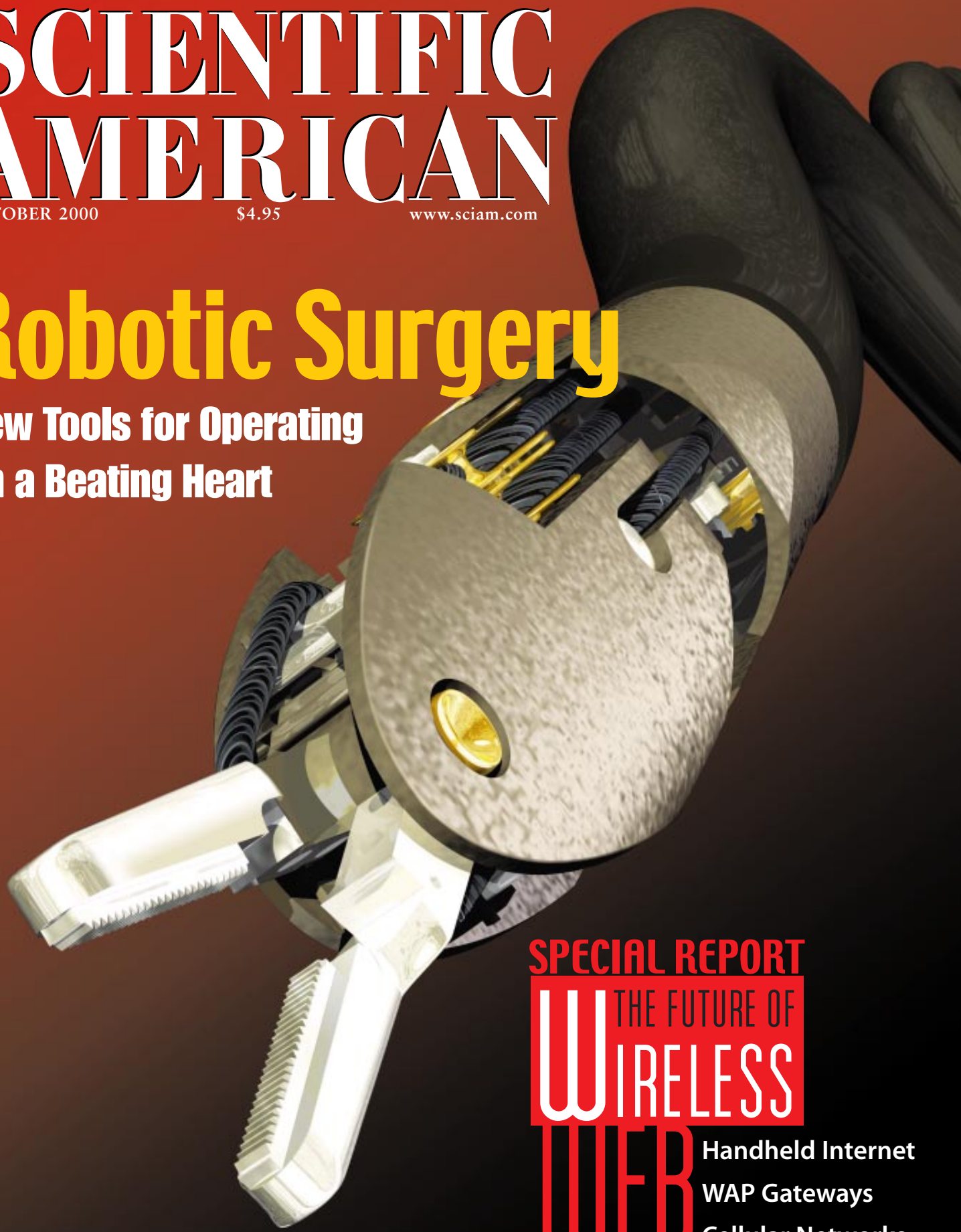
Memes: The Genetics of Culture • Making Smarter Decisions • The Lost City of Nabada

SCIENTIFIC AMERICAN

OCTOBER 2000 $4.95 www.sciam.com

Robotic Surgery

New Tools for Operating on a Beating Heart



SPECIAL REPORT

THE FUTURE OF WIRELESS WEB

Handheld Internet
WAP Gateways
Cellular Networks

SCIENTIFIC AMERICAN

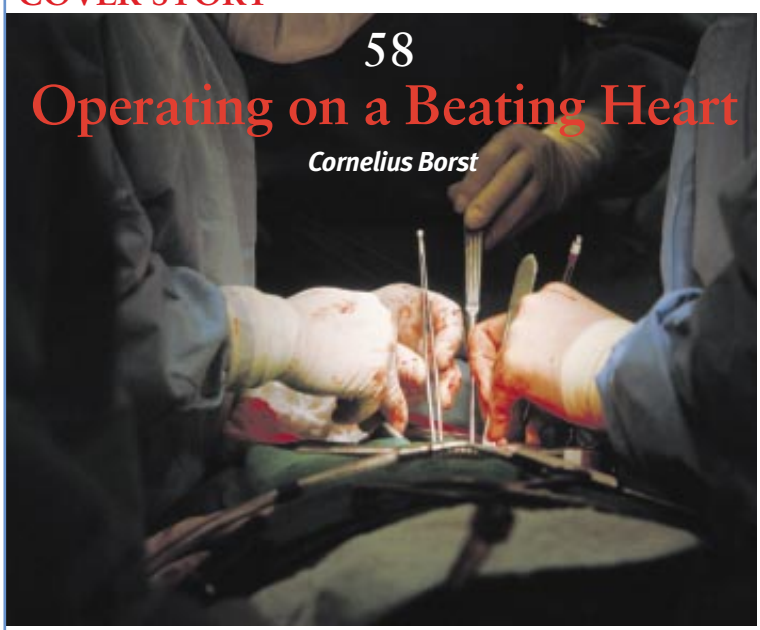
October 2000 Volume 283 www.sciam.com Number 4

