds, Gilead and F. Hoffman–La Roche finally found one, given the working name of GS 4071, that was as potent as zanamivir. Structural analyses revealed that it retained one major neuraminidase-binding group also found in zanamivir: a negatively charged carboxylate group that bound to positively charged amino acids in the enzyme's active site. In addition, the interaction of GS 4071 with the active site somehow caused one of the amino acids in the cleft to rotate and create a new hydrophobic (water-hating) pocket. This pocket became an anchor for a similarly hydrophobic component (a chain of carbons and hydrogens) of the drug.

NEWLY MADE VIRAL PARTICLES can emerge from the cell as filaments (*left*), although they can also be spherical or any shape in between. The particles can be seen to cluster ineffectually on the cell surface (*small rods and spheres at top of right micrograph*) when their neuraminidase molecules are inactivated.

Initially the substance functioned well in the test tube and did no harm to healthy cells. But it failed a crucial test: the negative carboxylate group kept it from crossing out of the gut into the bloodstream in animals. Happily, though, a small adjustment—in essence, placing a mask over the negative group—fixed that problem. The masked form, dubbed GS 4104, passed easily into the bloodstream, after which it was unmasked, primarily in the blood and liver. Having been converted back to its original (GS 4071) form, it proceeded to inhibit the activity of neuraminidase and the spread of the virus in the respiratory tract of test animals. Like zanamivir, it also showed signs of functioning well in people.

Last fall reports from large, controlled human trials of zanamivir and GS 4104 confirmed and extended the earlier findings. If begun within about a day and a half after symptoms develop, both the inhaled zanamivir and the swallowed GS 4104 can reduce the time that people feel ill by about 30 percent (1.5 to three days). The compounds also reduce the severity of symptoms. In a trial of GS 4104, for instance, rating diaries kept by patients indicated that symptoms were 25 to 60 percent milder than in those patients who took a dummy pill. Further, the drugs lower the risk for potentially lethal secondary bacterial infections, such as bronchitis, by half or more. Such complications are a major cause of flu deaths, especially among the elderly and people with concurrent disorders.

A small study of zanamivir has recently indicated that an injected form of the drug is helpful as well. Separate investigations into prevention are encouraging, too. When zanamivir was tested, 6 percent of nonusers but only 2 percent of users came down with the flu. GS 4104 produced comparable results. So far neither drug has caused serious side effects.

On the basis of such studies, Glaxo Wellcome has requested approval to market the inhaled form of zanamivir in Australia, Europe, Canada and the U.S. Gilead and F. Hoffman–La Roche are expected to submit a similar application for GS 4104 in the U.S. and Europe this year. Other neuraminidase inhibitors developed by BioCryst Pharmaceuticals in Birmingham, Ala., have done well in studies of animals and have been licensed to Johnson & Johnson for further development.

To the casual observer, shaving a few days off a bout of the flu may seem a minor accomplishment, yet the achievement is more profound than it first appears. The fatigue and other discomforts felt during the last days of the flu

What Accounts for Virulence?

When the flu kills, it usually takes the lives of people whose immunity is already compromised, such as by advanced age or some preexisting disorder. At times, though, it cuts down vigorous young people as swiftly and surely as it does the infirm.

Such was the case in 1918, when the "Spanish flu" pandemic killed more people than died fighting World War I. During that pandemic, even robust soldiers perished. Some felt a bit sick in the morning, went to bed in the afternoon and were dead by nighttime. Vigorous young individuals also became victims in 1997, when six of 18 people stricken by a novel strain of influenza died in Hong Kong.

What makes one strain inherently more lethal than others? Part of the answer seems to be an ability to infect a number of different tissues instead of the restricted set usually preferred by influenza viruses—namely, the respiratory tract in mammals and the gastrointestinal tract in birds. Many investigators interested in understanding the transformation to virulence are therefore hunting for features that enable some strains to become promiscuous, or pantropic, in the cells they attack.

About 15 years ago one of us (Webster) and his colleagues at St. Jude Children's Research Hospital in Memphis uncovered a possible clue. In 1983 a virus that had been causing mild gastrointestinal disease in chickens in Pennsylvania suddenly began killing entire commercial flocks. The team found that substitution of just one amino acid for another in a viral surface protein—hemagglutinin—was the culprit. That small change had somehow enabled the virus to replicate in, and damage, organs throughout the birds' body.

Subsequent work revealed why this tiny structural change had such a profound effect on viral activity. When an influenza virus first enters a cell, the virus is initially sequestered in a kind of intracellular jail (an endosome). The virus manages to reproduce nonetheless because hemagglutinin molecules on the viral surface help the viral membrane to fuse with the endosome cage. As fusion occurs, viral genes and proteins escape from the endosome and set about mass-producing new copies of the virus. Hemagglutinin can facilitate fusion only if it has been cleaved into two parts before the virus enters the cell. This cleav-

"SPANISH FLU" PANDEMIC that began in 1918 not only killed at least 20 million people, it also sent family members of flu victims to food lines. The scene below occurred in Cincinnati.

CORBIS-BETTMANN

stem mainly from the lingering activity of the immune system after it has eliminated most of the virus in the body. The anti-flu drugs cannot affect that activity. But by halting viral replication, the compounds can certainly minimize the duration and severity of the early, most miserable part of a flu episode.

As a rule, the earlier a person begins therapy, the better the outcome is—partly because the body will harbor a smaller, more manageable amount of virus and partly because the drugs cannot undo tissue damage caused by the virus before treatment starts. This finding, combined with the research into prevention, implies that the best way to use these drugs may be in conjunction with instant tests that will signal the presence of an incipient infection before a person becomes symptomatic. It is quite conceivable that in the near future people will, say, dab their tongue each morning with a small strip of influenza-detecting test paper. If the strip turns a color, the individuals will know they harbor the influenza virus and need to begin taking a flu-fighting drug to ward off symptoms. Quick detection kits are now becoming available for use in doctors' offices. Manufacturers are trying to perfect tests for home use as well.

Determining that a person has contracted the flu and not merely a cold or some other disorder is important because the drugs have power only against influenza viruses. They would be wasted if taken to combat colds (which are caused by other kinds of viruses), allergies or bacterial infections that produce flulike symptoms.

Questions

Although the neuraminidase-blocking agents are generating great excitement, certain questions remain. Logic dictates that inhibition of viral replication and reductions in secondary complications should save lives, but whether the drugs will actually prevent deaths is not yet known.

Further, almost all drugs that humans lob at viruses and bacteria eventually become undermined by drug resistance. As the microorganisms struggle for survival, those that alter the drug target and make it unrecognizable can escape the drug's effects [see "The Challenge of Antibiotic Resistance," by Stuart B. Levy; Scientific American, March 1998]. Amantadine and rimantadine are among those plagued by the resistance problem. Are there any reasons to think the new flu drugs will escape this fate?

In fact, there are. Scientists have tried diligently to produce strains of influenza in the laboratory that are resistant to zanamivir and GS 4104. So far they have had only limited success. Some strains gained resistance, as might be anticipated, by altering amino acids in the active site of neuraminidase. But these changes made the enzyme less stable or less active than usual, suggesting that viral survival in the body would be impaired.

Other strains became resistant not by altering neuraminidase but by changing the chemical makeup of hemagglutinin. Recall that neuraminidase is needed to strip sialic acid off nascent viral particles, so that hemagglutinin will not stick to sialic acid on neighboring particles and prevent the virus from spreading to other cells. The change in hemagglutinin reduced its affinity for sialic acid, thereby obviating the need for neuraminidase to intervene. Although it thrived in cell culture, this second group of mutants showed no resistance to the drugs in animals. Conceivably, the impaired binding age is accomplished by particular enzymes, in the serine protease family, that are made in the avian digestive tract and the mammalian respiratory tract but are less evident in most other tissues.

The amino acid substitution found in the lethal avian virus altered the cleavage site on hemagglutinin in a way that made it accessible to cutting by enzymes (furinlike proteases) that are common in tissues throughout the body. Such increased susceptibility to cleavage enabled the virus to infect tissues systemwide.

This discovery suggested that the 1918 human pandemic might have become deadly because the responsible influenza strain carried a mutant form of hemagglutinin that was susceptible to cleavage by common proteases found outside human airways. To address this possibility and to search for other sources of virulence, Jeffery Taubenberger and his colleagues at the U.S. Armed Forces Institute of Pathology have been studying genetic material recovered from three victims of the long-ago pandemic: two soldiers (from whom tissue samples had been saved) and an Inuit woman whose body was exhumed from the permafrost of Alaska in August 1997.

The genetic work has revealed the amino acid sequence of the virus's hemagglutinin molecule, and Taubenberger's group has almost completed work on the sequence of a second surface molecule: neuraminidase. The hemagglutinin molecule turns out to be unremarkable at the cleavage site. In addition, Taubenberger says his as yet unpublished analysis of the neuraminidase gene indicates that neuraminidase lacks another kind of mutation that had been proposed as a possible route to viral promiscuity. That intriguing proposal suggested that a particular mutation in neuraminidase would basically allow it to hoard serine proteases and use them to cleave hemagglutinin molecules in tissues that did not provide those proteases.

For now—and perhaps always—the reasons for the extreme virulence of the 1918 pandemic will remain mysterious. But investigators do have a sense of why the influenza strain that appeared in Hong Kong in 1997 was so deadly.

That virus, which originated in fowl, did possess a form of hemagglutinin that is highly susceptible to cleavage. Still, researchers do not have absolute proof that this form of hemagglutinin accounts for the virulence. For that reason, they are continuing to comb the viral genes for hints to other explanations. *—W.G.L., N.B. and R.G.W.*

CHICKENS were trucked to Hong Kong in February 1998, replacing some of the millions killed to halt the spread to humans of a lethal influenza virus. Officials were about to test blood from the birds.

ROBYN BECK *AP Photo*

between hemagglutinin and sialic acid reduced the infectivity of the strains by weakening their ability to dock with cells.

Neuraminidase-blocking drugs are not the only ones under study. Scientists have long been trying to build plug drugs that target the sialic acid binding site on hemagglutinin, for instance, but up to now the effort has failed. Amantadine and rimantadine, the drugs already on the market, were found to work before their mechanism of action was known. It is now evident that they interfere with the activity of a viral protein called M2, which functions as an ion channel. This inhibition of M2 explains why the drugs have no effect against type B influenzas: those viruses do not carry an M2 molecule. As work on new drugs continues, so, too, do efforts to improve flu vaccines [*see box on pages 84 and 85*].

Between pandemics, smaller but still substantial influenza epidemics are common. In 1994, according to the U.S. Centers for Disease Control and Prevention, an estimated 90 million Americans (about 35 percent of the population) contracted the flu. Collectively, those people spent 170 million days in bed and lost 69.3 million days of work. More typically in the U.S., the flu afflicts 10 to 20 percent of the population every year and causes some 20,000 deaths from influenza-related complications.

These numbers, like those associated with pandemics, are likely to fall in the next several years as new anti-flu drugs become available and as vaccines become more widely used and faster to produce. Indeed, society may finally be heading for an era in which the human species gains the upper hand against the much dreaded influenza virus. SA

The Authors

W. GRAEME LAVER, NORBERT BISCHOFBERGER and ROBERT G. WEBSTER have all contributed to progress in the control of influenza epidemics and pandemics. Laver, professor of biochemistry and molecular biology at the Australian National University in Canberra, produced the first crystals of neuraminidase and continues to provide them to researchers who are designing new neuraminidase inhibitors. Bischofberger is senior vice president of research at Gilead Sciences in Foster City, Calif. Webster, who with Laver first traced pandemic strains of influenza viruses to lower animals, is Rose Marie Thomas Professor and chairman of the department of virology and molecular biology at St. Jude Children's Research Hospital in Memphis. He is also director of the World Health Organization's Collaborating Laboratory of the Ecology of Influenza Viruses in Lower Animals and Birds.

Further Reading

THE PLAGUE OF THE SPANISH LADY. Richard Collier. Atheneum, 1974.

EPIDEMIC AND PEACE, 1918. Alfred W. Crosby, Jr. Greenwood Press, 1976.

INFLUENZA. Edwin D. Kilbourne. Plenum Medical Book Company, 1987.

INFLUENZA VIRUS. Frederick G. Hayden and Peter Palese in *Clinical Virology.* Edited by Douglas D. Richman, Richard J. Whitley and Frederick G. Hayden. Churchill Livingstone, 1997.

TEXTBOOK OF INFLUENZA. Edited by Karl G. Nicholson, Robert G. Webster and Alan Hay. Blackwell Science, 1998.

Y2K: So Many Bugs ... So Little Time

Fixing Y2K seems simple: change all two-digit years to four digits. But that tedious—and unexpectedly difficult—process takes more time than is left

by Peter de Jager

An explanation of the psychology behind the Year 2000 computer problem can be found in a perhaps unlikely place: Lewis Carroll's *Alice in Wonderland*. In the popular children's classic, the Mad Hatter asks, "Does your watch tell you what year it is?" and Alice replies, "Of course not, but that's because it stays the same year for such a long time together."

There are many reasons why programmers, including me, chose to represent years by using just two digits, 55 for 1955 and 10/23/76 or 23/10/76 for October 23, 1976, for example. Decades ago digital real estate was scarce: computer memory was expensive, and typical punch cards were only 80 columns wide. People also rationalized the shortcut by citing the efficiency of reduced keystrokes. Of course, the absence of standards played an enabling role, and many of us truly believed (incorrectly so) that the software we were writing would long be retired before the new millennium. Thanks to sheer inertia and lingering Tea Party logic (why store more than two digits when the century stays the same for such a long time together?), the practice continued long after computer memory and cost constraints were legitimate concerns.

The net result? Computers are now riddled with representations of dates that are frighteningly ambiguous. Simply put, how will digital machines know whether 00 means 1900 or 2000? Already the confusion has led to a variety of problems. Back in 1993, Boeing noticed errors in its application that handles orders with seven-year lead times. A system at Amway Corporation, a global manufacturer of soap and other personal care products, rejected chemicals that it had mistakenly thought to be almost a century old. And some computerized cash registers have crashed when customers tried to use credit cards expiring in 00. In fact, according to one industry study conducted a year ago, more than 40 percent of the organizations surveyed had already suffered some kind of Year 2000, or Y2K, failure.

The Soul of the Y2K Bug

Exactly what is causing such problems? An obvious example is a savings deposit made in 1999 and withdrawn in 2000. If the accounts program at the bank calculates the interest earned by first subtracting 99 from 00, the computer will mistakenly think that the term of deposit is −99 years.

Consider a more obscure Y2K pitfall: An insurance company must routinely look for policies that have been dormant for more than five years so that it can delete them from its databases. To locate the inactive policies, the firm runs an application that relies on data stored as LAST-ACCESS, which contains time stamps of the most recent dates that customers conducted a transaction.

When the program examines a LAST-ACCESS value, it adds five to the two-digit year. If the result is less than the current year, it deletes the record. Therefore, when LAST-ACCESS is 93, then $93 + 5 < 99$, and the policy will be correctly declared inactive. If LAST-ACCESS is 96, however, the processing gets dicey. If only two digits are allowed and extra information is truncated, $96 + 5$ will become 01, which is less than 99. The result would be the incorrect termination of the policy. (Because this example describes sloppy programming, people might be tempted to dismiss it. But this type of failure has already occurred.)

In other instances, a Y2K error could literally have fatal consequences. Using a medical program that specifies the dosage of certain drugs, a doctor types in "03-16-00" for an infant's birth date. The computer, however, as-

FAILURES THAT HAVE ALREADY OCCURRED

Unum Life Insurance Company deleted 700 records from a database that tracks the licensing status of brokers because a computer mistook 00 for 1900.

Mary Bandar, a centenarian, received an invitation to attend kindergarten in Winona, Minn.

In 1993 Boeing experienced errors in a system that used seven-year lead times for orders.

PC-based mixing system at Amway Corporation rejected a batch of chemicals because it mistakenly believed the expiration date to be in 1900.

sumes the patient is a centenarian and recommends a dosage that would have been fine for an elderly adult but turns out to be deadly for the tiny newborn.

This, then, is the soul of the Year 2000 problem. People have stored years ambiguously, and that confusion is now beginning to wreak havoc. Many organizations have simply replaced the offending programs with newer software that is Y2K-ready. For embedded systems [*see box on page 92*], that tactic has sometimes been the only viable option. But replacing the myriad computer programs already in use would be too costly and time-consuming. Companies are thus trying to mend what they have.

But What's a Date?

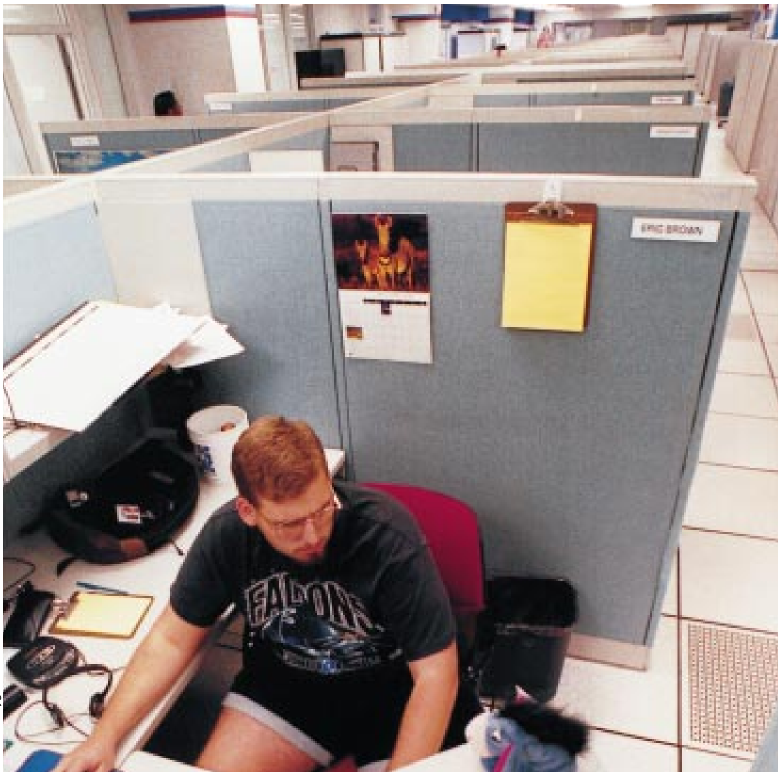
When people first became aware of Y2K, they quickly proposed a legitimate, even optimal, repair strategy: if two digits were left out, just put them back in. Setting aside for now the logistical complexities involved in modifying the terabytes of data in a multinational corporation, this solution raises the crucial yet seemingly facetious question, What's a date? After all, before people can add two digits to every year, they must first find the dates.

It is important to note that computers do not know what any of the information they reference actually represents. The meaning of the data is supplied by—and, more to the point, remains with—humans. And programmers did not adhere to a set of standards when creating names to identify dates. For instance, dates can have obscure labels, such as SNARK and WUMPUS, as well as more obvious choices, such as BIRTH_DATE and EMP_START_DT.

Nor can people rely entirely on the numerical information itself. One approach has been to sift through data and pick out columns with numbers that range from 1 to 12, from 1 to 31 and from 0 to 99 and then assume that the last piece of information denotes years. Although justified, this assumption is not always correct. The 0-to-99 data might instead represent percentages of a certain quantity. And often two-digit years are buried deep inside other data, such as within long product serial numbers.

Automated software tools have been developed to locate dates. Some of the more advanced products deploy deductive reasoning

SLIM FILMS; SCOTT CAMAZINE *Photo Researchers, Inc.* (honeycomb); SCOTT SMITH AND SANDRO VANNINI *Corbis* (bees)



CHIP EAST *Sygma*

COMPUTER PROGRAMMERS have been busy correcting Y2K problems. In this software "factory" in Charlotte, N.C., more than 250 technical staff of contractor Alydaar Software fix millions of lines of code each month. Much of this work is for Fortune 500 companies.

to achieve astounding rates of success. Still, no tool has achieved error-free operation.

Finding dates is merely the first step. One problem with extending all years to four digits is that programmers might have to redesign the layouts of certain reports, forms and screen displays to accommodate the extra digits. A larger complication is that software applications that refer to the expanded data would also have to be modified.

Consider a personnel database in which employee names are stored in columns 1 through 30, birth dates from 31 through 36, salaries from 37 through 42 and so on. If the birth dates are expanded to accommodate four-digit years, then the salary information and all subsequent data will also be bumped two columns to the right. As a result, every program that retrieves that information must be adjusted to obtain data from the correct locations.

Such changes are typically made to "source code," which programmers write using languages like COBOL and C. The software is then converted into a form—called an object module—that a computer can comprehend. The translation is performed by a program called a compiler. Problems can arise because compilers are continually upgraded. In some cases, programming techniques that an earlier compiler accepted as legitimate might now be forbidden, in much the same way that a document created by the third version of a word-processing software program might be rejected by version five of the same product. Consequently, old source code that has recently been corrected to handle Y2K may no longer compile properly into a new object module unless additional modifications are made.

To make matters worse, many companies have lost parts of their source code. Although the quantity of missing software is typically less than 3 to 4 percent, even this tiny amount can be extremely troublesome because programmers cannot easily modify an object module directly. They must re-create the source code, either from scratch (a difficult task because the accompanying documentation is most likely also missing) or from the object module itself (a heinous process that has been compared to retrieving a pig from sausage).

Once source code has been repaired and recompiled, it must also be tested. Because software revisions almost always introduce new bugs, the verification of modified programs is now the largest part of any Year 2000 project.

In the early 1990s many experts asserted that date expansion was the best way to tackle Y2K. But the unexpected necessity of having to recompile every program that refers to a date in any file, even when the application does not perform date calculations, has made this approach too expensive and time-consuming for most companies. (Of course, expansion to four digits would, in the transition from A.D. 9999 to 10,000, also lead to the Y10K problem, but that is a subject for another article.)

Doing Windows

An alternative solution is to teach computers that 00 means 2000. Programmers have extended this simple idea into a strategy called windowing. They have taken all the years from 00 to 99 and divided them into two groups based on a carefully selected pivot (45, for example). Two-digit years greater than or equal to that number are considered to reside in the current century (68 becomes 1968). Everything else is considered to lie in the 21st century (13 becomes 2013).

Using this concept, a programmer can delve into the source code, find all date references and modify accordingly the calculations involving that information. Because the actual two-digit years themselves do not need to be altered (just the calculations involving those dates are adjusted to place years in the appropriate centuries), windowing requires less work than date expansion and is currently the technique most commonly used to fix Y2K.

The approach, however, has weaknesses. It obviously fails for any data that span more than 100 years, such as birth dates and long-term leases. It also poses some interesting prob-

Y2K Litigation

After computerized cash registers at Produce Palace in Warren, Mich., repeatedly crashed when customers tried to use credit cards expiring in 00, the grocer sued TEC America, the manufacturer of the system.

lems when information must be moved between systems that have used different pivot points. For example, a pivot of 25 might be appropriate for a program that handles the invoices of a company founded in 1928. But another application that does long-term sales projections at the same firm might deploy a pivot of 70. The problem, of course, is when information must be moved between the two systems: the year 2031 will become 1931, or vice versa, causing mayhem in the calculations.

To complicate matters, programmers have also used sliding windows, in which the pivot changes over time. This strategy makes sense for certain types of programs, such as those that handle 30-year home mortgages. For such an application, the pivot might be set equal to 40 years from the current year. Obviously, it is crucial to keep track of sliding pivots to avoid possible conflicts with other systems. Furthermore, programmers must recognize that representing the year 1968 in a program with 70 as its pivot will require additional software tweaking.

Time Shifting

In another approach, basic arithmetic is used to finesse Y2K. Consider the calculation 00 – 99 = –99. If the operation was intended to represent 2000 – 1999, then the answer of –99 is obviously incorrect. But note that 00 – 99 is equivalent to (00 + 5) – (99 + 5). If that expression is calculated in two-digit math, it will yield 5 – 4 = 1. Adding 5 to both 00 and 99 has, in effect, shifted both dates into the same century, so that the computation of 2000 – 1999 could be performed correctly using just two digits.

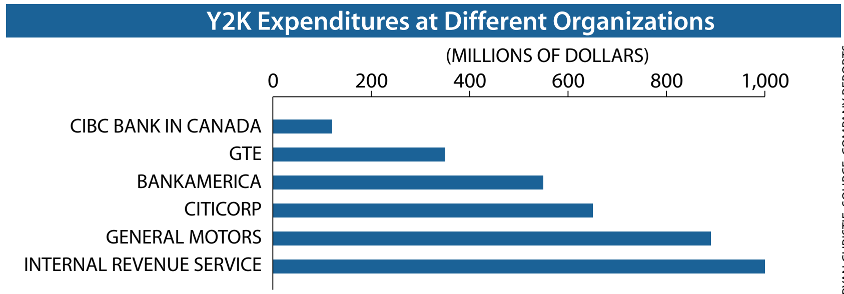
But dates are more than mere numbers. January 1, 2000, is a Saturday; January 1, 2005, is not. Thus, the approach of adding 5 to all years will fail for programs that need to distinguish the days of the week. Still, all is not lost. Basically, there are two cycles controlling the days of the week: a cycle of seven for the different days and one of four for leap years. Multiplying the two gives an overall cycle of 28 years. January 1, 2000, will be a Saturday just as January 1, 1972, was and as January 1, 2028, will be. Taking advantage of that pattern, the "encapsulation" technique adds 28 to two-digit years before performing further calculations. Once those computations are complete, 28 is then subtracted from the dates.

Although encapsulation can be used to sidestep many Y2K problems, the technique becomes unwieldy for complex computations. An example is an application that works in parallel with one or more other programs: during its own computations, the application may need to send information to other systems that manipulate the data in their own ways; the results might then be combined and recombined at various intervals. Programmers would have difficulty determining and keeping track of when to add and when to subtract 28.

In addition, encapsulation fails with dates that are buried in other information that uses certain digits for validation purposes. Consider a product stock number such as 7289-47-99-5, in which 99 is the expiration year and the last digit, 5, is used to verify the overall sequence. In this instance, the 5 is obtained from adding 7289, 47 and 99 and taking the last digit of the resulting sum of 7435. So-called check digits are often used to verify credit card, bar code and Social Security numbers. Obviously, cavalierly adding 28 to two-digit years would gum up such validation calculations.

The three techniques—date expansion, windowing and encapsulation—have thus far accounted for more than 95 percent of Y2K fixes to existing software. Many large companies with thousands of computer programs have deployed a mixed approach. Software tools exist to automate the three solutions, but none has achieved error-free success.

Perversely, one factor that severely hampers Y2K fixes is that some software already takes into account the century change. The original programmers have used their own windowing or encapsulation scheme in anticipation of Y2K problems. But these precautions may not have been deployed throughout an entire program, leaving the system betwixt and between a solution. For such software, implementing additional remedies could result in a digital



BRYAN CHRISTIE; SOURCE: COMPANY REPORTS

Other Rollover Dates

***August 22, 1999:* Global Positioning System (GPS) will reset to week 0 after 1,023 weeks.**

***January 19, 2038:* Unix systems will roll over, because it will be 2^{31} seconds from their start date of January 1, 1970.**

***February 6, 2040:* Older Macintosh computers may revert to January 1, 1904, losing 2^{32} seconds.**

How Prepared Is the U.S. Federal Government?

	Total Number of Critical Systems	Systems Fixed as of August 1998	Estimated Systems Fixed by March 1999	Estimated Year When All Systems Are Fixed
SSA Social Security Administration	308	93%	99%	1999
FEMA Federal Emergency Management Agency	49	69%	82%	1999
Treasury Department of Treasury	323	45%	61%	2000
DOD Department of Defense	2,965	42%	54%	2001
Justice Department of Justice	207	31%	31%	2030+
NASA National Aeronautics and Space Administration	158	63%	74%	2000

BRYAN CHRISTIE; SOURCE: SUBCOMMITTEE ON GOVERNMENT MANAGEMENT, INFORMATION AND TECHNOLOGY

Problems Embedded Everywhere

Hidden inside a medical system that doctors use for radiation treatments are computer chips that must ensure that the right dosage is delivered to a patient. But because the calculations are performed with two-digit years, the device will not be able to compute the correct strength of a radioactive sample that decays from the current century into the next. Fortunately, the manufacturer has already discovered the problem and has contacted its customers to recall the product.

Without a doubt, the dangerous wild cards of Y2K are preprogrammed computer chips that have been built into a wide variety of electronic equipment, including industrial machinery, monitoring devices, traffic lights, security alarms, navigation tools and countless consumer products, such as automobiles, watches, VCRs and microwave ovens. Such "embedded systems" are everywhere—in factories, nuclear power plants, water and sewage systems, hospitals, office buildings and homes. The Gartner Group, a consultancy in Stamford, Conn., puts the worldwide total at 32 billion to 40 billion.

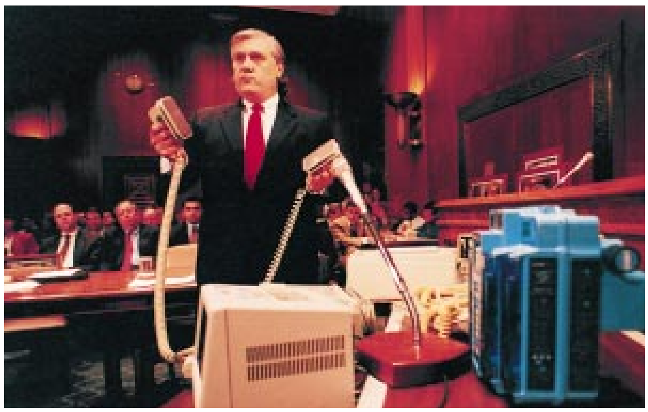
Of course, most of these systems do not depend on any knowledge of dates and will therefore not be affected by Y2K. Of those that are date-sensitive, only a very tiny fraction will suffer anomalous processing. Nevertheless, even that minute percentage is still cause for concern. Although nobody knows for sure the exact number, Gartner estimates that millions of those embedded systems are vulnerable.

Such enormous uncertainty exists because it is often extremely difficult for someone—even an experienced engineer—to determine exactly how a particular embedded system might fail. Often equipment that does not at first appear to depend on date calculations actually does. Consider a device that for safety reasons shuts itself down on January 1, 2000, because it mistakenly thinks that its last inspection was nearly a century ago. An even more insidious problem could occur with a monitoring system that issues emergency warnings. The crucial messages might be ignored because they are date-stamped 00 and have thus been pushed off the operator's screen by seemingly more current messages dated 99. Indeed, Y2K problems could occur in the unlikeliest of places. Firefighters in Baton Rouge, La., had to test their fire trucks to ensure that the mechanisms controlling the water pumps and ladders do not require Y2K repairs. Fortunately, they don't.

Personal computers may be among the easiest products to correct embedded problems. Most PCs purchased before 1996 have no intrinsic knowledge of the century turnover: on January 1, 2000, many will automatically revert to their reset date in 1980. But software patches are available to correct that flaw.

Unfortunately, other fixes are not so simple. Many of the companies that manufactured embedded systems are no longer in business, or if they are, the people who designed the faulty products have left and the documentation is missing or incomplete. Furthermore, some of the buggy systems have been installed in remote sites, such as on offshore oil rigs.

A common solution is to make one-for-one replacements of the offending chips, including real-time clocks (which keep track of time with a crystal oscillator) and microprocessors and controllers (which instruct a device to perform certain actions). The process might be as simple as pulling the parts off printed circuit boards and plugging in their newer counterparts that are inscribed with revised software designed to handle Y2K. Often, though, the entire piece of equipment must be scrapped and replaced—obviously an expensive course of action but sometimes a necessary one, particularly if the device will have trouble calculating a patient's radiation dosage. —*P.d.J.*



CHIP EAST *Sygma*

KENNETH W. KIZER of the Veterans Health Administration testified before Congress last July that certain medical equipment, including cardiac defibrillators, would not work properly because of Y2K problems in their computer chips.

Y2K in Japan

Japan has been criticized for its perceived inaction on Y2K. But one factor in the country's favor is that many companies have used the alternative emperor system to record years, in which 1999 is Heisei 11.

mess—double windowing with different pivot years or a combination of encapsulation and windowing on the same data.

Determining whether a software application has preexisting fixes can be difficult, particularly if the code is poorly written and badly documented. As a programmer, I have sat in front of a small amount of code (merely 10 to 15 lines) for hours, struggling to determine what the software was supposed to accomplish. And the danger of automated tools is that they blindly apply a cookie-cutter solution with no understanding of what the code is actually doing.

There are additional monkey wrenches. A common programming practice has been to give dates or years that are close to 2000 a special meaning. Specifically, programmers have often used 9999 or simply 99 to mark the end of a file or a record that should be expunged or archived. Now, of course, that practice leads to confusion because the two quantities might instead legitimately mean September 9, 1999, or the year 1999, respectively. For instance, a sales application might prompt clerks to enter 99 as a year if they want to delete the corresponding customer order. The program must now be rewritten to enable that request to be made in a different way.

Another complication involves leap years. Because the earth takes a little less than 365.25 days to orbit the sun (the more exact number is 365.242199), leap years do not adhere to a strict four-year cycle. Exempt are century years—such as 1700 and 1800, which are not leap years—but exceptions to that exception are centuries evenly divisible by 400. Thus, there will be a February 29 in 2000 (even though there was no such day in 1900), which will further confuse computers that have not been properly programmed with that knowledge.

A Digital Disaster?

These various factors, among numerous others, have led to widespread uncertainty and heated controversy over Y2K. At one end of the spectrum lies extreme silliness—"There is a possibility we will lose electrical power forever...." (a statement actually made by a speaker at a recent Y2K conference). At the other end is ill-informed complacency—"Y2K is a one-day event. People will fix any problems over the weekend."

In my view, both extremes are equally naive. The former ignores society's ability to recover from adversity. The notion that people will somehow lose forever the intellectual capacity to produce electricity does not merit serious discussion. In fact, organizations such as financial institutions that have devoted the necessary resources have made great strides in combating Y2K. Last summer Wall Street simulated what stock trading would be like on January 3, 2000, and uncovered only minor date-sensitive problems. Additional tests are scheduled this spring.

On the other hand, pooh-poohing Y2K ignores the technological vulnerability of modern society, which is supported by an intricate foundation of interlocking codependencies. In particular, single points of failure can ripple quickly through a system, with disastrous results. Galaxy IV, just one of many communications satellites, had a problem last spring, and millions of pagers died a sudden death. A single cable failed in Auckland, New Zealand, overloading the system, and the city lost power for six weeks.

These events happened; none was expected. Of course, Y2K is different—it has been predicted. Today computer professionals around the world are modifying much of their existing software. The CIBC bank in Canada has 1,000 people working on its project with a budget of about $120 million. AT&T has already spent over $500 million, Citicorp will shell out about $650 million, and the Internal Revenue Service's expense will be roughly $1 billion.

These are huge efforts, but if people have learned anything about large software projects, it is that many of them miss their deadlines, and those that are on time seldom work perfectly. To deny this is to forget the lessons of past software debacles, including the computer fiascoes at the Atlanta Olympics and the Denver International Airport. Indeed, on-time error-free installations of complex computer systems are rare. The excruciatingly painful aspect of Y2K projects is that the deadline is immovable.

All that said—and considering other factors, including the amount of work already completed and the planned contingencies and compromises people will have to make as the century turnover nears—I believe that severe disruptions will occur and that they will last perhaps about a month. Additional problems, ranging from annoyances to more serious issues, will continue cropping up throughout 2000. This prediction might be optimistic; it assumes that people will have done what is necessary to minimize the number of single points of failure that could occur. Accomplishing that alone in the time remaining will require a Herculean effort unprecedented in the history of computers. SA

The Bottom Line for Y2K in the U.S.

	Best Case	Expected Case	Worst Case
Software applications with Y2K problems	10 million	12 million	15 million
Percent of Y2K problems that will not be fixed in time	5%	15%	25%
Infrastructure failures because of Y2K*			
Power systems	5%	15%	75%
Transportation systems	5%	12%	50%
Telephone systems	5%	15%	65%

**Percent of households that will be affected*

BRYAN CHRISTIE; SOURCE: SOFTWARE PRODUCTIVITY RESEARCH/ARTEMIS MANAGEMENT SYSTEMS

Cash-Flow Problems

In anticipation of people hoarding money, the Federal Reserve Board is planning to add $50 billion to the government's usual supply of cash.

The Author

PETER DE JAGER is a technology consultant whose essay "Doomsday 2000," published more than five years ago in *Computerworld,* a computer trade weekly, is widely credited with alerting industry and governments to the Y2K computer bug. In 1996 he was summoned before the science subcommittee of the U.S. House of Representatives to testify on Y2K, and he is currently a special adviser to the U.K.'s Year 2000 task force. He is the co-author of *Managing 00: Surviving the Year 2000 Computing Crisis* (John Wiley & Sons, 1997) and *Countdown Y2K: Business Survival Planning for the Year 2000* (John Wiley & Sons, 1998). Years ago critics dismissed de Jager as a Chicken Little who was preying on the fears of technophobes. Today he is viewed as a moderate, particularly in comparison with doomsayers who have proclaimed the end of the world because of Y2K. A year from now, on New Year's Eve, de Jager hopes to spend the night with his family at a pub in Doolin, County Clare, Ireland, where his mother lives.