Contents

SPECIAL INDUSTRY REPORT

The Wireless Web 38

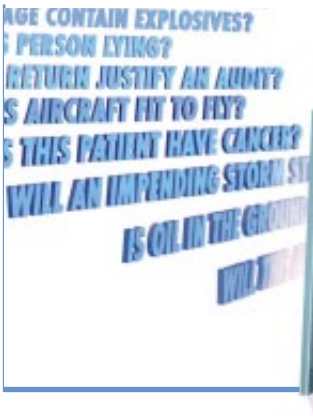
COVER STORY



Contents

SCIENTIFIC AMERICAN

October 2000 Volume 283 www.sciam.com Number 4



Better Decisions through Science 82

John A. Swets, Robyn M. Dawes and John Monahan

Every day, important and complex yes-or-no diagnostic decisions are made throughout medicine, industry and society. Statistical methods of making those choices could dramatically improve the outcomes.

About the Cover

Illustration by Slim Films.

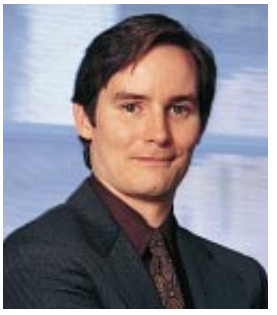
EDITOR_JOHN RENNIE

Smart Choices

Decisions, decisions. We all make them every day, and thank heaven mine are always right, but can you imagine the anxiety felt by those people with flawed judgment?

Take the stressful lives of diagnosticians in medicine and industry, whose choices tip the balance between life and death. In their perfect world, diagnoses would be easy because the evidence would unambiguously and without fail point to the true underlying condition. In their slightly less perfect world, knowledge of the occasional misdiagnoses would be tempered by certainty that they had caused a minimum of damage. But our world is the planet Earth, where the motto is "Not for the Squeamish."

And yet there is hope. As the authors of "Better Decisions through Science" convincingly argue, beginning on page 82, statistical aids can often improve diagnoses. Moreover, this mathematical approach—don't worry, it's fairly simple—works even with decisions that have traditionally been seen as qualitative and subjective, such as parole assessments of violent felons. I strongly recommend this article to politicians, managers, physicians, educators and anyone else routinely making tough choices; it will make you think.



A fusion of math and psychology can improve decision making.

This article inaugurates a series of collaborations between Scientific American and *Psychological Science in the Public Interest,* a new journal from the American Psychological Society (APS, www.psychologicalscience.org). Leadership in the APS recognized that the public's awareness of psychological research is poor. The best and most reliable findings are lost in the haze of headline-grabbing reports that often make conflicting or spurious claims. *PSPI* will therefore publish "white papers" summarizing the conclusions of a jury of experts that has weighed the published evidence on topics of national concern. Future issues may consider such matters as: Do smaller class sizes improve students' academic achievement? Is controlled drinking a safe alternative to abstinence for alcoholics? Can ginkgo and other herbal products enhance cognitive function?

To help disseminate these findings as widely as possible, Scientific American is working with the authors of the *PSPI* scholarly papers to publish versions aimed more at the general public. Our hope is that these articles will inform political and social discussions to good effect.

Speaking of smart decisions, voters in the Kansas primaries have rebuffed the anti-evolutionists seeking reelection to that state's board of education. Last year, you will recall, that panel rewrote the curriculum guidelines to eliminate requirements that evolution be taught to biology students—and it had the hubris to pretend that this raised the educational standards. Lunacy.

I'd make a quip about cryptocreationists being extinct except that, sadly, these Kansas specimens are far from the last of their species. Here's hoping nonetheless that the Sunflower State's response echoes elsewhere.

editors@sciam.com

SCIENTIFIC AMERICAN®
Established 1845

EDITOR IN CHIEF: John Rennie

MANAGING EDITOR: Michelle Press
ASSISTANT MANAGING EDITOR: Ricki L. Rusting
NEWS EDITOR: Philip M. Yam
SPECIAL PROJECTS EDITOR: Gary Stix
SENIOR WRITER: W. Wayt Gibbs
EDITORS: Mark Alpert, Graham P. Collins, Carol Ezzell, Steve Mirsky, George Musser, Sasha Nemecek, Sarah Simpson
CONTRIBUTING EDITORS: Mark Fischetti, Marguerite Holloway, Madhusree Mukerjee, Paul Wallich

ON-LINE EDITOR: Kristin Leutwyler
ASSOCIATE EDITOR, ON-LINE: Kate Wong

ART DIRECTOR: Edward Bell
SENIOR ASSOCIATE ART DIRECTOR: Jana Brenning
ASSISTANT ART DIRECTORS: Johnny Johnson, Heidi Noland, Mark Clemens
PHOTOGRAPHY EDITOR: Bridget Gerety
PRODUCTION EDITOR: Richard Hunt

COPY DIRECTOR: Maria-Christina Keller
COPY CHIEF: Molly K. Frances
COPY AND RESEARCH: Daniel C. Schlenoff, Myles McDonnell, Rina Bander, Sherri A. Liberman