Further Reading

YEAR 2000 PROBLEM: STRATEGIES AND SOLUTIONS FROM THE FORTUNE 100. Leon Kappelman. International Thomson Computer Press, 1997. ISBN: 1-85032-913-3

THE YEAR 2000 SOFTWARE CRISIS: CHALLENGE OF THE CENTURY. William M. Ulrich and Ian S. Hayes. Prentice Hall, 1997.

COUNTDOWN Y2K: BUSINESS SURVIVAL PLANNING FOR THE YEAR 2000. Peter de Jager and Richard Bergeon. John Wiley & Sons, 1998.

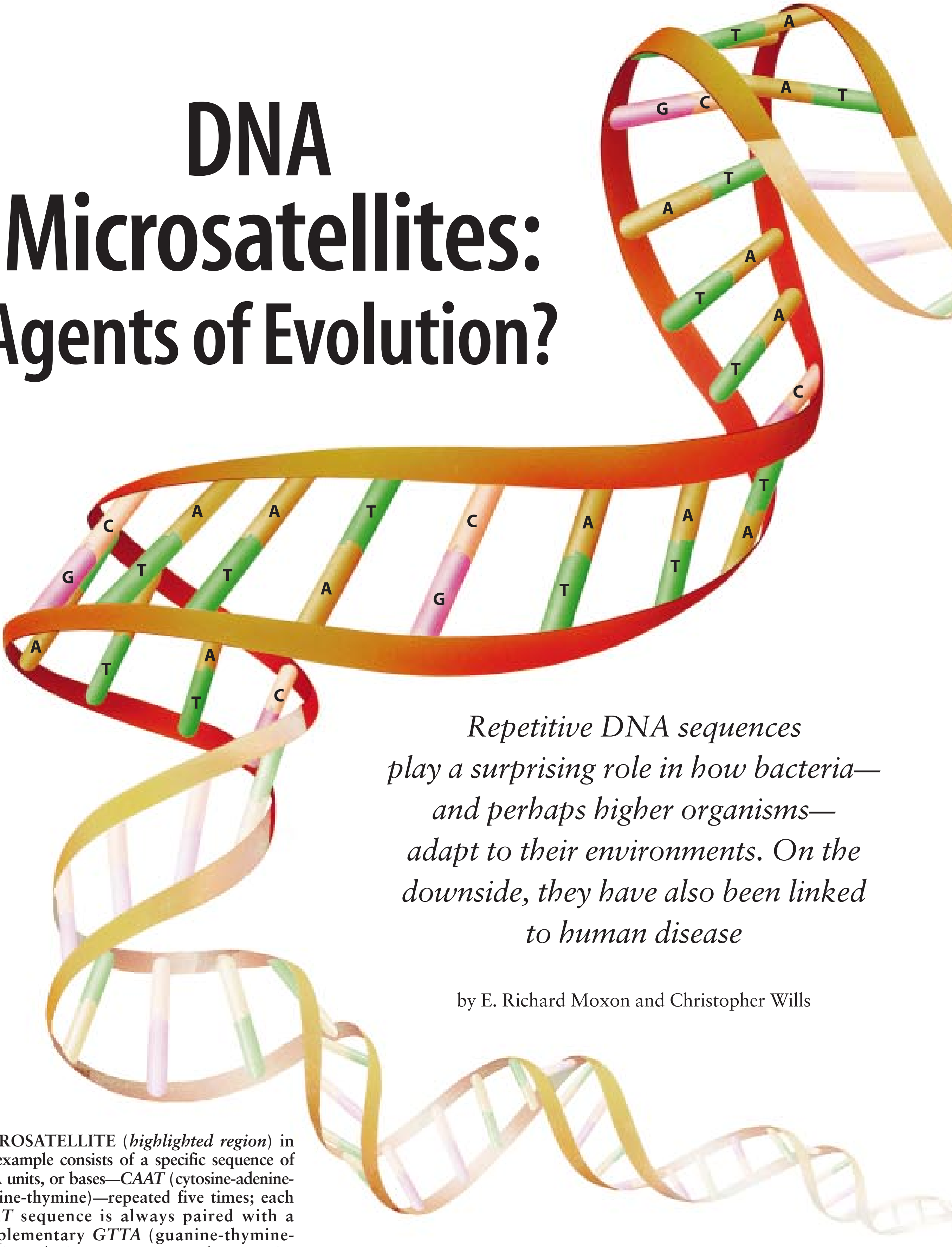
EVALUATING SUCCESS OF A Y2000 PROJECT. Howard Rubin and Brian Robbins. Information Economics Press, 1998.

YEAR 2000 COMPUTER CRISIS: LAW, BUSINESS, TECHNOLOGY. Michael D. Scott and Warren S. Reid. Glasser LegalWorks, 1998.

THE YEAR 2000 SOFTWARE PROBLEM: QUANTIFYING THE COSTS AND ASSESSING THE CONSEQUENCES. Capers Jones. Addison-Wesley, 1998.

For more on Y2K see the author's Web site www.year2000.com

DNA Microsatellites: Agents of Evolution?



Repetitive DNA sequences play a surprising role in how bacteria— and perhaps higher organisms— adapt to their environments. On the downside, they have also been linked to human disease

by E. Richard Moxon and Christopher Wills

TOMO NARASHIMA

MICROSATELLITE (*highlighted region*) in this example consists of a specific sequence of DNA units, or bases—*CAAT* (cytosine-adenine-adenine-thymine)—repeated five times; each *CAAT* sequence is always paired with a complementary *GTTA* (guanine-thymine-thymine-adenine) sequence on the opposite strand of the DNA "ladder." The repeating sequences, or motifs, in microsatellites can contain up to six bases, and each sequence can be present in multiple copies.

A human's genetic code consists of roughly three billion bases of DNA, the familiar "letters" of the DNA alphabet. But a mere 10 to 15 percent of those bases make up genes, the blueprints cells use to build proteins. Some of the remaining base sequences in humans—and in many other organisms—perform crucial functions, such as helping to turn genes "on" and "off" and holding chromosomes together. Much of the DNA, however, seems to have no obvious purpose at all, leading some to refer to it as "junk."

Part of this "junk DNA" includes strange regions known as DNA satellites. These are repetitive sequences made up of various combinations of the four DNA bases—adenine (*A*), cytosine (*C*), guanine (*G*) and thymine (*T*)—repeated over and over, like a genetic stutter. In the past several years, researchers have begun to find that so-called microsatellites, those containing the shortest repeat sequences, have a significance disproportionately great for their size and perform a variety of remarkable functions.

Indeed, scientists are discovering that the repetitive nature of microsatellites makes them particularly prone to grow or shrink in length and that these changes can have both good and bad consequences for the organisms that possess them. In certain disease-causing bacteria, for example, the repeat sequences promote the emergence of new properties that can enable the microbes to survive potentially lethal changes in the environment. Some microsatellites are also likely to have substantial effects in humans, because at least 100,000 occur in the human genome, the complete complement of DNA in a human cell. Although the only function assigned so far to human microsatellites is negative—causing a variety of neurological diseases—microsatellites may be surviving relics of evolutionary processes that helped to shape modern humans.

While some investigators search for the reasons humans carry so much repetitive DNA, many are now learning to exploit microsatellites to diagnose neurological conditions and to identify people at risk for those disorders. They are also finding that microsatellites change in length early in the development of some cancers, making them useful markers for early cancer detection [*see box on page 98*]. And because the lengths of microsatellites may vary from one person to the next, scientists have even begun to use them to identify criminals and to determine paternity—a procedure known as DNA profiling or "fingerprinting" [*see box on page 97*].

Satellite DNA was first identified in the 1960s. Researchers discovered that when they centrifuged DNA under certain conditions, it settled into two or more layers: a main band that contained genes and secondary bands that came to be known as satellite bands. The satellite bands turned out to be made of very long, repetitive DNA sequences. In 1985 Alec J. Jeffreys of the University of Leicester found other, shorter repetitive regions of DNA, which he dubbed minisatellites, that turned out to consist of repeats of 15 or more bases. (Jeffreys and his colleagues also determined that the number of repeats in a given minisatellite differs between individuals, a finding that allowed them to invent the DNA-fingerprinting technique.) In the late 1980s James L. Weber and Paula L. May of the Marshfield Medical Research Foundation in Marshfield, Wis., and Michael Litt and Jeffrey A. Luty of the Oregon Health Sciences University isolated satellites made up of still shorter DNA repeats and named them microsatellites; these, too, would prove useful for DNA fingerprinting.

Today scientists generally consider microsatellite DNA to consist of sequences of up to six bases repeated over and over, end to end, like a train made up of the same type of boxcar. What makes microsatellite DNA so important for evolution is its extremely high mutation rate: it is 10,000 times more likely to gain or lose a repeat from one generation to the next than a gene such as the one responsible for sickle cell anemia is to undergo the single-base mutation leading to that disease. And although it is quite rare for the single-base mutation that underlies sickle cell anemia to mutate back again to its benign state, microsatellites can readily return to their former lengths, often within a few generations.

"Smart" Microbes

The role of microsatellites in the diversity of pathogenic bacteria was uncovered in 1986 in the laboratory of Thomas F. Meyer of the Max Planck Institute for Biology in Tübingen. Meyer and his colleagues were studying *Neisseria gonorrhoeae,* the bacterium that causes the sexually transmitted disease gonorrhea. *N. gonorrhoeae,* a single-celled organism, possesses a family of up to 12 outer-membrane proteins that are encoded by genes called *Opa*s. (The name of the genes is derived from the opaque appearance of bacterial colonies that make Opa proteins.) The proteins produced by the *Opa*s are important because they allow the bacterium to adhere to and to invade epithelial cells, such as those that line the respiratory tract, as well as cells of the immune system called phagocytes. Each of the *Opa* genes contains a microsatellite composed of multiple copies of the five-base motif *CTCTT.*

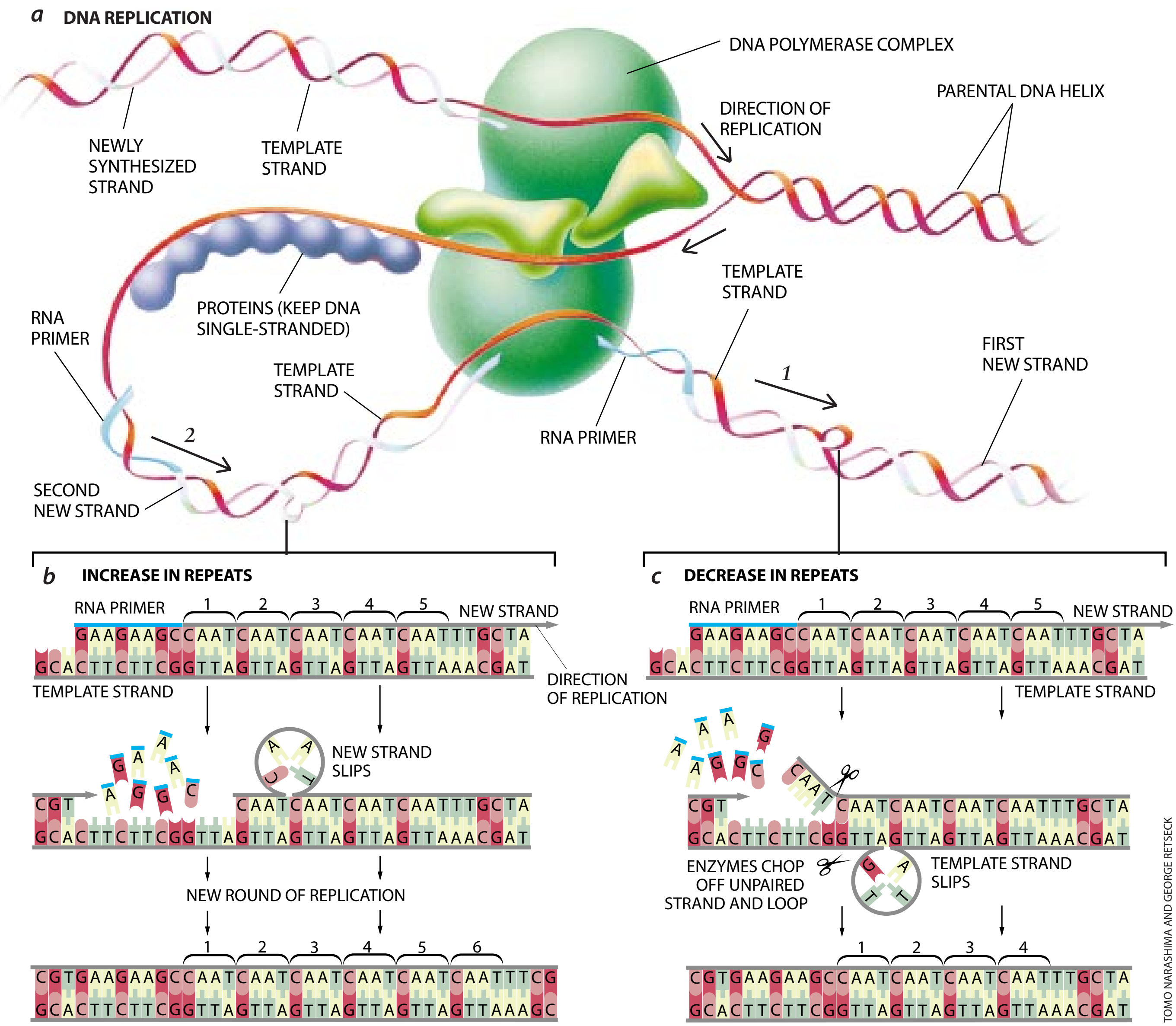
The enormous variation conveyed by microsatellite repeats results from the fact that the repeats are especially prone to DNA-replication errors, often through what is called slipped-strand mispairing. Before a cell—bacterial or otherwise—can replicate, it must make a duplicate set of its DNA. This is a complicated process because each DNA molecule is a double helix resembling a twisted ladder, where the rungs of the ladder are base pairs. The genetic code is spelled out by the bases on one side of the ladder; the bases along the other side are complementary (*A* always pairs with *T,* and *C* with *G*).

During DNA replication, the ladder splits down the middle, separating the base pairs, as enzymes called DNA polymerases copy each strand [*see box on next page*]. As the new strand is made, it pairs with its template. Slipped-strand mispairing can occur when either the old, template strand or the newly forming, complementary strand slips and pairs with the wrong repeat on the other strand. This slippage causes the DNA polymerase to add or delete one or more copies of the repeat in the new strand of DNA.

The frequency of such slippage mechanisms is very high in *N. gonorrhoeae:* each time the bacteria divide, approximately one

Slipped-Strand Mispairing

In this process, the number of microsatellite repeats increases or shrinks when a cell copies its DNA before dividing. During DNA replication (*a*), enzymes called the DNA polymerase complex unzip the parental DNA helix and copy both strands. One of the copies is made piecemeal: the polymerase complex synthesizes a short fragment (*1*) beginning with an RNA primer, then skips ahead to generate a second short fragment (*2*). When the polymerase finishes the second fragment, the RNA primer is removed, and the two fragments are connected with DNA. Increases in the number of microsatellite repeats (*b*) occur when the new strand slips down one repeat in its binding to the old, template strand, causing the polymerase to add an extra repeat in the new strand to fill the gap. Decreases (*c*) happen when the old strand slips, which results in repair enzymes deleting a repeat.



out of every 100 to 1,000 daughter cells will carry a mutation that changes the number of *CTCTT* repeats. This change can have a dramatic effect on the *Opa* genes, because genetic information is read in "words" of three bases, called codons. Proteins are strings of amino acids, and each codon specifies a particular amino acid in the protein chain. Because the repeat is not three bases long, an increase or decrease in the number of repeats shifts the meaning of all the subsequent codons.

In the case of the *Opa* genes, deleting a *CTCTT* repeat leads to the production of a protein that is shortened and cannot adhere to host cells; in consequence, the bacterium bearing the shortened protein becomes unable to enter those cells. But subsequent slippage has a good chance of adding the repeat back, thereby allowing the *Opa* gene to produce a functional protein once again.

This reversible switching, called phase variation, has been found in many disease-causing bacteria. By switching its various *Opa* genes on and off from one generation to the next, *N. gonorrhoeae* can increase its chances for survival. There are times, for instance, when it is useful for the microbe to stick to and enter host cells, such as when the bacterium is spreading to a new host. At other times, it is strategically more advantageous for the bacterium not to interact with host cells—particularly phagocytic cells, which engulf and destroy bacteria.

The implications of slipped-strand mispairing for the ability of a bacterium to vary its surface molecules have also been studied extensively in *Hemophilus influenzae*. Type b strains of this bacterium are a primary cause of the life-threatening brain infection bacterial meningitis. Until the advent of a vaccine in the late 1980s, roughly one in every 750 children younger than five years of age contracted *H. influenzae* meningitis.

The outer membrane of *H. influenzae* is studded with molecules of fats and sugars joined together to make a molecule called lipopolysaccharide (LPS). One part of LPS, called choline phosphate, helps *H. influenzae* stick to cells in the human nose and throat, where the bacterium normally lives without eliciting symptoms. At least three of the genes required for making LPS contain microsatellites built from the four-base sequence *CAAT*. As is true of the microsatellites of the *Opa* genes of *N. gonorrhoeae,* changes in the number of *CAAT* repeats in these genes can cause *H. influenzae* to make LPS that either has or lacks choline phosphate.

Jeffrey N. Weiser of the University of Pennsylvania has shown that strains of *H. influenzae* that have choline phosphates on their LPS molecules—so-called ChoP+ strains—colonize the human nose and throat more efficiently than strains without them, which are referred to as ChoP– strains. Without ChoP, however, the bacterium is more resistant to being killed by various factors present in the host's blood and in other tissue fluids. The bacterial cells can switch between the two states, depending on whether they are being left undisturbed to grow in the respiratory tract or are spreading through the blood to other sites, where they are likely to be attacked by components of the immune system.

Most *H. influenzae* bacteria isolated from humans are ChoP+ variants, which are susceptible to the immune attack. ChoP– variants inevitably arise through slipped-strand mispairing, but they usually do not persist in the respiratory tract, because they adhere less efficiently to host cells than ChoP+ strains. But if the host contracts a viral infection that inflames the nasal tissues, the inflammation can increase the exposure of the bacteria to defense proteins of the host's immune system. In that case, ChoP– variants would have an advantage because they can fend off such an attack. Once the viral infection subsides, ChoP+ mutants generated by further slipped-strand mispairing of microsatellite DNA will once again predominate.

Genes such as these that can switch on or off readily have been named contingency genes for their ability to enable at least a few bacteria in a given population to adapt to new environmental contingencies. The variety of traits encoded by contingency genes includes those governing recognition by the immune system, general motility, movement toward chemical cues (chemotaxis), attachment to and invasion of host cells, acquisition of nutrients and sensitivity to antibiotics. Contingency genes make up a very small fraction of a bacterium's DNA, but they can provide a vast amount of flexibility in functioning. If only 10 of the 2,000 genes in a typical bacterium were contingency genes, for instance, the bacterium would be able to display 2^{10}—1,024—different combinations of "on" and "off" genes. Such diversity ensures that at least one bacterium in a population can survive its host's immune or other defenses and then can replicate to produce a new, thriving colony.

Causing disease—which can backfire by killing the life-giving host—may be one of the prices that bacteria pay for their ability to produce so many variants. The occasional variant may stray beyond its usual ecological niche in the host. It may penetrate the cells lining the respiratory or intestinal tracts, for example, to yield a potentially fatal infection elsewhere in the body. Provided that such events occur rarely, however, the

Searching for Papa Chimp

Besides the well-publicized use of microsatellite DNA to nab criminals through DNA fingerprinting, microsatellites are also being used to aid conservation efforts through study of the sex lives of endangered animals.

DNA fingerprinting, which distinguishes people by differences in selected regions of their DNA, is possible because the lengths of microsatellite DNA sequences differ between individuals. Scientists create DNA fingerprints by using special enzymes to make millions of exact copies of various microsatellites from each subject and then separating the copies by size on a gel. The result is a pattern of bands that looks much like a bar code—and that is almost as unique to each individual as a fingerprint is.

Pascal Gagneux and David S. Woodruff of the University of California at San Diego—together with Christophe Boesch of the Zoological Institute of the University of Basel—have used DNA microsatellites as tracers to probe the mating habits of a group of wild chimpanzees in the Taï Forest of Ivory Coast. They collected hairs from the temporary treetop nests each animal built to sleep in and extracted DNA from cells clinging to the roots of the hairs. By comparing the microsatellite DNA fingerprints of the adult males and females with those of 13 offspring, Gagneux, Woodruff and Boesch found that seven of the babies could not have been fathered by males in the group. Although the researchers had never seen them doing it, at least some of the female chimpanzees must have sneaked into the surrounding forest during the night for trysts with males in other groups nearby.

NIGEL J. DENNIS *Photo Researchers, Inc.*

FEMALE CHIMPANZEES, such as this one with her baby, often sneak away to breed with males outside their own groups, according to DNA-fingerprinting tests based on microsatellites.

Such nocturnal adventures might explain how even small groups of chimpanzees maintain a great deal of genetic diversity. Diversity is valuable for providing resistance to disease and is strongly suspected to aid survival in many other ways.

Preserving such variety is likely to be essential to the survival of wild chimpanzee populations. Unfortunately, as these populations become more and more fragmented and separated by longer distances, the ability of females to find males in other groups and to bring new genes into their group will be curtailed drastically. —*E.R.M. and C.W.*

Detecting Cancer

Microsatellite DNA may soon be affecting our lives in an important way: it may improve the early detection of cancer. Tests for mutations in genes that in their altered forms predispose to cancer, such as *p53* and *ras,* can now be used to detect as few as one cancer cell out of 10,000 normal ones. But mutations in these genes do not occur in all cancers or even in all cancers of a given type.

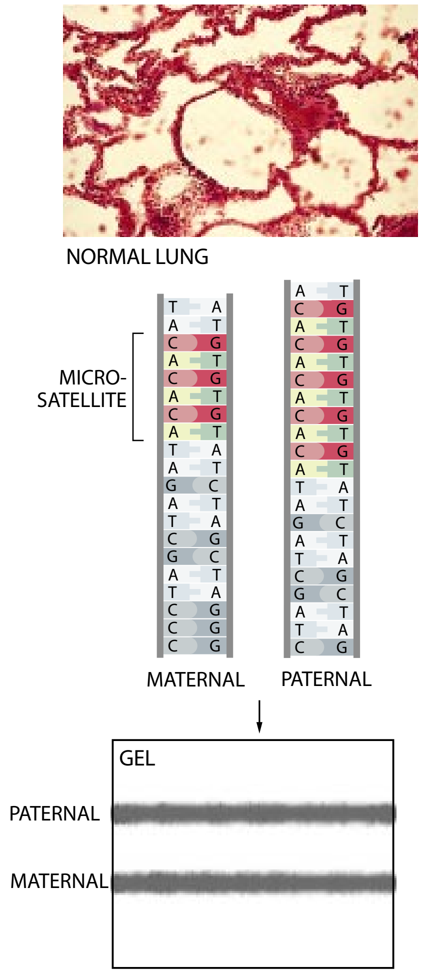
Microsatellites provide another method for early cancer detection because the overall rate of microsatellite expansion or contraction in cells turns out to be markedly increased in some types of cancers. Such bursts of change, often involving many different microsatellites, can be detected fairly easily. The approach can currently detect one cancerous cell out of about 500 normal ones.

Microsatellite changes in cancer cells were first found in 1993 by Manuel Perucho of the California Institute of Biological Research in La Jolla, Calif., who was studying hereditary nonpolyposis colon cancer. Perucho noted that many microsatellites from cancer cells were either longer or shorter than those in normal cells from the same patient. It was soon shown that one of the defects causing these alterations was in a gene encoding an enzyme responsible for correcting the length of microsatellites that grew or shrank during DNA replication; loss of the functional gene would presumably increase the likelihood that the errors would go uncorrected.

The circle of proof was closed when Richard C. Boland of the University of California at San Diego and others inserted a human chromosome carrying a normal DNA-repair gene into colon cancer cells grown in the laboratory. They observed that the inserted gene corrected the tendency for microsatellites in the cancer cells to mutate.

Striking as these findings are, however, microsatellite instability may be more a symptom than a cause of cancer. Although so-called knockout mice that lack the gene encoding one of the major mismatch-repair proteins live for only a short time and acquire many types of cancer, none of the cancer cells show increased levels of microsatellite mutations. It appears that such alterations form only part of the great variety of genetic changes that can cascade through the genome of a cell once the process of carcinogenesis has been set in motion—meaning they might be by-products of the carcinogenic process rather than contributors to it.

Nevertheless, the associations occur often enough that microsatellite instability provides clinicians with a new and powerful tool. Successful clinical trials of early detection systems employing microsatellites have been carried out for colorectal and bladder cancers and are now being extended to many other types of cancer, although none of the tests are now available outside research settings. As clinicians gain experience with these various patterns, not only will cancers be detected sooner than ever before, but the pattern of microsatellite variation will provide strong indications of the type of cancer involved. *—E.R.M. and C.W.*



TEST FOR CANCER spots changes in the lengths of microsatellites, such as those composed of repeats of the sequence *CA.* Normal lung cells (*left*), for example, have *CA*-repeat regions of two different lengths—one inherited from an individual's mother and one

benefits of contingency genes for the survival of a bacterial species outweigh the disadvantages of killing some hosts.

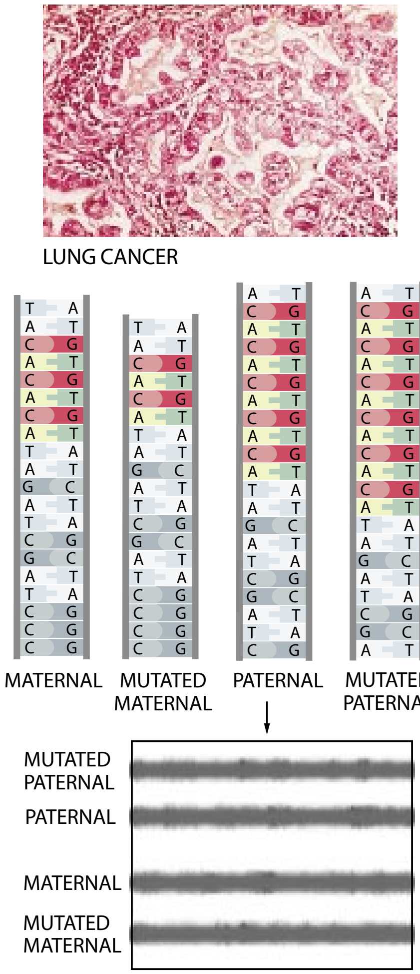
The microsatellites of these bacteria are true evolutionary adaptations. It is implausible that such unusual repeats could have arisen by chance; they must have evolved and been retained because they enable bacterial populations to adapt rapidly to environmental changes.

Microsatellites in People

Useful as they are, contingency genes are apparently confined to bacteria. The role of microsatellites seems to be very different in eukaryotic organisms like ourselves, whose cells contain a nucleus. None of the eukaryotic microsatellites identified to date appear to scramble the way DNA is read and to yield nonfunctional proteins. Most lie outside genes, but roughly 10 percent actually fall within them. Of this 10 percent, almost all are so-called triplet repeats, which tend to expand or contract in units of three bases. Just as adding or deleting an "and" or a "the" in a sentence rarely obscures its meaning, triplet repeats can expand or contract without disturbing a gene's message. Having the same length as a codon, they may simply lead to insertion or removal of a few repetitive amino acids without changing the sequence of all the others down the line.

So what are the functions of microsatellites in higher organisms? Scientists suspect that at least some of them must have uses, because eukaryotes have more microsatellites than bacteria and many of them happen to be in or near genes involved in pathways regulating fundamental cellular processes. Only a few hints have yet emerged, however, about what these purposes might be.

The few effects that have now been traced to eukaryotic microsatellites have generally been harmful. For example, the grim neurodegenerative disorder Huntington's disease—characterized by late-onset dementia and gradual loss of motor control—is triggered by a flawed version of a gene that codes for a large protein, huntingtin, of unknown function. The normal gene contains a long, triplet-repeat microsatellite that adds a string of amino acids called glu-



BIOPHOTO ASSOCIATES/SCIENCE SOURCE (*top left*); PARVIZ M. POUR *Photo Researchers, Inc.* (*top right*); GEORGE RETSECK (*illustrations*)

from the father—that can be sorted by size on a gel. Early tumors, such as adenoma of the lung (*right*), are made of both normal cells and cancer cells, which can be detected because their microsatellites have shortened or lengthened to yield more than two bands on a gel.

tamines near the start of the protein.

The number of glutamines at the beginning of the huntingtin protein usually ranges from 10 to 30. But people who have—or who are destined to develop—Huntington's disease carry a microsatellite coding for an unusually long run of 36 or more glutamines. Inheritance of just one copy of the flawed gene, from the mother or the father, is enough to ensure eventual illness. It is not yet clear how the long stretches of glutamines contribute to Huntington's.

More than a dozen such triplet-repeat diseases are now known; most are rare neurological diseases. About half the disease-causing microsatellite repeats are inside a gene, and most encode glutamines. The rest are sufficiently close to nearby genes that they can affect their function.

One of these rare neurological diseases—spinal bulbar muscular atrophy—results from expansion of a microsatellite inside a gene on the X chromosome; the gene codes for a receptor for the male hormone androgen. People with 40 or more triplet repeats in part of one of their androgen receptor genes develop the disease. But a group led by E. L. Yong of National University Hospital in Singapore has demonstrated that repeats that are even slightly longer than normal can also have medical effects. They reported in 1997 that men with between 28 and 40 repeats in the part of the androgen receptor gene that encodes glutamines were likely to be infertile.

Too few triplet repeats in the androgen receptor can also have untoward consequences. Several other research groups have shown that men with 23 or fewer repeats have an increased risk of prostate cancer. Such cases are unusual, however.

Evolving Evolvability

Why do we have all these genetic time-bombs ticking inside our genomes? It is striking that so many of our triplet-repeat diseases involve neurological function and that none of those linked to triplet repeats in humans have yet been reported in other primates, such as chimpanzees. If such diseases turn out to be unique to humankind, they might represent a genetic cost we have incurred because of the rapid evolution of our brains. It is possible that long microsatellites at or near certain genes might contribute to brain function and might therefore have persisted throughout evolutionary time even though they occasionally expand too much and cause disease.

In 1989 one of us (Wills) postulated on theoretical grounds that some genes have evolved the ability to evolve. According to the hypothesis, in an environment that fluctuates in some predictable way—such as growing warmer or cooler—possessing the genetic apparatus to evolve quickly would have advantages. The contingency genes of bacteria have turned out to be excellent examples of evolvability genes: their high rates of forward and backward mutation allow bacteria to adapt rapidly to predictable environmental changes and then to revert back again when the earlier conditions reappear.

Perhaps eukaryotic microsatellites exert a more subtle form of regulation than that provided by bacterial contingency genes. In humans, microsatellites within genes have been found that influence the production rate of a number of proteins, ranging from the bile pigment bilirubin to neurotransmitters, the chemicals that carry messages between nerve cells. David G. King of Southern Illinois University has suggested that such microsatellites may be "tuning knobs" that evolved to act as rheostats for gene function, turning up the amount of protein produced by a gene in some instances and decreasing it in others.

Indeed, Walter Schaffner and his colleagues at the University of Zurich have shown that adding microsatellites that encode runs of glutamines or prolines (another amino acid) at the start of a known gene can increase its ability to yield protein. Perhaps, because it is so much less disruptive than contingency gene switching, this form of gene regulation emerged during the evolution of complex, multicellular organisms.

Scientists have only begun to probe the roles of microsatellites in our own species. It may be that the repeats, with their ability to switch rapidly among a limited number of states, will provide insights into our own capacity to adapt to environmental change, just as contingency genes have done in bacteria. SA

The Authors

E. RICHARD MOXON and CHRISTOPHER WILLS approach the study of DNA microsatellites from two different vantage points. Moxon is a pediatrician and an infectious disease specialist; Wills is an evolutionary biologist. Moxon currently heads both the department of pediatrics at the University of Oxford and the Molecular Infectious Diseases Group of the Institute of Molecular Medicine at John Radcliffe Hospital in Oxford, England. Wills is a professor of biology at the University of California, San Diego. He is the author of several popular books on evolutionary biology, including *Yellow Fever, Black Goddess* (Addison-Wesley, 1996) and *Children of Prometheus: The Accelerating Pace of Human Evolution* (Perseus Books, 1998).

Further Reading

Adaptive Evolution of Highly Mutable Loci in Pathogenic Bacteria. E. R. Moxon et al. in *Current Biology,* Vol. 4, No. 1, pages 24–33; January 1, 1994.

Evolutionary Tuning Knobs. D. G. King, M. Soller and Y. Kashi in *Endeavour,* Vol. 21, No. 1, pages 36–40; 1997.

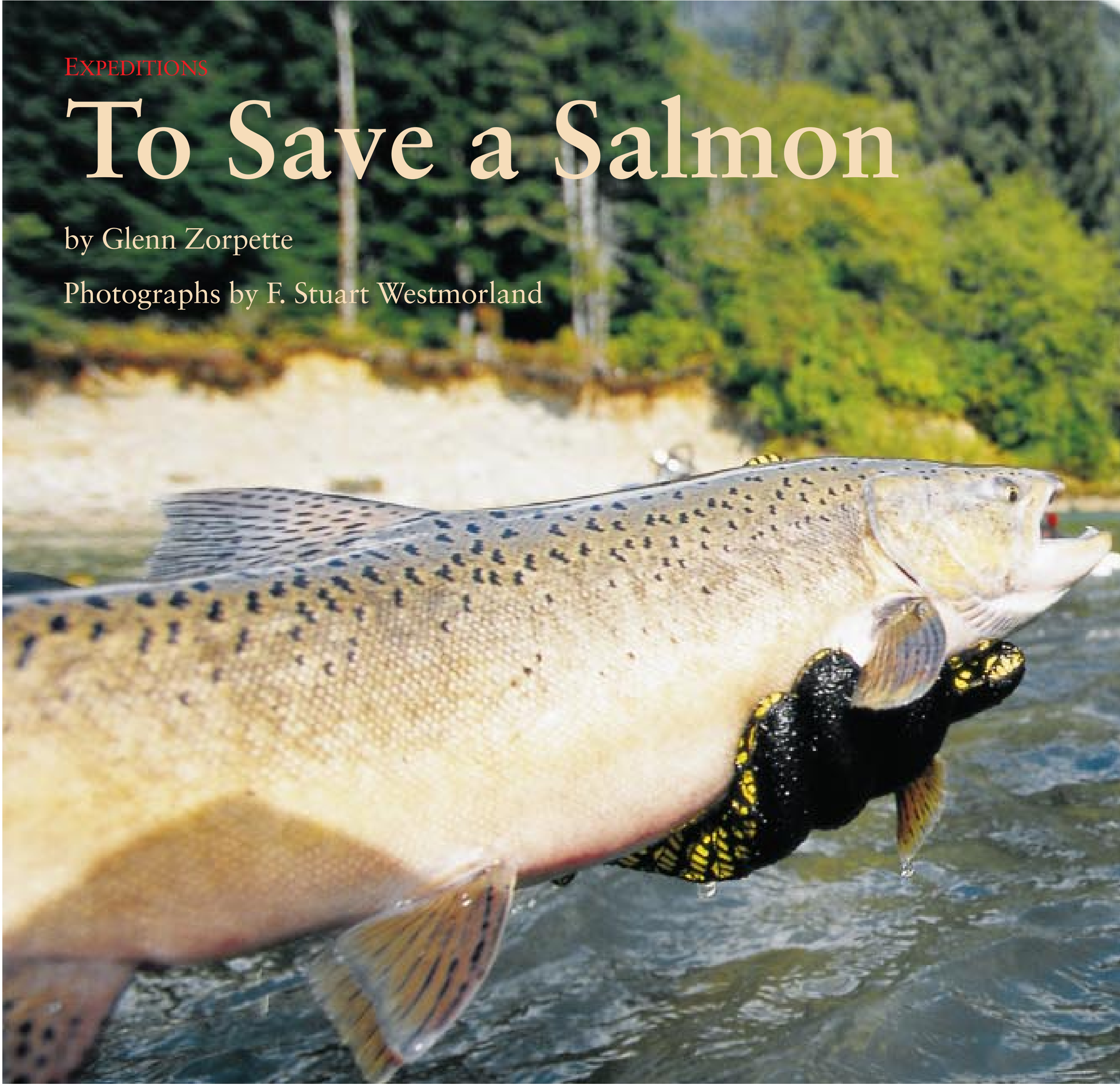
Furtive Mating in Female Chimpanzees. P. Gagneux, D. S. Woodruff and C. Boesch in *Nature,* Vol. 387, pages 358–359; May 22, 1997.

EXPEDITIONS

To Save a Salmon

by Glenn Zorpette

Photographs by F. Stuart Westmorland



On a misty, gray September morning in Port Alberni Canal, on the western coast of Vancouver Island, the 20-meter-long fishing boat *Ganges I* floats peacefully at anchor. Long, wispy clouds are grazing the tops of the rugged hills that line the edges of the inlet like the sides of a deep green canyon. Off the starboard side, a trawler has just puttered westward, but otherwise the only other disturbances are from jumping coho salmon, which break the dark, lightly rolling waves, and from the occasional livid, gesticulating sport fisherman cruising by.