EDITORIAL ADMINISTRATOR: Jacob Lasky
SENIOR SECRETARY: Maya Harty

ASSOCIATE PUBLISHER, PRODUCTION: William Sherman
MANUFACTURING MANAGER: Janet Cermak
ADVERTISING PRODUCTION MANAGER: Carl Cherebin
PREPRESS AND QUALITY MANAGER: Silvia Di Placido
PRINT PRODUCTION MANAGER: Georgina Franco
PRODUCTION MANAGER: Christina Hippeli
ASSISTANT PROJECT MANAGER: Norma Jones
CUSTOM PUBLISHING MANAGER: Madelyn Keyes

ASSOCIATE PUBLISHER/VICE PRESIDENT, CIRCULATION: Lorraine Leib Terlecki
CIRCULATION MANAGER: Katherine Robold
CIRCULATION PROMOTION MANAGER: Joanne Guralnick
FULFILLMENT AND DISTRIBUTION MANAGER: Rosa Davis

ASSOCIATE PUBLISHER, STRATEGIC PLANNING: Laura Salant
PROMOTION MANAGER: Diane Schube
RESEARCH MANAGER: Susan Spirakis
PROMOTION DESIGN MANAGER: Nancy Mongelli

SUBSCRIPTION INQUIRIES sacust@sciam.com
U.S. and Canada (800) 333-1199,
Outside North America (515) 247-7631

DIRECTOR, FINANCIAL PLANNING: Christian Kaiser
BUSINESS MANAGER: Marie Maher
MANAGER, ADVERTISING ACCOUNTING AND COORDINATION: Constance Holmes

DIRECTOR, ELECTRONIC PUBLISHING: Martin O. K. Paul
OPERATIONS MANAGER: Luanne Cavanaugh

DIRECTOR, ANCILLARY PRODUCTS: Diane McGarvey
PERMISSIONS MANAGER: Linda Hertz
MANAGER OF CUSTOM PUBLISHING: Jeremy A. Abbate

CHAIRMAN EMERITUS
John J. Hanley
CHAIRMAN
Rolf Grisebach
PRESIDENT AND CHIEF EXECUTIVE OFFICER
Gretchen G. Teichgraeber
VICE PRESIDENT AND MANAGING DIRECTOR, INTERNATIONAL
Charles McCullagh
VICE PRESIDENT
Frances Newburg

Scientific American, Inc.
415 Madison Avenue
New York, NY 10017-1111
PHONE: (212) 754-0550
FAX: (212) 755-1976
WEB SITE: www.sciam.com

ERICA LANSNER

EDITORS@SCIAM.COM

In "A Scourge of Small Arms," Jeffrey Boutwell and Michael T. Klare write that "insurgent leader Charles Taylor invaded [Liberia] with only 100 irregular soldiers armed primarily with AK-47 assault rifles."

That a few rebels or government-backed soldiers could take over a whole country or slaughter hundreds of thousands should be enough to shock us into taking action. But can we realistically expect to prevent future Charles Taylors from obtaining a scant 100 light weapons? Can we really prevent a government from obtaining and distributing weapons to its supporters?

It seems that a much more practical—and long-lasting—solution would be to ensure that civilian populations in threatened countries are permitted to defend themselves, their families and their homes from rebellion and insurrection, instead of being forced to remain helpless as they are lined up and executed. No way could a band of 100 hope to succeed in attacking a properly defended village, let alone an entire country.

JAMES TERPENING
via e-mail

I would suggest that there are tens of thousands of workshops in the world that could manufacture a simple recoil-operated automatic weapon but perhaps only a few score that could make ammunition for it. Surely ammunition control is a much better bet than gun control. The guns are there; the ammunition must be continually supplied.

F. D. REDDYHOFF
Ornaisons, France

Boutwell and Klare reply:

The telling point about Charles Taylor igniting a decades-long civil war in Liberia is not that his initial force was just 100 armed men but that his forces received continuous supplies of weapons through illegal channels year after year. Tighter control of ammunition to areas of conflict is but one of many options being seriously studied by governments and international organizations, although initial assessments are mixed regarding its ultimate feasibility.

In Sierra Leone, Liberia and other countries, arming civilians has only worsened the conflict, not prevented it. Such weapons are invariably stolen by rebels or used by criminals, or they ignite armed conflict where there was none. Only well-trained, well-disciplined soldiers of the national government or an international peacekeeping force can provide a real solution to internal conflict.

Incredible that a whole article could be written on this topic without mentioning the epidemic of small arms–related deaths in the U.S. How arrogant, blind and shameful.

BRYAN CEBULIAK
Mansfield, Queensland,
Australia

Having used an M16 rifle, I can verify that they do not fire at 750 to 900 rounds per minute, as listed in the chart "Supply and Demand" on page 50. The M16A2 rifle, for example, has a sustained rate of fire of only 12 to 15 rounds per minute. Given the fact that a soldier would need to constantly replace the magazine, it would be impossible to get a rate of fire of 750 to 900 rounds per minute.

ISAAC ERBELE
Cadet Private First Class,
U.S. Military Academy

Editors' note:

The "rounds per minute" number reflects the rate at which a gun can put one round through the reloading and firing cycle. As CPFC Erbele rightly points out, the actual number of rounds a soldier can fire in a minute is limited by factors such as how ammunition is fed into the gun.

In "The Human Cost of War," Walter C. Clemens, Jr., and J. David Singer attempt to make the case that in wars of this century, civilian deaths have outnumbered military deaths. I believe, though, that a more thorough review of the historical data will show that warfare has almost always been more devastating to civilian populations than to military opponents. It is only when you have a very disciplined, professional military force, under the direction of leaders who make a conscious effort to minimize civilian suffering, that you will find exceptions to this trend.