The leaping, silvery fish are on their way to the inland rivers of their birth to spawn and then die, a tiny part of the ancient and still mysterious pageant that has intrigued and sustained people along the borders of the earth's temperate zones for tens of thousands of autumns. The trawler, one of 13 on the canal, is there to intercept the fish—not to sell them, but as part of a weeklong experiment on commercial fishing techniques being run by researchers on board the *Ganges I*.

The angry sport fishers would also like to be snagging tasty cohos, one of the most celebrated of game fish. But at about U.S.$100 apiece, the fish would be expensive. That's the fine for taking a coho, because on the advice of scientists—including a few now on board the *Ganges I*—the government has made it illegal to keep the fish. Yet today, to the indignant amazement of the "sporties," scientists are out in the canal directing the small fleet of commercial boats as they haul in the magnificent creatures and then kill some of them.

Much of the sport fishers' misguided fury is directed at

On Vancouver Island, fisheries scientists are trying to find out whether commercial fishing and cohos can coexist

Brent Hargreaves, the 44-year-old biologist directing the project. They do not realize that this is no ordinary trawling fleet. The 13 boats in the canal today have been mustered to Hargreaves's command to catch thousands of coho and to kill several hundred of them, in the hope of eventually saving millions.

Populations of coho and, to a lesser extent, chinook salmon have declined sharply in British Columbia in recent years, even as their cousin species—pinks, chums and sockeyes—have held their own. While some biologists struggle to understand what is decimating the prized fish, others are taking admittedly desperate measures to resuscitate the flagging species. These fisheries experts are focusing their efforts on two critical points in the salmon's life span: their birth in rivers and streams and their death in the nets of fishing trawlers that intercept the fish as they swim back to freshwater to spawn.

There is more at stake here than just the future of fishing in a corner of the world where the only other industry of any size converts trees into paper pulp. "I see it as part of a larger picture," says Robert Bell-Irving, a fisheries technician with the Canadian Department of Fisheries and Oceans (DFO). "Look at the buffalo and the carrier pigeon in North Ameri-

CHINOOK SALMON was captured in a seine net in the Phillips River, north of Vancouver in British Columbia. At eight or nine kilograms (20 pounds), the fish is not very large for a chinook; the largest specimens can reach 40 kilograms.


FISHERIES BIOLOGIST Brent Hargreaves (*right*), on the fishing boat *Ganges I,* mulls over a fish-tagging problem on the second day of his experiment in Port Alberni Canal, on the western coast of Vancouver Island. Elsewhere in the canal, the *Island Spirit,* one of 13 commercial fishing vessels taking part in Hargreaves's experiment, deploys its seine net from a reel on the stern of the boat.

ca. Salmon are our last chance to show that we can live with a natural abundance without destroying it."

Bell-Irving, a burly, soft-spoken man with an easy smile, is trying to bolster chinook at their birthplace on the wild and remote Phillips River, north of Vancouver. Even in this uninhabited watershed, chinook runs have fallen by as much as 90 percent since World War II.

On many rivers in the more developed parts of southern British Columbia, the numbers of chinooks, the largest of the Pacific salmon that spawn in North America, have dropped from the thousands to a few hundred, threatening the genetic viability of the stock. For coho salmon, which reproduce in small streams, at least 70 percent of their favored spawning and rearing grounds have been eliminated in the more populated areas.

The declines, which began more than half a century ago, have been attributed to logging, hydroelectric development, mining, urbanization, overfishing and, lately, climate changes. The situation is so precarious that some experts assert that Canada's array of federal and provincial efforts to boost flagging populations will at best prop up the stocks until another climate shift—perhaps the end of the current El Niño cycle—enables more salmon to survive in the ocean and return to freshwater to spawn.

Commercial fishing for coho was essentially banned last autumn throughout the Canadian Pacific, and trawling for pink, chum and sockeye salmon was significantly reduced, even though those species were not nearly as threatened as coho. Why? Because commercial fishers have no way to catch only certain species of fish while assuring the well-being of others. Cohos that wind up in a net full of sockeyes generally die in the net, on board the boat or in the water shortly after they are thrown back.

Hargreaves thinks there is a better way. His $600,000 experiment here in Port Alberni Canal is testing new fishing gear designed to spare the coho while snaring the more abundant species. Five of the commercial trawlers participating in the experiment are fishing just as they normally would, using ordinary purse-seine nets or gill nets. Purse-seines trap the creatures in a long expanse of net that is trailed behind the boat for a while and then gradually winched into a large purse containing hundreds or thousands of fish. Hauled on board, the package typically crushes the fish on the bottom. Gill nets, which have a larger mesh, snare fish by the gills, drowning many of them.

The eight other boats are evaluating experimental gear or techniques that have been modified to make them more hospitable to the incidental catch. For example, two of the trawlers are using new kinds of "brailers." These devices, resembling huge, black wind socks, funnel the fish from the seine net to a sorting box on the boat. The new brailers are made of slick vinyl rather than the standard netting, so the fish slide down easily without having their scales scraped off.

On day two of the experiment, Hargreaves is up against the kind of calamities that harass any scientist who bothers to work in the field. On board the *Ganges I,* which has been converted into a floating command center, he is dividing his time between the wheelhouse and the galley, where two fisheries scientists are entering data about catches into laptop computers. In the wheelhouse, radios squawk almost continuously with tallies of the fish caught in each "set" (a single deployment of a net), as well as with questions and complaints about the tagging gear and, every so often, profane invective from sport fishermen.

The headache du jour is the tagging guns being used to tag each coho behind the dorsal fin. The scientists have been wasting many tags before getting one to attach. (Unbeknownst to Hargreaves at this point, the guns were delivered with the wrong tags.) So all morning there has been a stream of anxious griping as the supply of tags dwindles.

It is a little after 11 A.M. Hargreaves, a youthfully handsome man with a deep voice, firm jawline and piercing gaze, grabs the microphone of the wheelhouse radio and intones reassuringly: “Don’t panic about the tags here. We’ve got a bunch more tags in Nanaimo, and they’re coming out this afternoon.”

“Take a valium,” adds Ken Widsten, the captain of the *Ganges I,* after Hargreaves has released the microphone’s “talk” button. Then, to the unsmiling Hargreaves, he offers: “Just five more days.” Outside a window, over Hargreaves’s shoulder, a sport fisherman is glimpsed cruising by in a motorboat, middle finger upraised.

Sporties tend to regard cohos as “their” fish. Although some trolling boats do go after cohos—or did when there were enough to go after—most commercial vessels seek out the much more abundant pinks, sockeyes and chums. So the scuttlebutt around the piers—that a bunch of commercial fishers and government scientists are out in the canal killing cohos—galls the sporties to no end.

In fact, the fisheries people are killing relatively few fish to sample their blood for biochemical indicators of stress. Researchers will later compare the results with the survival rates of other cohos kept for observation in net pens on the inlet. “This is one of the rare occasions when we’ll be able to know the survival rates, so we’ll find out whether these blood-chemical tests mean anything or not,” Hargreaves explains.

Unfortunately, the sporties don’t know the reason behind the bloodletting. Fortunately, however, they also don’t know that lunch on the *Ganges I* is a delicious steamed coho with lemon juice and pita bread.

Hargreaves has already had his share of abuse from the sporties. The night before, after an exhausting first day, several of them accosted him on a pier in Port Alberni. To mollify them, he invited them to see the work today for themselves. But now, with the tagging crisis, Hargreaves has not yet sent someone to meet the sporties, who have been waiting and fuming on the pier for about an hour. After one of them begins haranguing Hargreaves over the radio, he calmly but wearily replies, “I think I’m hearing more attitude than I want to deal with today.” The belligerent backs down a bit, and a boat is soon dispatched to pick up the group.

A couple of hours later the morning’s mist has burned off to leave a dry, spectacular afternoon, with sunlight bouncing blindingly off the waters of the inlet. On board the seiner *Ocean Venture,* captain Glenn Budden and his crew are testing one of the experimental brailers, which Budden designed himself.

Budden steers the boat along an arc as the net unspools into the water from a huge reel on the back of the boat. After the net soaks for about 15 minutes, the crew reels in a cable attached to the far end of the net. The long arc of mesh tightens into a circle that slowly shrinks into a small pen, bulging with a few hundred fish, hanging off the starboard side of the boat.

Now it’s show time for Budden’s brailer. The crew scoops fish by the dozen into the mouth of the five-meter-long chute, through which the animals flop, thrashing, into a plywood sorting box. The men quickly pick out the

COHO TAGGING starts off with a shock applied by a metal paddle (*top*) to stun the fish. A bit of flesh is then taken from the gill plate (*middle*) for DNA testing. Finally, a tag is inserted behind the dorsal fin (*bottom*).

NETTED SALMON on the remote Phillips River include several chinooks, including the one at the left, and some pinks. In the background, volunteer workers prepare to pull in the ends of the net, shrinking the size of the "purse" containing the fish.

eight or so silvery cohos, letting the several hundred other fish, mostly sockeyes, fall from the box into the water.

Budden holds one of the writhing, slippery, 12-kilogram packages of muscle on a paddlelike device made of a metal screen. He presses a button on the device and a mild shock stuns the fish. Working fast, a fisheries official measures and tags the fish and uses a paper punch to get some flesh from the gill cover for DNA testing. The tags will be important several weeks from now, when observers will look for them on cohos in inland rivers. Data from the tags will help establish which fish survived long enough to spawn. (The DNA data provide a backup way to identify fish whose tags are lost.) Cohos are being killed for blood samples on a single trawler, a different one each day, so that the stress levels for the different experimental techniques can be compared.

After the set, Budden, in rubber waders and a baseball hat, rests and smiles through a scruffy beard in the wheelhouse of his boat. He has been a commercial fisherman for 26 years. "I need to make a living, but I'll sure need to make a living next year and four years from now, too," he says, as radios crackle and chatter in the background. "We want to prove to the DFO that we can change our gear to be more friendly to whatever they deem to be in trouble. Lots of people say it can't be done. We're going to prove it can be."

Budden may be right. But even if he is, selective fishing won't be enough by itself to revive runs of coho and chinook. So on many rivers, fisheries experts and volunteers are working a complementary angle, trying to increase the number of very young salmon, called smolts, that swim downstream.

The juveniles are taken alive and kept in a net pen until the females are ready to spawn. Then workers kill the fish, remove their eggs and stir the roe in a bucket with semen to fertilize them. The resulting creatures are incubated in a hatchery. For chinooks, this incubation occurs from October until March or April, after which the minute fish are released into the river as "fry" weighing about half a gram. By June, the fish have grown into smolts that, at five to 10 grams, are large enough to migrate down to the sea. Hatchery incubation can boost egg-to-smolt survival rates for the chinooks from roughly 5 percent to 80 to 90 percent. So more fish make it to the ocean—where, generally, less than 1 percent will survive to come back in a few years to spawn.

MAN AND FISH—specifically, salmon expert Robert Bell-Irving and a male pink salmon—meet in the Phillips River.

Compelling numbers like these are why, on another bright September day, 11 people have volunteered their time to don thick wet suits and round up chinooks in the glacially cold Phillips River. The mountains towering all around are covered with Douglas firs, red and yellow cedars, Sitka spruces, hemlocks and alders. There is also a fair amount of grizzly bear scat, which to this city dweller seems distressingly fresh. Reassuringly, though, one of the volunteers has brought a rifle. The others are apparently relaxed enough to have brought children and dogs, who play happily along the banks.

The volunteers have deployed a seine net into the river and pulled it into a bag of a few thousand closely packed, wriggling, splashing salmon. Bell-Irving and some of the crew are standing around the net, lunging and grasping for the 20 or 30 big, brassy, spotted chinooks swimming among the pinks that outnumber them 100 to one. Underwater the pinks are visible through the brown murk kicked up by all the activity, behind the net's green mesh. Distinctively humpbacked at this stage of life, they are gaping at the edge of the net, on which several of them have entangled their baroquely hooked lower jaws.

Bell-Irving, in black neoprene and a blue hood, leans over the white polystyrene floats holding up the net. He dips his face into the water and peers through his mask into the seething, swirling mass of fish. Suddenly, he lifts his head. "There's a buck here," he announces. "He's a dandy."

Deftly, he embraces the slippery chinook, as long as a car door, and slides it into a cylindrical bag. The salmon is one of 18 from the set that are mature enough to be added to the dozens in the net pen nearby on the river. The total will ultimately reach 70 chinooks—a good season's tally, according to Wallace R. Parker, one of the volunteers. All last year, he adds, they caught only four.

Hargreaves also found more fish than expected. One of the seine boats in his experiment pulled in more than 900 cohos in one set, an astounding number. "A week ago everyone was arguing that we wouldn't get enough coho to make the experiment work," he recalls. He and I are sitting on the bow of the *Ganges I* as day two of his experiment comes to an end. We're squinting into the setting sun and breathing the heavy perfume of the cedar logs floating and creaking around us in booms in the inlet. Twenty minutes ago one of his assistants showed him a copy of the day's local newspaper, the *Alberni Valley Times,* with Hargreaves's photograph on page one along with a story about his experiment. Hargreaves had only enough time to grin at it before getting a report on a flooded outboard motor.

For the contained Hargreaves, even the grin was surprising. The only other sign that he is more relaxed now than six hours ago is the laugh that erupts from him when, asked to list some of the day's stressors, he rattles off five or six before including "finding time for the media"—a pointed reference to me.

He is relieved that the tagging crisis ruined only three sets. Basically, it was the kind of routine contingency that separates the field-workers from the theorists. "It's easy to come up with experiments while you're sitting in an office," he explains. "But it's out in the real world—with weather and boats and fish and with politics attached to it—that's when you find out if you can do fieldwork or not."

There is good news on the political front as well. After their tour of the experiment, the sport fishermen went from furious to fulsome. "They were extremely hostile last night," Hargreaves recalls. "When they left today, they said if there's anything they can do to help, just ask."

Perhaps it is his adversaries' sudden contrition, or perhaps it is the smell of the cedar and the sight of the big, dazzling sun dropping into the mountains, setting the water ablaze in pale yellow glitter, but the normally reserved Hargreaves seems almost, well, inebriated. "I love it out in the field," he says at last. "It is the best part of this job." SA

THE AMATEUR SCIENTIST

by Shawn Carlson

Taking the Earth's Magnetic Pulse

Most folks think about the earth's magnetism only when they need to find their way in the wilderness. When we look at a compass, the earth's magnetic field appears to be a steady guide. But in reality the magnetic field is anything but stable. Ephemeral undulations, called micropulsations, ripple about the ionosphere and generate magnetic disturbances that reach down to ground level. Although they are common and sometimes last from seconds to minutes, these magnetic disturbances are hard to detect, having barely one ten-thousandth the strength of the earth's average magnetism.

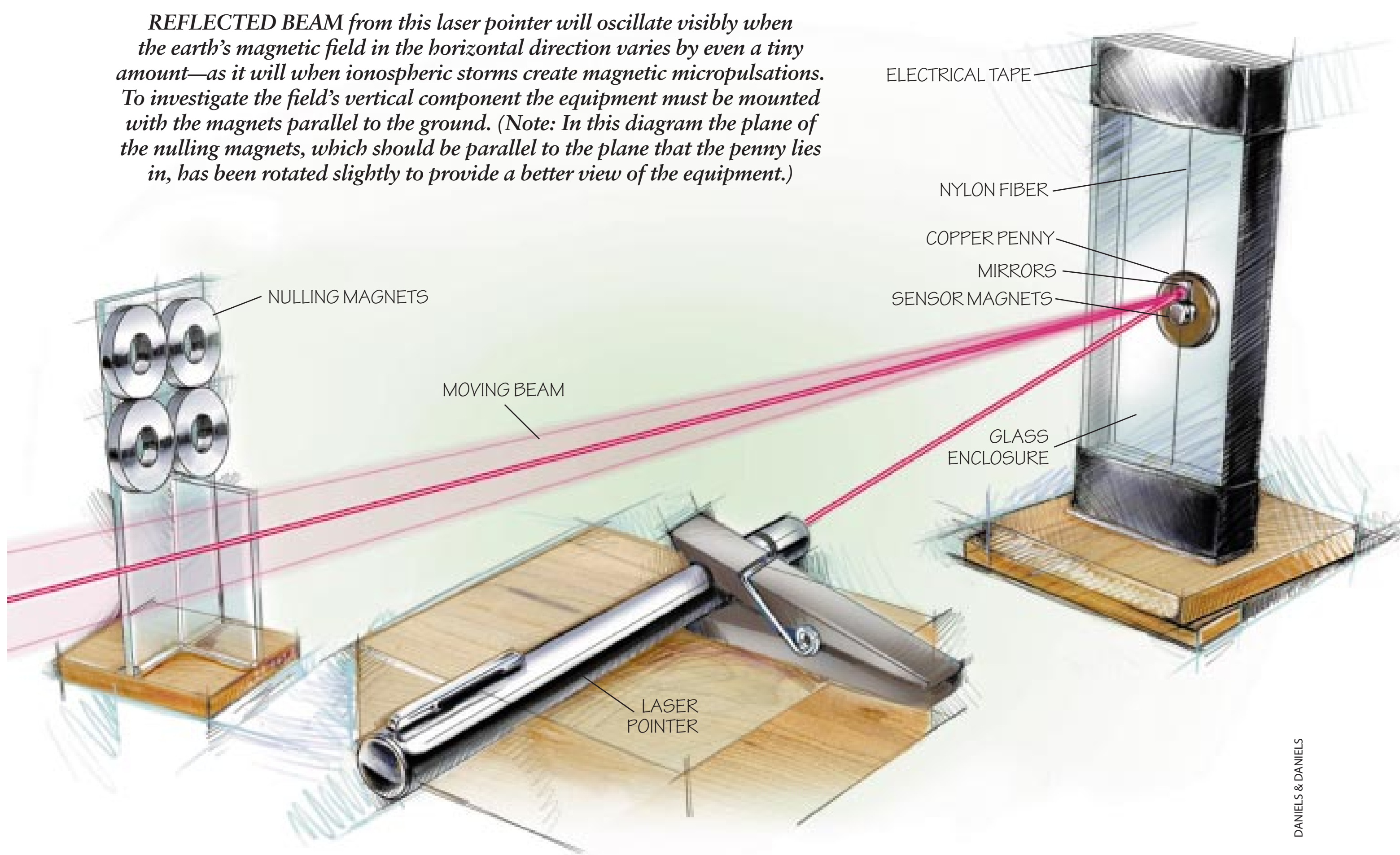
For decades, the high cost of sensitive magnetometers has made tracking these signals the exclusive privilege of professionals. But now, thanks to the creative genius of Roger Baker, a gifted amateur scientist in Austin, Tex., anyone can easily study magnetic micropulsations. Baker's magnetometer costs less than $50, yet it can easily capture those tiny pulsations as well as the occasional dramatic effects of a magnetic storm high in the ionosphere.

Baker's device employs one of the most sensitive instruments in science. It's called a torsion balance, and it measures a force by using it to twist a fine filament. The thread gently resists rotation with a torque that grows until it just balances the torque created by the applied force. The resulting angle of deflection, which is found by bouncing a light beam off a small mirror attached to the filament, is proportional to the force under study. With the beam from a laser pointer and a match-head-size mirror, one can in principle resolve deflections as small as one ten-millionth of a degree.

Most professional torsion balances use fine quartz fibers, which are incredibly strong and insensitive to changes in humidity and temperature. Sadly, quartz fibers are difficult for amateurs to come by. But Baker has found that nylon fibers also work quite well. Start with silky, multifilamented nylon twine, which you can purchase at any hardware store, and cut a 30-centimeter (one-foot) length. Next, gently unravel it and use tweezers to select the finest strands, which should be about 25 microns (0.001 inch) thick.

Baker installs the nylon filament into a simple case made from window glass. Cut two strips five centimeters

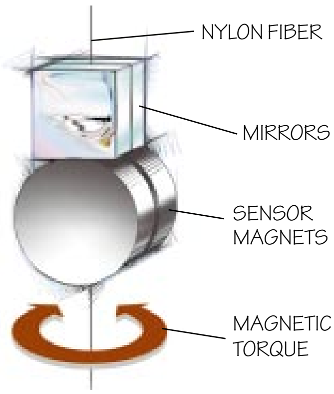
REFLECTED BEAM from this laser pointer will oscillate visibly when the earth's magnetic field in the horizontal direction varies by even a tiny amount—as it will when ionospheric storms create magnetic micropulsations. To investigate the field's vertical component the equipment must be mounted with the magnets parallel to the ground. (Note: In this diagram the plane of the nulling magnets, which should be parallel to the plane that the penny lies in, has been rotated slightly to provide a better view of the equipment.)



DANIELS & DANIELS

wide and 15 centimeters long using a glass cutter. (Most hardware vendors sell glass and will also cut it for a nominal fee.) These pieces serve as the vertical walls of the case. Then cut eight glass strips one centimeter wide by five centimeters long and use silicone cement to glue pairs together back to back. Finally, glue one pair of these small glass strips to the top and one to the bottom of each of the longer glass walls. The smaller pieces will act as spacers between the walls [*see illustration on page 106*].

When the glue sets, cover the horizontal spacers of one of the walls with a layer of stretchy, black vinyl electrical tape. The tape prevents the glass from cutting the fiber. Next, lay one end of the fiber across the center of the top spacer and tack it in place with a small dollop of epoxy; secure it with another strip of tape. The epoxy will keep the thread from slipping over time. Baker generates the necessary tension in the fiber by dangling four nickels attached temporarily to the end. He then epoxies and tapes the lower end of the thread into place against the bottom spacer, locking in the tension.



***MIRROR** pieces and a pair of powerful magnets attached to a nylon fiber rotate in response to changes in the magnetic environment.*

DANIELS & DANIELS

To coax the nylon to twist in response to minute changes in the ambient field, you need to affix a powerful magnet to the fiber. Because a large and massive magnet responds only sluggishly, the ideal attractor would possess a powerful field and yet be extremely lightweight. Such magnetic miracles exist; they are called rare-earth magnets because they contain rare-earth elements, such as samarium. These marvels are tiny and yet harbor at their surface magnetic fields that are 10,000 times stronger than the earth's. Best of all, you can pick up a pair of them at any Radio Shack for less than $2 (part number 64-1895).

Deposit a thin smear of silicone cement on one rare-earth magnet and sandwich the filament between the two of them. Make sure that they overlap each other completely and are perfectly centered on the fiber as the glue hardens.

Baker fashioned the reflector from a small vanity mirror. With a glass cutter, he cut 1.5-millimeter-square chips, which he cemented together, back to back, centered on the fiber, just above and in contact with the rare-earth magnets. The mirror and magnets will then rotate as a well-balanced unit. Baker notes that the instrument will work better if you use the reflective surface on the back of the mirror to reflect light, which avoids the possibility that passage through the glass will distort the beam. To remove the lacquer that covers the reflective coating, use a Q-Tip to rub a little methyl ethyl ketone (MEK) on the back surface. If you would rather not work with a potentially toxic chemical, then install the mirrors in the usual way, with light passing through the glass.

Baker next glues a solid-copper penny (one minted before 1982, when the purity of the metal was still high) to the glass, just behind the magnets. When the magnets move, they induce electrical eddy currents in the copper that in turn produce their own magnetic fields, which oppose the motion of the magnets. Baker's clever trick quickly damps unwanted oscillations, making the magnetometer much easier to read.

Next, encase the sensor by gluing the second glass wall with its spacers on top of the first and then seal off the sides with black electrical tape to protect your magnetometer from pesky air currents. Mount the entire assembly vertically to a smooth flat base. Your sensor is now an accurate compass. As you walk around, the magnets should align to magnetic north and display little oscillation.

Because the earth's relatively large field absolutely overwhelms the coveted magnetic micropulsations, you must first null the instrument before it will register signals from the ionosphere. Baker's procedure for doing so requires another trip to Radio Shack to acquire four doughnut-shaped magnets (part number 64-1888). Attach them side by side to a small piece of glass or wood using silicone cement. (Note that you'll need to use small clamps to hold them in place against their mutual magnetic repulsion until the glue sets.) Turn the array of doughnut magnets upright and cement it to a freestanding base so that the center of the assembly aligns with the rare-earth magnets in the sensor.

As the nulling magnets are brought close to the magnetometer (approximately 30 centimeters), the sensor will begin to wobble and then rotate quite freely when the combined forces of the earth's magnetic field and those of the doughnut magnets almost cancel each other out. The period of the oscillations should lengthen to a second or more when all the forces are nearly balanced. In this configuration the magnetometer will be the most sensitive.

The bright beam of a laser pointer completes the instrument. (Radio Shack sells one pointer for less than $30.) Position the laser so that the beam shines through the glass case and bounces off the mirror and onto a distant wall. Ripples in the earth's magnetic field will show up as deflections of the beam.

Although at this point you're ready to do real science, Baker has improved the device further. By wrapping wire onto a rolled-oats container, he has fashioned a pair of coils to calibrate his creation. He uses the so-called Helmholtz arrangement of two circular coils (ones separated by half their diameter) to make the magnetic field in the center uniform. Such Helmholtz coils can also be combined with photocells and electronic feedback to keep the laser beam fixed as the earth's field changes. Measuring the current needed to null the signal in this way shows the size of the magnetic fluctuations. Using a computer to monitor the current will produce a stream of measurements that you can record and later analyze. SA

For more details about this project, consult Roger Baker's Web site at www.eden.com/~rcbaker. And visit the Web forum hosted by the Society for Amateur Scientists at web2.thesphere.com/SAS/WebX.cgi. You may also write the society at 4735 Clairemont Square, Suite 179, San Diego, CA 92117, or call 619-239-8807.

by Ian Stewart

Division without Envy

Last month's column took a look at some of the mathematical issues that arise from the deceptively simple problem of dividing a cake fairly—meaning that if there are n people, each is convinced that his or her share is at least $1/n$ of the cake. This month we'll look at some related questions, which have given rise to the more modern parts of the theory.

A brief reminder of where we'd got. With two people, the time-honored algorithm "I cut, you choose" leads to fair division. With three or more people, there are several possibilities. The "trimming" method allows successive participants to reduce the size of a purportedly fair share of the cake, with the proviso that if nobody else trims that piece, then the last person to trim it has to accept it. In the "successive pairs" algorithm, the first two people divide the cake equally, and the third person secures what he or she considers to be at least a third of each piece by negotiating with each of the first two separately. And with "divide and conquer," participants try to divide the cake using one cut so that roughly half the people would be happy to have a fair share of one piece, while the rest would be happy to have a fair share of the other piece. The same idea is then repeated on the two separate subcakes, and so on.

These algorithms are all fair, but there is a more subtle issue. Even if everybody is convinced that he or she has a fair share of the cake, some may still feel hard done by, thanks to the Deadly Sin of envy. For example, Tom, Dick and Harry may all be satisfied that they've got at least a third of the cake; nevertheless, Tom may feel that Dick's piece is bigger than his. Tom's share is "fair," but he doesn't feel quite so happy anymore. A division of the cake is "envy-free" if no person thinks that someone else has a larger piece. An envy-free division is always fair, but a fair division need not be envy-free. So finding an algorithm for envy-free division is more difficult than finding a fair one.

Cut-and-choose for two people is easily seen to be envy-free, but none of the other algorithms mentioned above is. An envy-free algorithm for three people was first found by John Selfridge and John H. Conway in the early 1960s:

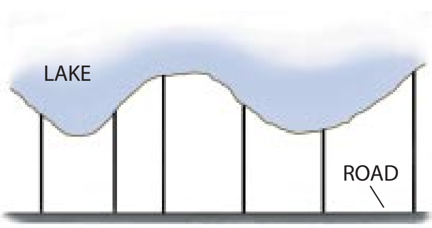
STEP 1: Tom cuts the cake into three pieces that he considers of equal value.

STEP 2: Dick may either (a) do nothing, if he thinks that two or more pieces are tied for largest, or (b) trim what he perceives as the largest piece to create such a tie. Set aside any trimmings; call the accumulated trimmings "leftovers."

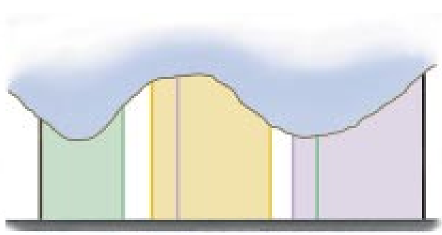
STEP 3: Harry, Dick and Tom, in that order, choose a piece—one that they believe to be either the largest or tied for largest. If Dick trimmed a piece in step 2, then he must choose the trimmed piece unless Harry has already done so.

At this stage, part of the cake has been divided in an envy-free manner. So it remains to divide the leftovers in an envy-free manner, too.

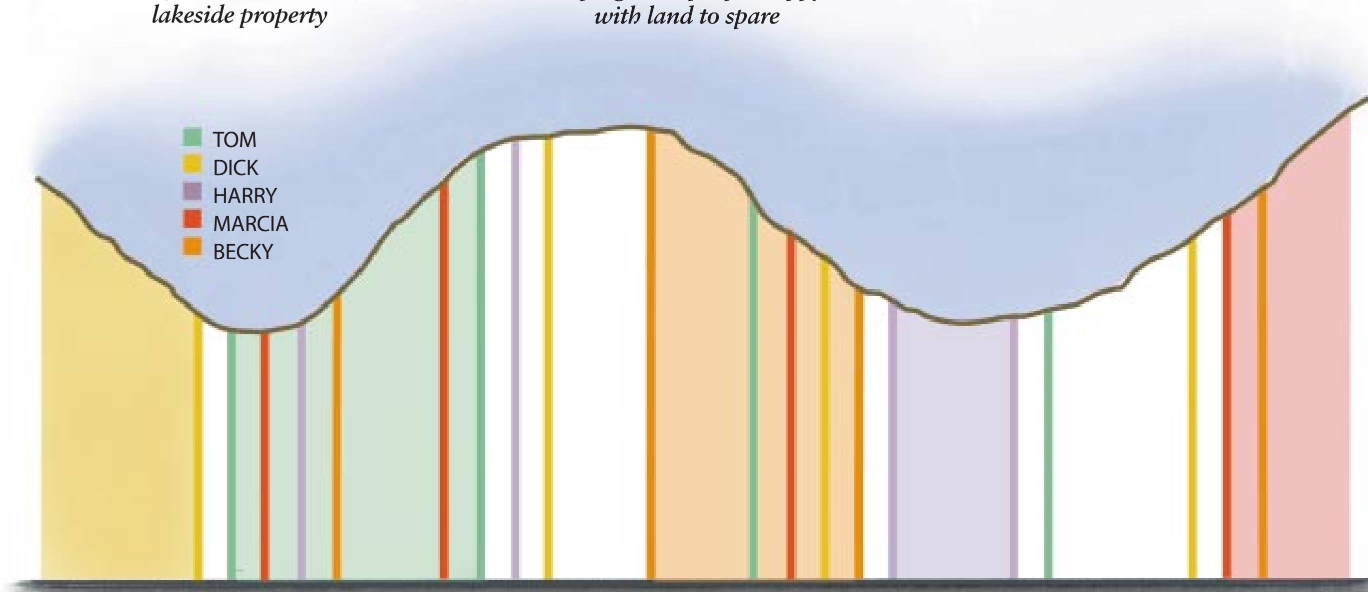
STEP 4: If Dick did nothing in step 2, there are no leftovers and the cake has been divided. If not, either Dick or Harry took the trimmed piece. Suppose



One person's division of lakeside property



Keeping three people happy, with land to spare



Keeping five people happy, with land to spare

MATT COLLINS

Dick took it (if Harry did, interchange those two people from now on in the description of what to do). Then Dick divides the leftovers into three pieces he considers to be equal.

STEP 5: All that remains is for Harry, Tom and Dick, in that order, to choose one piece from the leftovers. Then Harry has first choice and so has no reason to be envious. Tom will not envy Harry however the leftovers are divided, because the most that Harry can get is a piece that Tom is already convinced is worth a third. And he won't envy Dick, because he chooses before Dick does. Dick has no grounds for complaint, because it was he who divided the leftovers anyway.

At this point, everyone got stuck for 30 years. Is there an envy-free protocol for *n* people? In 1995 Steven J. Brams of New York University and Alan D. Taylor of Union College found a remarkable protocol for any number of players. It is distinctly complicated, and I won't give it here: either see their article "An Envy-Free Cake Division Protocol" (*American Mathematical Monthly,* Vol. 102, January 1995) or the marvelous book *Cake Cutting Algorithms,* by Jack Robertson and William Webb (A. K. Peters, Natick, Mass., 1998).

One of the most interesting features of the theory of cake division is what Robertson and Webb call the "serendipity of disagreement." At first sight, it might seem that fair division is simplest when everybody is in agreement about what each bit of the cake is worth—after all, there can then be no disputes about the value of a given share. Actually, the reverse is true: as soon as participants disagree about values, it becomes easier to keep them all happy.

Suppose, for example, that Tom and Dick are using the cut-and-choose algorithm. Tom cuts the cake into two pieces, which he views as having equal value, one half each. If Dick agrees with those valuations, nothing more can be done. But suppose that Dick values the two pieces at $3/5$ and $2/5$. Then he might, for some altruistic reason, decide to give Tom $1/12$ of what he considers to be the larger piece (which he values at $1/20$ of the whole cake). He still has $3/5 - 1/20 = 11/20$ of the cake, according to his valuation. One way to do this is for Dick to divide the larger piece, in his estimation, into 12 parts that he considers to be of equal value. Then he offers Tom the choice of just one of them.

Whichever one Tom chooses, Dick still thinks he is left with $11/20$. Tom, on the other hand, is faced with 12 choices, and he values their total as one half. Therefore, at least one of them is worth $1/24$ in his estimation. By choosing that one, he ends up with what he considers to be at least $13/24$ of the cake. So now both Tom and Dick are satisfied that they have more than a fair share.

The intuition here is not that disagreement about values must lead to disagreement about what constitutes fair division. That might happen if a third party divided the cake and then insisted that Tom and Dick accept one of those predetermined shares.

Another instance of the same principle arises in the problem of dividing beachfront property. Suppose that a straight road runs east to west past a lake and that the land between the road and the lake is to be divided by north-south property lines. The problem is to divide the property among *n* people so that each gets a connected plot of land that he or she considers to be at least $1/n$ of the total value.

The solution is disarmingly simple. Make an aerial photograph of the property and ask each participant to draw north-south lines on it so that in his or her estimation the land is divided into *n* plots of equal value [*see illustration at top left on opposite page*]. If all draw their lines in the same places, then any allocation will satisfy them all. If there is any disagreement about where the lines go, however, it is possible to satisfy all of them that they have a fair share and to have some of the property left over. The illustration at the top right shows a typical case in which Tom, Dick and Harry have carried out such a procedure. Clearly, we can let Tom have his first plot, Dick his second and Harry his third—with some bits left over.

The large illustration shows a more complicated example in which Tom, Dick, Harry, Marcia and Becky each seek $1/5$ of the land. In 1969 Hugo Steinhaus proved that the same thing happens for any choices of dividing lines where there is the slightest disagreement. A proof using the principle of mathematical induction can be found in Robertson and Webb's book.

Sadly, not all tasks can be divided fairly, at least not with reasonable restrictions. Take washing up. If each person must wash and/or dry a complete dish, then in extreme cases no fair allocation is possible. Imagine two participants with one huge dish and one small one. Both will want to deal with the small one and won't accept the huge one. So even in a perfect world in which all disputes are settled by negotiation, some disagreements seem unavoidable. SA

FEEDBACK

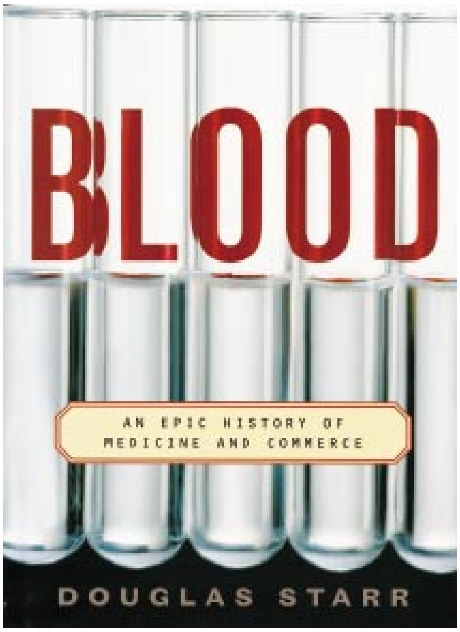
In my April 1998 column, "Repealing the Law of Averages," I mentioned that in a two-dimensional random walk on a square grid, the probability of eventually returning to the origin is 1 but that on a three-dimensional cubic grid the probability is less than 1, roughly 0.35. My reference for this fact was the book *An Introduction to Probability Theory and Its Applications*, Vol. 1, by William Feller. Several readers pointed out that the figure given in Feller's book is not quite correct. David Kilbridge of San Francisco says that in 1939 English mathematician George N. Watson gave the value as $1 - 1/[3(18 + 12\sqrt{2} - 10\sqrt{3} - 7\sqrt{6})(K(2\sqrt{3} + \sqrt{6} - 2\sqrt{2} - 3))]^2$, where $K(z)$ is $2/\pi$ times the complete elliptic integral of the first kind with modulus z^2. (If you don't know what that is, you probably don't want to! For the record, elliptic functions are a grand classical generalization of trigonometric functions like sine and cosine, which were very much in vogue a century ago and are still of interest in a number of contexts. But they are seldom studied in today's undergraduate math courses.) The numerical value is approximately 0.340537329551.