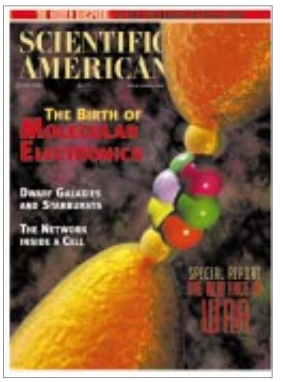
ROBERT HENDERSON
Portland, Ore.

CROAT SOLDIER fires from what was a bedroom window onto a street in Mostar, Bosnia-Herzegovina, in 1993.

JAMES NACHTWEY *Magnum*

THE_MAIL

READERS HAVE ON MORE THAN ONE OCCASION expressed the opinion that this magazine has no business wading into political waters, and, indeed, the special report "Waging a New Kind of War" [June] drew many such objections. Editor in chief John Rennie responds: "Science and technology are pervasively influential on culture and politics. People who take a detached, utopian view of science might prefer that *SA* confine itself to discussing 'pure science' rather than any of these social concerns. But this is narrow-minded and wrong. Science has an obligation to determine the facts, such as they are, but it also has an obligation to discuss the consequences of its findings. So if research can address, for example, whether turning children into soldiers causes them psychological harm that can be understood in the long term as a giant public health issue, you bet we'll cover it."



Clemens and Singer reply:

We agree that, throughout history, wars have often killed more civilians than warriors. Each war, however, is different. During America's Civil War more civilians died in the South than in the North, but overall most casualties were soldiers. As the chart in our article shows, relatively few civilians died in the Franco-Prussian War of 1870–71. Many times more Austrian and German soldiers died in World War I than civilians. The Korean War was bloody, but China and the U.S. suffered only military losses.

Recent bloodshed in the Balkans, in Africa and in Chechnya has probably raised the ratio of civilian to combatant deaths—precisely because the combatants are not disciplined professional troops. Many are part-time soldiers, and many deaths are the result not of combat but of outright massacres. When fighting is between neighbors, civilian suffering will be high.

Perhaps the recruiting of children as soldiers ["Come Children, Die," by Neil G. Boothby and Christine M. Knudsen] has "never played so large a role in warfare as it does today," but neither were child soldiers "bit players" in the past, and this was true long before the 1930s.

Drummer boys accompanied troops into combat from the 1600s through the late 19th century, and many preteenage males served on warships in the age of fighting sail, including Admiral David Farragut, who commanded a warprize vessel at age 14. Apprentice seamen and cadets in their early teens saw action in World War I. Prior to the 1930s, children and young teenagers fought and performed hazardous duties in many imperial settings, in Latin American wars and in the bitter guerrilla fighting of 1807–14 in Spain.

ROGER BEAUMONT
Bryan, Tex.

Letters to the editors should be sent to editors@sciam.com or to Scientific American, 415 Madison Ave., New York, NY 10017.

ERRATUM

"Deconstructing the Taboo," by Gary Taubes [SCIENTIFIC AMERICAN PRESENTS: *Building the Elite Athlete,* Fall 2000], incorrectly states that a review by Jonathan Marks of the book *Taboo*, by Jon Entine, had appeared in *Human Biology.* The review has not been published but is scheduled, according to that journal.

SCIENTIFIC AMERICAN

Sandra Ourusoff
PUBLISHER
saourusoff@sciam.com

NEW YORK ADVERTISING OFFICES
415 Madison Avenue
New York, NY 10017
212-451-8523 fax 212-754-1138

Denise Anderman
Associate Publisher
danderman@sciam.com

Wanda R. Knox
wknox@sciam.com

Darren Palmieri
dpalmieri@sciam.com

DETROIT
Edward A. Bartley
Midwest Manager
248-353-4411 fax 248-353-4360
ebartley@sciam.com

LOS ANGELES
Lisa K. Carden
West Coast Manager
310-234-2699 fax 310-234-2670
lcarden@sciam.com

SAN FRANCISCO
Debra Silver
San Francisco Manager
415-403-9030 fax 415-403-9033
dsilver@sciam.com

CHICAGO
ROCHA & ZOELLER MEDIA SALES
312-782-8855 fax 312-782-8857
mrrocha@aol.com
kzoeller1@aol.com

DALLAS
THE GRIFFITH GROUP
972-931-9001 fax 972-931-9074
lowcpm@onramp.net

CANADA
FENN COMPANY, INC.
905-833-6200 fax 905-833-2116
dfenn@canadads.com

U.K. AND SCANDINAVIA
Roy Edwards
International Advertising Director
Julie Swaysland
Chester House, 25 Ecclestone Place
London SW1W 9NF, England
+44 207 881-8434/35 fax +44 207 881-8503
redwards@sciam.com
jswaysland@sciam.com

GERMANY
Rupert Tonn
John Orchard
PUBLICITAS GERMANY GMBH
Oederweg 52-54
D-60318 Frankfurt am Main, Germany
+49 69 71 91 49 0 fax +49 69 71 91 49 30
rtonn@publicitas.com
jorchard@publicitas.com

MIDDLE EAST AND INDIA
PETER SMITH MEDIA & MARKETING
+44 140 484-1321 fax +44 140 484-1320

JAPAN
PACIFIC BUSINESS, INC.
+813-3661-6138 fax +813-3661-6139

KOREA
BISCOM, INC.
+822 739-7840 fax +822 732-3662

HONG KONG
HUTTON MEDIA LIMITED
+852 2528 9135 fax +852 2528 9281

OTHER EDITIONS OF SCIENTIFIC AMERICAN

SPEKTRUM DER WISSENSCHAFT
Verlagsgesellschaft mbH
Vangerowstrasse 20
69115 Heidelberg, GERMANY
tel: +49-6221-50460
redaktion@spektrum.com

POUR LA SCIENCE
Éditions Belin
8, rue Férou
75006 Paris, FRANCE
tel: +33-1-55-42-84-00

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

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

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

ΕΛΛΗΝΙΚΗ ΕΚΔΟΣΗ
SCIENTIFIC AMERICAN HELLAS SA
35–37 Sp. Mercouri St.
Gr 116 34 Athens, GREECE
tel: +301-72-94-354
sciam@otenet.gr

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

Galveston Hurricane, Bruno Bettelheim on Prejudice

OCTOBER 1950

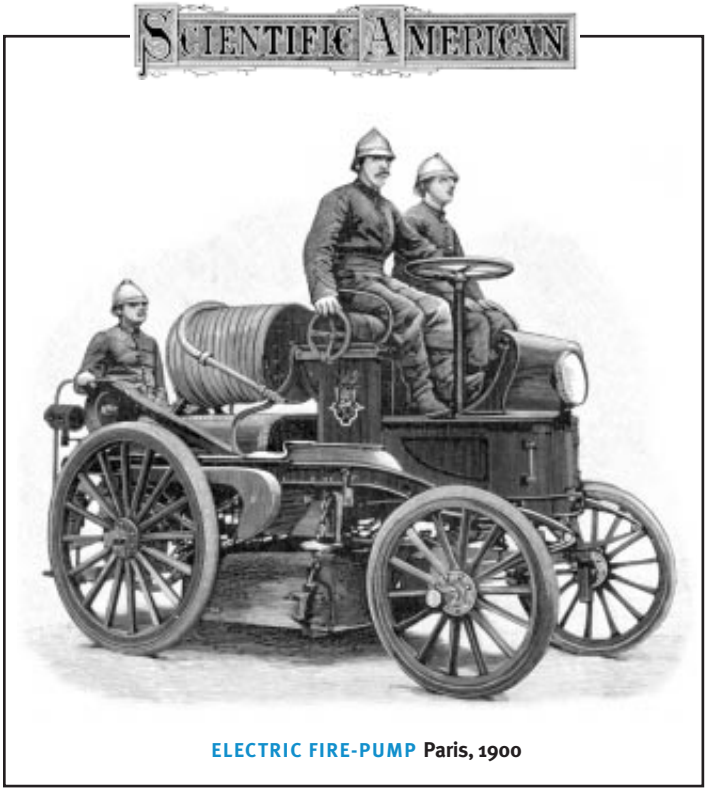
MICROSURGERY—"To work in the Lilliputian world of the cell, one obviously needs Lilliputian instruments. In recent years such high-precision instruments have been developed, and microsurgery on cells, known as micrurgy, has become an important part of the study of protoplasm. With a micromanipulator one can cut a cell into tiny fragments, remove the cell's nucleus or even its chromosomes, and inject fluids into either the nucleus or the cytoplasm."

PREJUDICE—"Of the many tension areas within our society the one we investigated was ethnic hostility—a polite term for racial prejudice. Our research supports the hypothesis that the individual's stereotypes are not only vitally needed defense mechanisms but are persistent, even under the impact of such immediate and realistic experiences as service with Jews and Negroes under conditions of war. Once a stereotype is formed, it is not easily changed. One can assume that as long as anxiety and insecurity persist as a root of intolerance, efforts to dispel stereotyped thinking or feelings of ethnic hostility by rational propaganda is at best a half-measure. —Bruno Bettelheim and Morris Janowitz"

OCTOBER 1900

THE GALVESTON DISASTER—"The special report by Isaac M. Cline, the Local Forecaster of the Weather Bureau at Galveston, verifies in the main the press reports of the recent disastrous hurricane. For a short time after 6:15 P.M. a maximum velocity of 100 miles per hour was recorded before the anemometer blew away. The tide at the Forecaster's residence at 7:30 P.M. is estimated in the report to have been 15.2 feet and rose, during the next hour, nearly 5 feet additional. 'These observations,' says Mr. Cline, 'were carefully taken.' By 8 P.M. a number of houses which had drifted and lodged against the Forecaster's house overthrew the building, and thirty-two persons out of the fifty who had taken refuge in it were hurled into eternity, including Mrs. Cline. The report states 'conservative estimates place the loss of life at the appalling figure of 6,000.' "

ALEUTS OF ALASKA—"The Aleuts of Alaska are skillful hunters, and the Russian government used to furnish them with supplies in exchange for furs. The disappearance of the food-bearing animals owing to persistent slaughter has made existence more difficult to them. In the construction of their native boats, or 'bidarkas,' the Aleuts are extremely ingenious and they manage them with wonderful skill. The bidarka is made of a frame of wood, covered over with the skin of the sea lion. The highest honors of chieftainship are conferred upon the best 'oarsman.' "



ELECTRIC FIRE-PUMP Paris, 1900

ANIMAL PLAGUES—"In the Department of Agriculture's Biochemical Division, the first work in the production of an antitoxine serum for hog cholera and swine plague was carried out experimentally in 1892. A soluble ferment was discovered in the cultures of the hog cholera germ and the relation of this enzyme to immunity was demonstrated. During the past three years anti-toxine serum for both hog cholera and swine plague have been prepared on a large experimental scale with very satisfactory results."

ELECTRICITY FIGHTS FIRE—"The firemen of Paris have an improved apparatus, which has been recently introduced for the purpose of enabling the firemen to reach and extinguish fires more rapidly than heretofore. Our illustration shows the electric fire-pump under way. The same electric motor which propels the carriage drives the pump as soon as the vehicle comes to a standstill at a fire."