Kilbridge also tells me that the answer to my "dice eventually even out" problem is approximately 0.022. For dice with 2, 3, 4 or 5 sides, the analogous probability is approximately 1, 1, 0.222 and 0.066, respectively. Yuichi Tanaka, an editor for our Japanese edition, used a computer to work out the probability of eventually returning to the origin on a four-dimensional hypercubic grid. After three days, his program printed out the approximate value 0.193201673. (Is there a formula like Watson's? Any elliptic function experts out there?) *—I.S.*

LIFE AND DEATH IS IN THE BLOOD

Review by Claire Panosian

Blood: An Epic History of Medicine and Commerce
BY DOUGLAS STARR
Alfred A. Knopf, New York, 1998 ($27.50)



Ask any medical student. The first time you watch those maroon globules of life-giving sap drip solemnly from plastic pouch through translucent IV tubing and merge with the tide of native cells flowing through your anemic patient's depleted veins, it is a holy moment. "For the life of the flesh is in the blood," writes the Old Testament author of Leviticus. To a physician, infusing whole blood or one of its components into an ill or injured person is an act of incredible power . . . and seduction.

The power is undeniable, even to the medically naive. "Type and cross, stat!" is the cry that thunders through trauma centers, ERs and operating rooms throughout our land, whether they be real or the mock variety depicted weekly on network television.

Soon, if all goes well, an orderly rushes in, identity tags are hastily checked and double-checked to assure the compatibility of donor and recipient, hemoglobin-rich goo begins to course through a large-bore catheter or needle, and, thanks to modern blood banking and its technical advances, a human life that might have ebbed away just a few decades ago is revived.

Or consider the hemophiliac whose Y chromosome dictates a poignant fate: forever lacking a crucial glycoprotein needed to knit together a sturdy blood clot. Before the availability of blood-derived clotting factors, an innocent bump of the knee could spell days of joint ooze followed by crippling arthritis; a tonsillectomy or appendectomy, almost certain death from hemorrhage. Then, building on blood-separation techniques developed in the 1950s and 1960s, a new wave of remedies hit pharmacy shelves.

The crowning achievement was factor VIII concentrate, produced by pooling hundreds—and later thousands—of units of plasma. In the 1970s conservative doctors employed factor VIII treatment on demand, meaning patients would self-inject when bleeding began; a few aggressive advocates even prescribed prophylactic use. The end result? Suddenly, "bleeder boys" previously confined to overstuffed sofas and storybooks were trying out for sports, set free from their genetic destiny by a small vial of crystalline powder. That's power.

But now, you think, how is blood seductive? First sweep your mind of fanged romancers from the Carpathian Mountains with their lustful fantasies and blood-smeared lips. To anyone who has seen blood shed in large quantity, hemorrhage is not sexual, it is merely horrifying. Yet from medicine's earliest record, the possibility of in vivo manipulation of blood has entranced certain practitioners with a fascination verging on passion. Witness the 2,500-year tradition of bloodletting that killed George Washington as recently as 1799 or, in 17th-century Europe, experiments that attempted to reverse violent mental disorders with the infusion of "gentle" blood from calves or lambs. Could it be that, at a time when metaphysics outweighed knowledge, blood's symbolically redemptive power bestowed on physicians an addictive *goût* of divine authority? One wonders.

Most recently, in the latter half of the 20th century, blood's mystique has lured not only spiritual and scientific seekers but also commercial bloodhounds with a nose for money. This new generation of blood disciples includes global traders sometimes guilty of greed and human exploitation, not to mention, in the eyes of certain jurists and medical watchdogs, gross neglect of the sacred Hippocratic tenet: *primum non nocere* ("first do no harm"). For just as seaborne exploration in the 15th century launched a worldwide pandemic of syphilis, international trafficking in blood products has ironically and tragically coincided with the global spread of the blood-borne viruses HIV, hepatitis B and hepatitis C. Of course, there is at least one major difference between our era and that of Columbus. Today, as a worldwide medical community, we have tools to screen biological cargo. But do we use them? Sad to say, the answer is not always. Although the current HIV risk from a single transfusion in the U.S. hovers between one in 100,000 and one in 500,000, elsewhere the story is different. In countries of Africa and the Indian subcontinent, for example, experts speculate that transfusion of single-donor or multidonor blood products is still responsible for 10 percent or more of all HIV infections.

Heroes and Antiheroes

The preceding facts and footnotes in the annals of blood, including the ongoing specter of overseas blood contamination, are richly represented in Douglas Starr's new book, *Blood: An Epic History of Medicine and Commerce*. But were it not for some appalling scandals—such as the 1992 conviction and imprisonment of three leading French doctors for knowingly distributing HIV-contaminated factor VIII to their country's hemophiliacs, and the 1977 exposé of ruthless overbleeding of malnourished Nicaraguans to acquire plasma for drug companies in the industrial world (a chilling aside: newspaperman Pedro Joaquín Chamorro was assassinated two months after one of his editorials about "red gold" hit newsstands)—I fear that veteran science writer and journalism professor Starr might have had a tough sell of his work. As history would have it, however, and as Starr so ably reports, the past two decades have witnessed an unprecedented entwining of the high-stakes global plasma business, the worldwide HIV conflagration and a First World public increasingly demanding disclosure, debate and legal retribution for transfusion-related injuries.

While making full use of this material, at the same time Starr is committed to chronicling transfusion medicine's rich back-story. The result is a fair and thorough book, balancing altruism and corruption, prescience and blindness, and respecting, along with newsbreaking headlines, key historical milestones such as the discovery of blood groups, the development of anticoagulants and the design of apparatus such as lancets, cannulas and storage containers.

Strange heroes and antiheroes also step from the pages of Starr's saga. Among them: Alexis Carrel, the French expatriate who, in New York City in 1908, performed the first modern transfusion by suturing a baby's leg vein to an artery in her father's arm. Norman Bethune, the passionate Canadian revolutionary who hand-carried blood to the front lines during the Spanish Civil War. U.S. Army Surgeon General Norman T. Kirk, an orthopedist who tragically opposed the shipment of whole blood to the battlefield at Normandy. Ryoichi Naito, the Japanese bacteriologist who spent World War II secretly designing "plague bombs," then dedicated himself to life-restoring transfusion medicine and in March 1951 opened the for-profit Japan Blood Bank (which soon dwarfed the activities of the local Red Cross, itself plagued by the unwillingness of ordinary Japanese citizens to embrace blood donation with civic pride). And well-meaning but warring leaders in the 50-year history of American blood banking.

Of the many known and unsung heroes who prophesied the perils of blood and multidonor "plasma mills," Starr highlights one Cassandra who merits special mention. In the 1950s J. Garrott Allen, a surgeon at Stanford University, was perhaps the first to foresee the risk of covert viruses in blood and plasma from paid donors. In 1966, publishing years of meticulous research, he confirmed that hepatitis was 10 times more common in recipients of blood from paid donors than in patients who had received blood from volunteers. For years, Allen then pleaded his findings to federal and state health agencies, asking that donation centers disallow all incarcerated and skid-row donors and that doctors limit their use of blood products by "giving one transfusion instead of two, two instead of three." Unfortunately, his message would not translate into action for many years. As Starr points out, in France in particular, prison blood collections continued well into the 1980s, despite indisputable evidence of their risk.

In vignettes such as this, Starr shows how hard it is to separate the medical from the human morality play that has recently surrounded life's most precious fluid. Recalling my own youthful awe at the magic of transfusion, I now believe that clear-eyed critiques such as *Blood: An Epic History of Medicine and Commerce* should be included in all first- and second-year medical curricula. And to my colleagues, I would add this postscript: let us never forget the human cost of forgoing vigilance against commercial and bureaucratic influences on our profession. Modern medicine is a high-stakes game, and we would be derelict indeed to trust that outside agencies will invariably act in our patients' best interests.

CLAIRE PANOSIAN is a medical journalist and a professor of medicine and infectious diseases at U.C.L.A. School of Medicine. Her reviews and essays have recently been published in Discover, *the* Los Angeles Times *and the* Chicago Tribune.

THE EDITORS RECOMMEND

ANNALS OF THE FORMER WORLD. John McPhee. Farrar, Straus and Giroux, New York, 1998 ($35).

"It's a real schlemazel," geologist Anita Harris said to McPhee as they examined geologic formations at a road cut along Interstate 80 near the Delaware Water Gap. "Not by accident is geology called geology. It's named for Gaea, the daughter of Chaos." The rocks are often chaotic, but the study of them is not in McPhee's pellucid presentation. His meaty book, adorned with 25 stunning landform maps, is the result of a 20-year project in which he set himself the goal of portraying geology and its practitioners in a way that would "arrest the attention of other people while achieving acceptability in the geologic community." He started with the intention of setting forth "a sort of cross section of North America at about the fortieth parallel" but wound up casting a much wider net. A measure of the scope of his tale is provided by the structure of Book 2: "In Suspect Terrain," which begins with a profile of Harris, examines the Delaware Water Gap as a fragment of the Appalachians, discusses the Appalachians and plate tectonics and presents the theory of continental glaciation. Book 2 and the four others fill out an absorbing picture of the former world—the North America of past geologic eras back to the beginning of the Mesozoic some 245 million years ago.


RUTH MARTEN

IF A LION COULD TALK: ANIMAL INTELLIGENCE AND THE EVOLUTION OF CONSCIOUSNESS. Stephen Budiansky. Free Press, New York, 1998 ($25).

"How animals think, and what they think about, are ancient questions that have proved both irresistible and maddeningly elusive," Budiansky writes. They are the questions he explores, cautioning repeatedly against the strong human tendency to accept tales of animals that seem to display human motives, understanding, reason and intentions. He gives many examples of what horses, dogs, cats, pigeons, chickens, apes and other animals do in various situations and says: “The study of animal cognition might be defined as the science of How Do We Know for Sure, for learning and evolution do a sensational job of generating intelligence without conscious intention or insight.” The intelligence displayed by every species “is wonderful enough in itself; it is folly and anthropomorphism of the worst kind to insist that to be truly wonderful it must be the same as ours.”


BRITISH MUSEUM

The Mummy in Ancient Egypt: Equipping the Dead for Eternity. Salima Ikram and Aidan Dodson. Thames and Hudson, London, 1998 ($45).

Ikram, an Egyptologist, and Dodson, an archaeologist specializing in Egyptian funerary practices and aspects of dynastic history, combine to produce a comprehensive account of how Egyptian mummies were made, wrapped, adorned and sheltered for an afterlife. Their story covers 33 dynasties and a span of some 3,000 years, beginning about 3050 B.C. The 485 illustrations show tombs, mummies, the mummifiers’ methods of embalming and many funerary accessories. “Everything,” the authors say, “depended on the Egyptian belief in eternal life, and the need to provide for it.” The account also traces the study of mummies from the often clumsy procedures of the 19th century to the x-rays, CT scans, endoscopy, scanning electron microscopy and DNA-testing techniques available today.

Calendar: Humanity’s Epic Struggle to Determine a True and Accurate Year. David Ewing Duncan. Avon Books, New York, 1998 ($23).

The days march along, one by one, oblivious of the human effort to impose order on them—chopping them into seconds, minutes and hours and then grouping them into weeks, months, years, decades, centuries and millennia. The effort stretches from the markings someone made on an eagle bone about 13,000 years ago to the oscillations of the cesium clock. And still we have to tinker to make the calendar fit the tropical or solar year of 365.242199 days, or 365 days, five hours, 48 minutes and 46 seconds. Duncan traces the story engagingly, telling of calendars based on the seasons of the Nile and the cycles of the moon, of the Julian and Gregorian reforms and of the consternation that ensued in England when the nation finally switched from the Julian to the Gregorian calendar in 1752, necessitating a correction that took 11 days from the “Old Style” calendar and made many people believe their lives had been shortened. As *The Ladies Diary: or, Woman’s Almanack* reminded its readers: “1752 September hath only XIX Days in this Year” because “The Account of Time has each Year run a-head of Time by the Sun.” It still does, by about 25.96 seconds a year, which means that tinkering will continue.

Dinosaur Impressions: Postcards from a Paleontologist. Philippe Taquet. Cambridge University Press, New York, 1998 ($24.95).

Taquet is a French paleontologist with a gift for making his readers feel what it is like to be painstakingly extracting dinosaur bones from the spot where the creature died many millions of years ago. At one such site, in the desert of Ténéré in Niger, collecting a complete dinosaur skeleton entailed digging out 274 bones, each broken into several pieces—some 600 fragments in all. Then Taquet had to package them and oversee their transportation to his laboratory at the National Museum of Natural History in Paris, where he reconstructed the skeleton. “So the bone hunter has to be a successful geologist, rambler, and naturalist,” he writes. “And then he has to transform himself into a roadworker, a sculptor, a plasterer, and a trucker.” Taquet also considers the four main hypotheses on what carried off the dinosaurs—an asteroid impact, an episode of severe volcanism, competition among species, and a combination of lowered sea levels and reduced temperatures—and concludes that “there is no simple answer to the riddle.” The book originally appeared in French; Kevin Padian, curator of the Museum of Paleontology at the University of California at Berkeley, has turned out a smooth translation.

Phantoms in the Brain. V. S. Ramachandran and Sandra Blakeslee. William Morrow and Co., New York, 1998 ($27).

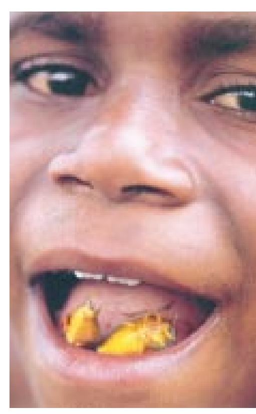
People who suffer from certain illusions, such as sensations from a missing limb or a conviction that look-alike impostors have replaced one’s parents, are often treated as psychiatric cases or neurological curiosities. Ramachandran, a brain researcher, sees them instead as “our guides into the inner workings of the human brain.” He tells the stories of several such people and what their illusions suggest about how the brain works. Along the way the reader learns of Charles Bonnet syndrome, the vivid visual hallucinations experienced by some blind people (James Thurber probably among them); hemineglect, a condition that often follows a stroke in the right brain and causes the patient to be profoundly indifferent to objects and events on her left side; and pseudocyesis, or false pregnancy. Ramachandran thinks the line of research he describes may reach an epochal goal—the answer to “a question that has been steeped in mysticism and metaphysics for millennia: What is the nature of the self?”

The Invisible Computer: Why Good Products Can Fail, the Personal Computer Is So Complex, and Information Appliances Are the Solution. Donald A. Norman. MIT Press, Cambridge, Mass., 1998 ($25).

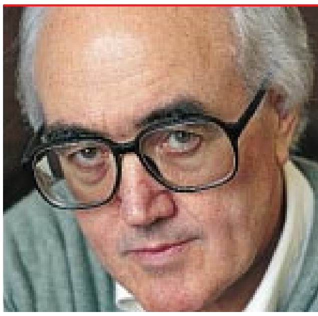
Norman’s thesis is that a number of modern products, most notably personal computers, come with a technological overload that makes them hard to use. He pictures a “vast chasm” that separates the early users of a device, who will put up with complex technology in order to reap the benefits offered, from later users, who want reliability and simplicity. “Alas, the aging teenagers who rule the computer companies of the world are still stuck on the youthful side of the chasm,” he says. Norman maintains that personal computers and other technologically daunting devices should be as easy to use as such “information appliances” as watches, telephones, digital cameras, and compact-disc and tape players.

Man Eating Bugs: The Art and Science of Eating Insects. Peter Menzel and Faith D’Aluisio. Ten Speed Press, Berkeley, Calif., 1998 ($19.95).

Entomophagy, the eating of insects, is not for every palate, but a surprising number of people do it. Menzel and D’Aluisio, husband and wife, have visited insect eaters in 13 countries, sampling the menu at each stop. “Our view of the culinary potential of invertebrates broadened as we ate raw scorpion in China, roasted grubs in Australia, stir-fried dragonflies in Indonesia, tarantulas on a stick in Cambodia, and live termites in Botswana,” they write. “Perhaps the most memorable meal was *Theraposa leblondi,* a tarantula big enough to hunt birds, which we ate with Yanomami Indians in the Venezuelan rain forest.” Adventurous readers will find recipes for such delicacies as Witchetty Grub Dip and Stink Bug Pâté. Menzel, a photographer, enlivens the book with many of his pictures. D’Aluisio, identified as “a reluctant bug eater,” nonetheless learned that she could do it. The experience led her to conclude that “the shelves of the [American] supermarket carry only a narrow slice of what the world has to offer.”


PETER MENZEL

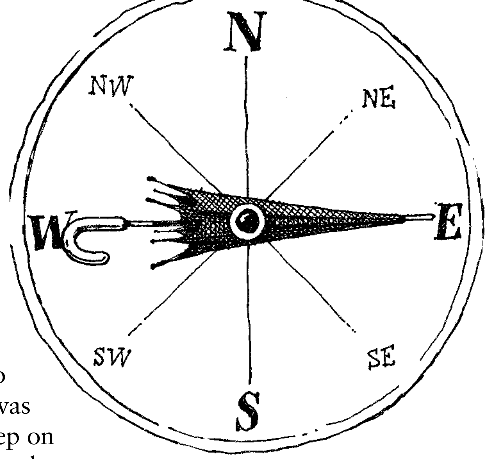
COMMENTARY



CONNECTIONS

by James Burke

And Now the Weather



DUSAN PETRICIC

It was ironic that as I recently dashed into the Paris church now known as the Panthéon, out of yet another rainstorm sweeping in from the Atlantic, I was filming a sequence about the guy who told us why the rain always comes from that direction. Because in the Panthéon hangs Léon Foucault's great experiment of 1851, in which he dangled a 62-pound cannonball on 220 feet of piano wire, pulled it to one side with a cord, then burned the cord to release the ball without influencing its movement. The pattern followed by the swing of the ball over the next few hours was the first physical proof that Copernicus had been right. As the pendulum swung in inertial space, a stylus attached under the ball traced a line that shifted as the earth turned beneath it. Point being, this demonstration would become the basis for the meteorological thinking of Buys Ballot and others about how global weather was driven by the west-east rotation of the earth. It made Foucault instantly famous.