OCTOBER 1850

MOONSCAPE—"The most remarkable feature of the lunar surface is the great number of rings, or craters, which almost entirely cover it, overlaying, intersecting and apparently elbowing each other out of the way. It is now pretty well demonstrated that these rings were the result of intense volcanic action at some remote period. In six-eighths of the lunar volcanic mountains, there was a cone in the center of the ring. The same thing is observed on extinct volcanic mountains on the earth, the cone in the center being the fruit of the last efforts of the expiring volcano. The moon has a proportionately larger surface area in relation to its mass than does the earth, and this fact was sufficient to explain the greater number of volcanic discharges that cover the surface of the moon."

KOPS—"The city of Pittsburgh has now two bodies of night watchmen, one appointed by the Mayor, and the other by the Police Committee. Their duties, so far, have been confined to arresting each other."

FROM SCIENTIFIC AMERICAN

OCEANOGRAPHY_HAZARD PREDICTION

Killer Waves on the East Coast?

Underwater landslides off the mid-Atlantic could trigger a tsunami, but the likelihood appears slim

If you perused any of several metropolitan newspapers along the Eastern seaboard this summer, you might have imagined a disaster of hurricane proportions striking the coast on a clear, blue day. With a sudden crumbling of the seafloor, the Atlantic Ocean would rise up and flatten Virginia Beach and Cape Hatteras. Giant waves might even surge up the Potomac River and flood the U.S. capital.

The notion of a tsunami striking the mid-Atlantic coast is startling—those disasters tend to hit earthquake-prone locales of the Pacific Rim, where land slipping along underwater faults sloshes the sea into threatening swells. But despite the breathless news reports, a long string of ifs and buts stretches between an imminent threat of an East Coast tsunami and its newly discovered potential cause: underwater landslides.

The landslide concern stems from new indications of looming instability atop the slope between the shallow continental shelf and the deep sea, off the coasts of North Carolina and New Jersey. Enormous cracks northeast of Cape Hatteras could be an underwater landslide in the making, three scientists suggested in the May *Geology*. Mud suddenly breaking loose and tearing downslope could displace enough water to swamp the nearby coastline with tsunami waves some five meters (15 feet) high—an event comparable to the storm surges of Hurricane Fran, which ravaged North Carolina in 1996.

The day after the media caught wind of the report, television helicopters were landing on the lawn of the Woods Hole Oceanographic Institution in Massachusetts, the workplace of the report's lead author, Neal W. Driscoll. Elsewhere, Driscoll's colleagues Jeffrey K. Weissel of Columbia University's Lamont-Doherty Earth Observatory and John A. Goff of the University of Texas at Austin were also fielding calls from eager reporters. "We underestimated the excitement the paper would cause," Weissel says.

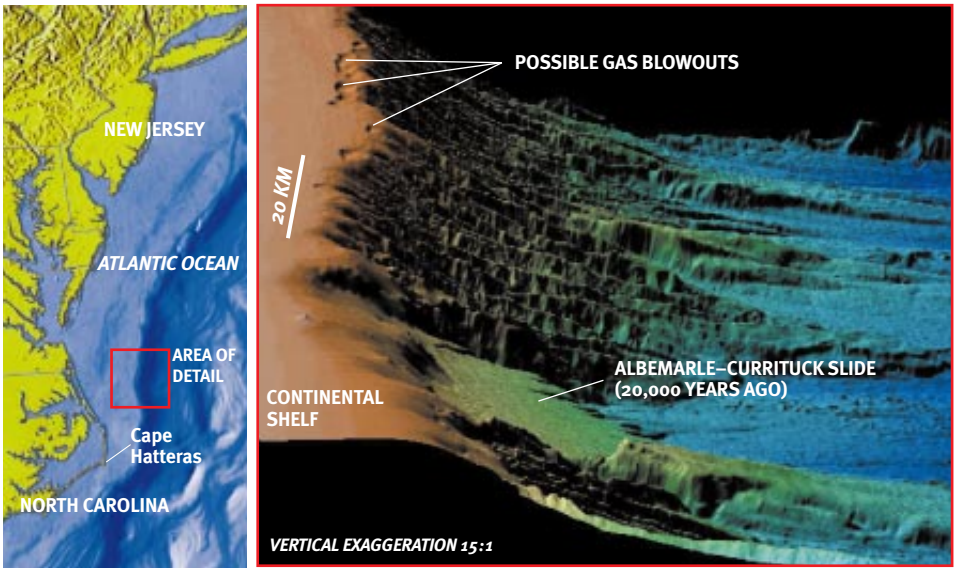
What the scientists knew—and what many news accounts failed to emphasize—was that although a tsunami would be devastating, the potential risk was remarkably unclear. At the time, the researchers had no idea when a landslide might occur (if ever), no mathematical predictions of the waves that might be generated and no evidence of a tsunami ever having struck the mid-Atlantic coast in the past. Still, Weissel maintains that "the paper would have been incomplete without a portion on tsunamis." At the heart of the scientists' concern is the growing evidence that underwater landslides—not earthquakes alone—pose a tsunami threat [see "Tsunami!," by Frank I. González, Scientific American, May 1999].

Oceanographers conducted the first intensive investigation of this theory after the 1998 Papua New Guinea tsunami. At least 2,200 people died—drowned, impaled on mangrove branches or bludgeoned by debris—when waves up to 15 meters high struck the country's north coast. A magnitude-7.1 earthquake had rocked the area only minutes before, but the waves were up to five times larger than expected for a quake that size. When oceanographers inspected the nearby seafloor, they found evidence of a landslide that could have enlarged the tsunami.