Foucault's name went up in lights for other reasons, however, thanks to his improved regulator, which kept arc-light carbon rods at the right distance apart as they burned. And so light became efficient enough to use in large public places like theaters, where, in 1892, people such as chemist Henri Moissan went for an evening off. Back in his lab Moissan was using the arc for a very different purpose: to power his electric arc furnace, in which the carbon rods burned so hot they nearly made him artificial diamonds (see the November 1998 column). Moissan was *le noodleur extraordinaire* who got the Nobel in 1906, isolated fluorine (without killing himself) and wrote more than 300 scientific papers. In the winter of 1897 he also provided a young Polish physicist, living in Paris, with some uranium powder, and she began to investigate its mysterious ability to "electrify" the air around it. This was Marie Skłodowska Curie's first step on the way to her discovery of what the "electrification" might turn out to be. Later she called it "radioactivity."

Marie had recently married into the Curie family, where there was a strong tradition of egalitarianism, science and emancipation. Much of this had come from the grandfather, Paul Curie, who early in the century had been a follower of New Christianity, a sect started by a weirdo named Henri de Saint-Simon, who at one point, while temporarily destitute and starving, had shot himself in the head six times and survived. De Saint-Simon's idea was to update religion so that it was closer to the modern world. His prospectus talked about spiritual power belonging to men with practical knowledge and in general glorified work and the capitalist ethic. No surprise that by 1830 engineers, financiers and businessmen were dressing up in New Christian costume and talking about free love (this particular aspect of the new religion would ultimately lead to its downfall).

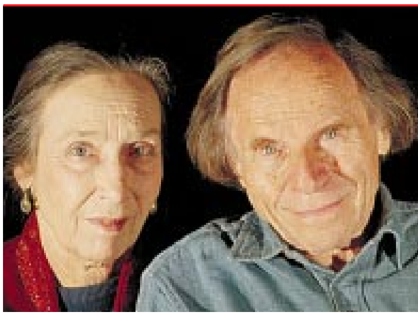
The religious sect leader shot himself in the head six times and survived.

One of these New Age techies was somebody who made even de Saint-Simon look normal: the corpulent banker Barthélemy-Prosper Enfantin, who took over after de Saint-Simon's death and became New Christianity's messiah. Enfantin modestly saw himself as half of Jesus Christ. The other half would be a soul mate yet to be identified. At one point, traveling to Egypt in search of the Bride (he never found her), he became excited about one of the bees de Saint-Simon had had in his bonnet: building canals. Enfantin's 1847 survey for a Suez Canal (a vision he had about joining East and West) later gave him grounds for a dustup with the influential Ferdinand, vicomte de Lesseps, when Lesseps actually built the proposed canal.

Ferdinand was a diplomat whose father (as French consul for Napoleon) had taken an illiterate soldier and made him Khedive of Egypt. So Ferdinand could do no wrong in Cairo. When he suggested the Suez Canal, it was just what the Khedive was looking for to give his country a reputation better than dreadful. The canal took Ferdinand 10 years and 25,000 laborers, and the official opening in November 1869 was one of history's greater bashes. Glitterati wasn't the word. Dinner for 8,000, every one of whom cared about protocol and getting the seat he or she deserved. One of the few who didn't have to worry about these niceties was the wife of Napoleon III of France, Empress Eugénie, who also happened to be Ferdinand's cousin.

Avid readers of this column will be aware that as a small Spanish child, Eugénie had met and become great pals with the French writer Prosper Mérimée, who had charmed her rigid. Well, you know the old adage: be nice to people on the way up. It paid off. Once Eugénie had tied the imperial knot, Mérimée was in like Flynn and rapidly became *très* chic among the chattering classes of Paris. And London. Where he got to know Antonio Panizzi, top bookworm of the British Museum and mega-schmoozer. If there was anybody worth knowing in England, Panizzi was on first names. Panizzi himself had

Continued on page 117



WONDERS
by Philip and Phylis Morrison

Mound and Crater

DUSAN PETRICIC

Buildings rest on terra firma, on its soil or its rocks. Even the jets aloft pass their load onto the earth's surface via the downward push they give to all that passing air. Travelers have long admired ancient earth mounds, like the 1,200-year-old tumulus of Prince Krak, founder of Kraków. A great many dwellings around the globe are still earthworks, built up of rammed dirt and clay, adobe blocks, often with additives. Roads are built so snugly to the earth that it is hard to see them as distinct works; our grand highways are elaborate composites.

But the simple use of earth as a major material has not yet vanished from modern structures. Even today the largest dams are not those sculptural concrete walls but low embankments, piled up for miles long out of soil and rubble. The largest U.S. embankment dam, at Fort Peck on the Missouri River in Montana, is second in the world for size. It stores some five cubic miles of water, contained by an earth-fill dam whose volume is only a few parts per 1,000 of the stored water volume.

The serpent's earthen body stretches in seven folds for a full quarter of a mile.

Two unusual American "earthworks," one artificial, one natural, are in the foreground of memory for us. We were led to the first of these at the Peabody Museum at Harvard University one afternoon years ago. In an odd corner of that large archaeological collection, there was an elderly plaster relief map of an ancient earthen mound shaped like an enormous snake, winding down one hill and up the next. Harvard archaeologist Frederic W. Putnam led the scholars who in the 1880s finally brought reason to the interpretation of earthworks like this Great Serpent Mound of Ohio. The growing evidence was plain that these imposing structures were the work of the same First Americans whose descendants lived nearby. A crowd of authors—some poets and dreamers, some rascals—had sought the Mound Builders among the Lost Ten Tribes or the Vikings or almost anyone but the local people, who by then were displaced or even destroyed.

Putnam came to Serpent Mound, about 40 miles southwest of Chillicothe, Ohio, and recognized the site as the largest serpent effigy in the world. The farmer who owned the land had seen its importance and never plowed it, aware that it needed protection. Putnam persuaded some serious Boston ladies of means to purchase the land (for $6,000) and give it to Harvard (the map we saw followed). He restored the mound in 1887 to fit the most reliable old survey, fenced it and drained it, not exactly as it was but in reasonable approximation. In 1900 Harvard deeded it to the Ohio Historical Society, and it is now a welcoming state memorial with an observation tower that overlooks the creature.

The body of the serpent, a long, rounded pile from two to five feet high and nearly 20 feet wide, stretches in seven folds for a full quarter of a mile, from coiled tail downhill and up again to the head, where its mouth holds an obscure oval form. One May we visited it, after a prologue guided by a Columbus archaeologist. The serpent was a place of engaging charm, its mantle of spring-green grass ornamented by wild violets. The coils were just right for a quiet pair of picnickers. Try it yourself this spring or summer and don't fail to visit the Ohio Historical Center in Columbus, where the artifacts of earlier Ohioan cultures are beautifully presented, in particular some mica cutouts that adorned Hopewell times, made by Ohio artists before the fall of Imperial Rome.

It is not hard to estimate in context roughly how the serpent might have been made. Local soil was carried by the people of a small chieftaincy. The effigy required about 300,000 modest backloads, each man and woman with a basket. If we imagine that they worked during the few pleasant weeks after the harvest every year, it would require only five or 10 seasons for a settlement of a few hundred adults to create their serpent, a monument to the power of ideas persuasively shared by some leader. No grave goods or relics of any kind have yet been found within the serpent itself. Indeed, it has never been closely searched; it was dated unreliably only through a few artifacts in a small mound nearby. But in 1996 it was sampled for bits of charcoal, and the first radiocarbon dates were published, setting the serpent's construction in about A.D. 1100. Cahokia—the largest mound complex in the country, near East St. Louis, Ill.—bears a similar date and is of a successor culture.

A dazzling, incandescent bolt from the sky caused a remarkable underground explosion on the high plateau near Winslow, Ariz., probably some 300 centuries before anyone was there to witness. The event, with the energy yield of 20 megatons of TNT, left a spectacular hole: Meteor Crater, an instant natural excavation. What you see is a thrilling near-circular cavity about three quarters of a mile across, deep enough to shelter the Washington Monument below grade, all within a rim wall of turned-up strata and rock debris.

In this region of many volcanic craters, its origin was obscured until the turn of the century. The sheer absence of any of the lava and ash that abound in volcanic features signaled the distinct nature of this rimmed crater. A mining engineer, Daniel M. Barringer, first understood what the astronomers had begun to say. Could this be an impact crater, made by an errant solid nickel-iron meteor as big as the White House, now surely buried deep within? That small asteroid would have to have weighed a quarter of a million tons to carry the kinetic energy released on impact. The crater floor still holds the relics of Barringer's long industrial effort to reach that iron treasure; his family still owns the site.

The orbiting iron body is not there; it had indeed entered but left at once! In a mere fraction of a second the energy-rich metallic object and the rocks it penetrated all vaporized; the enormous pressure of that expanding hot gas, part iron asteroid and part Arizona rock, excavated the rock shattered by the growing shock wave. Up it all went, most to come down somewhere nearby. By 1948 Harvey H. Nininger had definitely found the telltale nickel-iron: tiny, gleaming, glassy globules were everywhere in the topsoil, amounting to thousands of tons of droplets condensed out of the vapor mix, a metallic fog. In past, wetter climates this crater was often a lake of water, but now it is a lucid "lake" of high desert air, the youngest, freshest, most accessible big impact crater on the earth, among the 100 or so known. We first saw it beautifully frosted by a light winter snow. SA

Connections, continued from page 115

hightailed it to England at one point in 1823, when the Italian secret service got too close for comfort after he'd been involved with some shady folk known as the Carbonari, who wanted all kinds of crazy stuff for Italy, like freedom of speech.

Panizzi's libertarian views (and accent) went over very big in Liverpool, his first port of call, where he had been sent with an introduction to William Roscoe. Apart from being an extreme Italophile (me, too—Italy is, in my opinion, a disease for which, happily, there is no cure) and biographer of Lorenzo de' Medici, Roscoe was at various times banker, botanist, antislavery activist, Member of Parliament for Liverpool, publisher and collector of rare books (and runaway liberals). In 1806 he also wrote a nursery classic: *The Butterfly's Ball, and the Grasshopper's Feast,* which immediately took the fancy of the king and queen and was then published by John Harris in his first, boffo-success series of children's books.

Harris was successor to John Newbery, among the first publishers in England to print illustrated books specially written for children at his London publishing house in St. Paul's Churchyard (where he also did people like Goldsmith and Johnson). In 1780 a list of Newbery copyrights included *Tales of Mother Goose,* translated from the French. It had been written by Charles Perrault, who also had the dubious distinction of keeping the books during the construction of Versailles. Charles was one of three brothers: Claude designed a bit of the Louvre and the Paris Observatory, and Pierre (a tax collector caught with his hand in the till) founded the science of hydrology and worked with a guy who wrote Pierre's stuff up as his own. Edmé Mariotte managed to put up a few backs with this trick; Huygens accused him of plagiarism.

In the 1670s one of the things Mariotte did that was almost undoubtedly his own was to organize a chain of data stations across Europe from which he was able to put together a report on global and European winds. And for the first time theorize about what Foucault finally helped to prove, which I mentioned at the beginning of this column—that the weather moves west to east because of the way the earth turns. SA

WORKING KNOWLEDGE

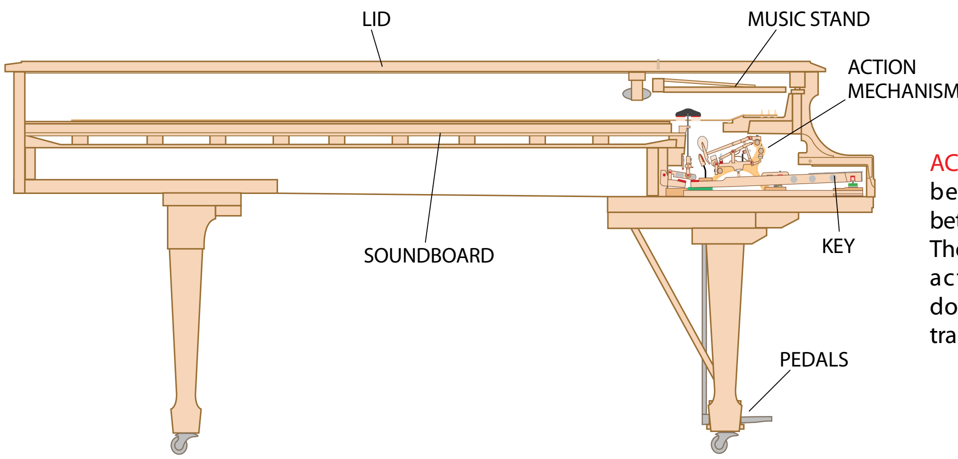
GRAND PIANO

by Michael Mohr
Director of Manufacturing, Steinway & Sons

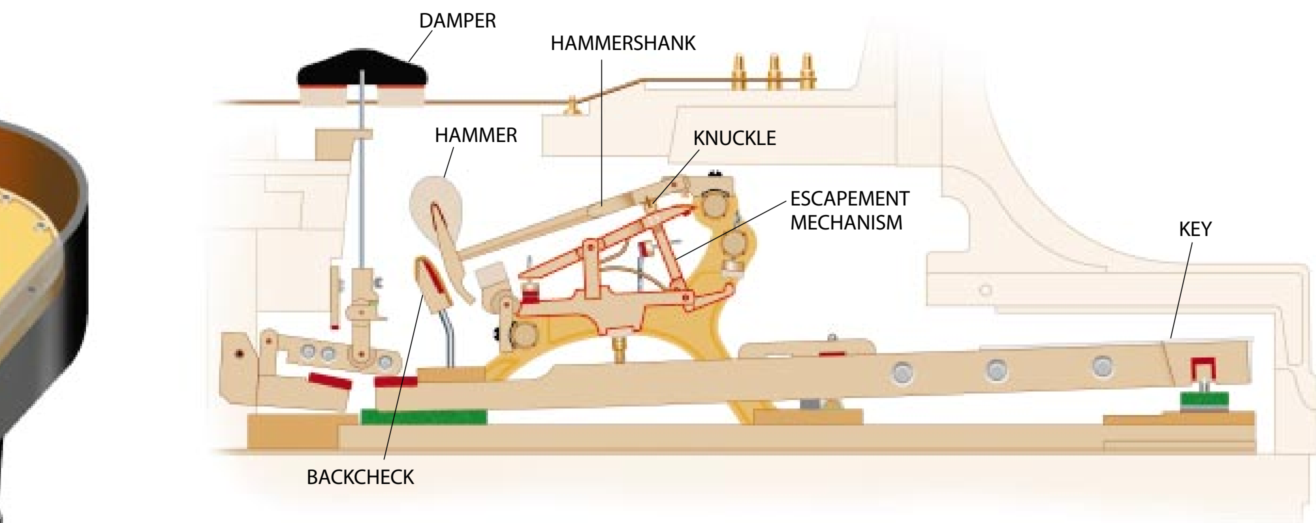
The piano dates back to the early 1700s, when a Florentine harpsichord maker named Bartolomeo Cristofori, seeking to improve the expressiveness of the harpsichord, devised the "escapement action"—a mechanism that, in modified form, lies at the heart of all modern pianos. Unlike the harpsichord, which features a plucking mechanism attached directly to a key, the piano has a mechanical interface between each key and a corresponding felt-tipped hammer. This complex system of levers, screws, springs and bearings allows the pianist to play quickly and to produce a wide range of dynamics and nuance.

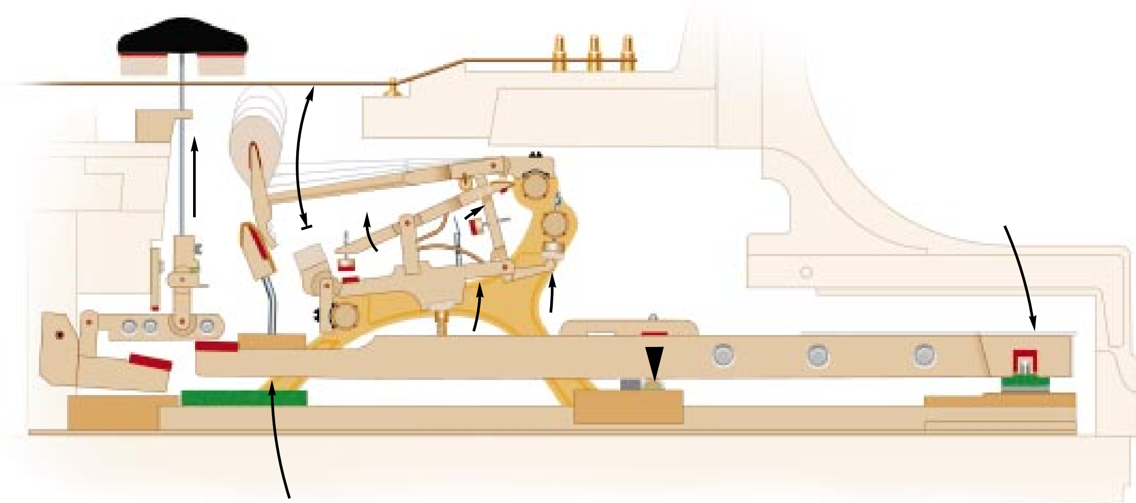
The lever system magnifies movement so that while the key travels less than half an inch from its rest position, the hammer travels one and three quarters inches to the string above. Just before the hammer strikes the string, the action mechanism must disengage, or "escape," from its contact with the key. The hammer then strikes the string on its own momentum and falls back instantaneously so the string can vibrate freely. Without escapement, the hammer would jam against the string, stifling vibration. Modern pianos have a double escapement action that allows the hammer to be reactivated even when the key has not yet been fully released; thus, a single note can be played in fast repetition.

Cristofori's instrument had a fairly wide dynamic range, but it was not so powerful as modern pianos. Over the years, the structure of the piano was strengthened to accommodate stronger strings for a more robust sound, and additional pedals (Cristofori's instruments featured only the "soft" pedal, if any) were introduced. Although details of the modern grand piano action differ from manufacturer to manufacturer, the basic concept is the same. In an upright piano, gravity alone does not return the action to the rest position. To compensate, special springs are incorporated into an upright's action, but the touch is still different from that of a grand piano.



ACTION MECHANISM IS HIDDEN behind the wood cabinet, between the keys and the strings. The pedals are connected to the action by a series of levers, dowels and springs called the trapwork (*not shown*).





ILLUSTRATIONS BY GEORGE RETSECK

DEPRESSING THE KEY pushes up on the escapement mechanism, which in turn pushes the hammershank upward. At two critical points in the hammer's upward motion, obstructions disengage elements of the escapement mechanism from contact with the hammershank's knuckle, so that during the last $^{1}/_{16}$ inch of its motion, the hammer is traveling entirely on its own momentum. Meanwhile the damper is lifted off the string, allowing the string to vibrate when struck. After the hammer strikes, it is caught and held by the backcheck. At the slightest release of the key, the backcheck will move away and free the hammer, and the components of the escapement mechanism will reset. The action is then ready for another strike from the key.