Two rare landslides in the western Atlantic also fuel the tsunami concern. In 1929 an earthquake-triggered landslide off Newfoundland's Grand Banks spawned a tsunami that killed 51 people. A similarly massive slide occurred some 20,000 years ago just to the south of the cracks discovered off the North Carolina coast.

Had scientists detected those cracks 10 years earlier, before underwater landslides were a suspected cause of tsunamis, their interpretations might have been different, Weissel says. But in light of this new historical evidence, his team couldn't ignore the possibility. Frank I. González, leader of the National Oceanic and Atmospheric Administration's tsunami research program in Seattle, agrees: "I think these guys were right on to call attention to the potential tsunami risk."

Based on sonar images, the cracks have



GIANT GAS BLOWOUTS, which may have cratered the seafloor off the North Carolina coast, could presage underwater landslides. Potential blowouts (*not shown*) also lie off New Jersey.

JOHNNY JOHNSON (*map*); BATHYMETRY IMAGE COURTESY OF JEFFREY K. WEISSEL *Lamont-Doherty Earth Observatory*

turned out to be giant craters—some five kilometers long and two kilometers across—that the team now thinks formed from eruptions of gas trapped in the sediments. What's more, additional gas is still waiting to blow. The researchers don't know when the past blowouts occurred, but they have reason to think they could have been explosive: such eruptions have destroyed oil rigs that penetrated gas deposits in the Gulf of Mexico and the North Sea.

In the July 14 *Science,* a second team reported another potential cause of seafloor blowouts. Peter B. Flemings and Brandon Dugan of Pennsylvania State University noted that explosions of waterlogged sediments could have carved several mysterious submarine canyons about 150 kilometers east of Atlantic City, N.J.

During an Ocean Drilling Program research cruise in 1997, Flemings and the crew drilled into one-million-year-old mud that contained up to 65 percent water. The soggy sediments were buried so fast that the water had nowhere to go. But the pressure caused by being buried 600 meters below the seafloor means that deep erosion could unleash the water with a bang. Flemings and Dugan didn't mention tsunamis in their journal article, but the media didn't miss the connection. "I just continually remind people that we haven't done any work on whether a tsunami would be generated," Dugan says.

Such a prediction would be difficult to make, anyway. It takes a sudden flow of a large volume of mud to create a tsunami; the scientists don't know whether the canyons formed quickly—in one explosive event—or eroded over tens of thousands of years. Even today, muddy seeps and geysers bleed off trapped water little by little.

Nor is it clear whether gas blowouts farther south would stabilize the slope by reducing the pressurized gas or destabilize it by rendering the shelf edge more precariously balanced than before, Weissel says. The fact is that landslides may never occur in either region. And until scientists can estimate the frequency of landslides—whatever the cause—it will be impossible to calculate the probability of a future tsunami. From Dugan's perspective, the bottom line is this: "Are these blowouts preventable? No. Should people be worried? No."

—Sarah Simpson

EARTHQUAKES_ENGINEERING

Riding the Rumble

A $400,000 house is given a good shaking in the name of science

SAN DIEGO—July 11 was a slow day for earthquakes in southern California—except at the Powell Structural Research Laboratory at the University of California at San Diego, where a magnitude-6.7 temblor battered and bruised a $400,000 experimental home. Built atop a giant shake table, the two-story, fully furnished wood-frame house rode out the simulated earthquake surprisingly well, but the building's contents were reduced to a shambles.

Part research, part public education and part silly-season entertainment, the event was covered live on several television channels and on the Internet (video clips are available at www.curee.org). Beforehand, principal investigator André Filiatrault explained that the shake test was part of a $7-million project, funded by the Federal Emergency Management Agency, to assess and improve the seismic behavior of wood-frame buildings—the kind in which nearly all Californians live. Such buildings sustained $12 billion in damage during the 1994 Northridge earthquake.

Four siren blasts then heralded the earthquake itself. As cameras rolled, a computer-controlled hydraulic ram shoved the table back and forth, precisely reproducing the motions recorded by accelerometers near the epicenter of the Northridge earthquake.

For the first five seconds, as the early arriving primary seismic waves rumbled through the structure, the house did little more than tremble. Then the larger-amplitude secondary waves arrived, jolting the house sideways with an acceleration of 1 g and a peak velocity of 40 inches per second. An unanchored water heater overturned, flooding the living room floor and rupturing its connection to the (nonexistent) gas supply. File cabinets, bookshelves and tables were also upended, while upstairs a speaker landed on a child's bed and bounced onto the floor. Heavy clay planters fell from upper-story window ledges into the direct path of anyone who might have been exiting the building.

The 15-second simulated earthquake had been billed as the "ultimate jolt" and was expected to "severely damage" the house. Some onlookers were thus disappointed when the house neither fell down nor suffered any visible structural damage beyond minor cracking in the exterior stucco and the interior drywall. Filiatrault himself expressed some surprise at the outcome, because earlier shake tests, conducted before the stucco and drywall were applied, had caused much more extensive damage. "This seems to indicate

MODEL HOME (*top*) survived a simulated magnitude-6.7 earthquake, although the furnishings didn't fare as well (*bottom*).

FRIEDER SEIBLE *U.C.S.D. Structural Engineering (top);* RICARDO GUTHRIE *U.C.S.D. Communications (bottom)*

that wall-finish materials, which are usually neglected at the design level, actually have a tremendous effect on a building's response to an earthquake," he said. But he reserved final judgment until the output of more than 300 sensors embedded in the structure has been analyzed, a task that will take several months.

According to project manager John F. Hall, a civil engineer at the California Institute of Technology, many wood-frame buildings would not perform as well as the one subjected to the U.C.S.D. test. A large number of homes were built prior to the introduction of modern construction codes and have insufficient foundation anchoring and shear-resistant elements, such as plywood panels. Also, Hall says, the stucco on older homes tends to be deteriorated and would not hold up to prolonged shaking. Researchers plan to construct buildings to precode standards and then study the effects of various retrofitting techniques on their behavior on the shake table.

A type of wood-frame building that performed especially poorly in the Northridge earthquake was the "tuck-under" apartment building, in which one side of the first floor consists of open-entry parking space. According to Hall, such structures yield unevenly as seismic waves pass through. So researchers at the University of California at Berkeley are devising an entire tuck-under apartment complex on a shake table even larger and more powerful than the one at U.C.S.D. Unlike that one, the Berkeley table moves along all three axes. Thus, the test, scheduled for a few months from now, should challenge the building with the closest thing possible to an earthquake—short of the long-awaited rupture of the East Bay's Hayward Fault. —*Simon LeVay*

SIMON LEVAY is co-author of The Earth in Turmoil: Earthquakes, Volcanoes, and Their Impact on Humankind.

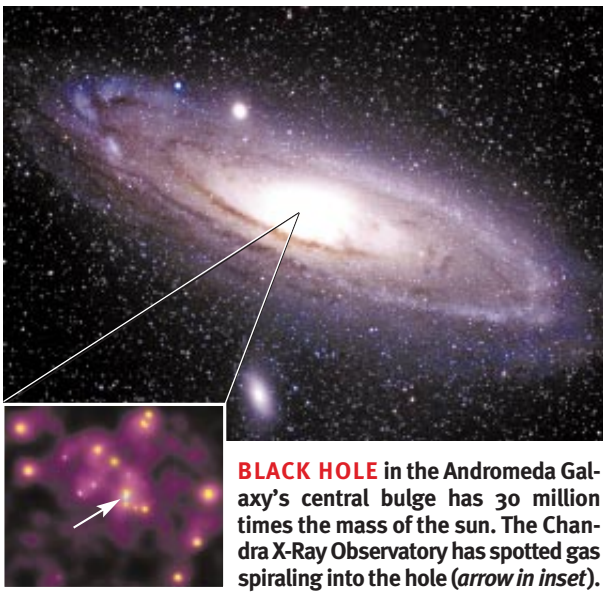
ASTRONOMY_BLACK HOLES

The Hole Shebang

Black holes and galaxies may be entwined from birth

One of the great accomplishments of astronomy over the past century has been to explain stars. Despite their amazing variety, stars all go through the same basic life cycle, driven by a few basic processes such as gravitational collapse and nuclear fusion. Now astronomers are on the verge of a similar synthesis for galaxies. From the diffuse band of light we call the Milky Way to the blindingly bright quasars near the edge of known space, galaxies all take shape in much the same way: through mutual interactions and—according to the latest findings—the exertions of supermassive black holes.

Black holes are notoriously destructive, but their creative potential has gradually come to light. Observations of quasars in the 1960s and 1970s suggested that only a giant hole—a mass of a million or a billion suns—could power them, and by the late 1980s mammoth holes had taken the credit for all kinds of anomalously luminous galaxies. The pace of discovery has accelerated lately, helped along by the ultrahigh resolution of the Hubble Space Telescope and the Very Long Baseline Array radio telescope. Stars and interstellar gas clouds near the center of many a galaxy are moving abnormally fast, whipped up by the gravity of a colossal unseen body—most probably a hole, although alternatives both prosaic (dense star clusters) and exotic (neutrino balls) have not been ruled out.



BLACK HOLE in the Andromeda Galaxy's central bulge has 30 million times the mass of the sun. The Chandra X-Ray Observatory has spotted gas spiraling into the hole (*arrow in inset*).

The count of supermassive holes is now up to 34, and two remarkable trends have emerged. First, supermassive black holes appear not just in quasar or quasar-like galaxies but in unostentatious ones, too. In fact, the only requirement seems to be an ellipsoidal shape: either an elliptical galaxy or a bulge in an otherwise flat galaxy, as in our Milky Way. Completely flat galaxies lack large holes. Second, the mass of each hole is roughly proportional to the mass of the ellipsoidal host (as estimated from its brightness). Holes weigh in at 0.15 percent of the mass of their elliptical galaxies or bulges. The Milky Way's modest hole befits its modest bulge.

A third trend has now been discovered by two teams writing in the August 10 *Astrophysical Journal Letters*: Karl Gebhardt and John Kormendy of the University of Texas at Austin, Douglas Richstone of the University of Michigan, and their colleagues; and Laura Ferrarese of the University of California at Los Angeles and David Merritt of Rutgers University. These researchers found that the mass of a black hole is related to the average velocity of stars within its ellipsoidal host, even in areas beyond the hole's direct influence. In fact, within the error bars, the velocity correlation is perfect. It almost has the status of a new law of nature, akin to Kepler's laws of planetary motion. "I'm surprised it's as tight as it is," Richstone says.

Although the two teams disagree as to the precise formula for this correlation, both analyses imply that black holes are somehow tailor-made for their galaxies. But how? Did the black hole come first and then determine the mass of the ellipsoid, or was it the other way around? Kormendy points out that the stellar velocity depends not just on the mass of a galactic

JASON WARE *Galaxy Photography:* NASA, CHANDRA X-RAY OBSERVATORY CENTER, AND STEPHEN S. MURRAY AND MICHAEL GARCIA *Smithsonian Astrophysical Observatory (inset)*

ellipsoid but also on its size: the smaller the ellipsoid, the faster its stars move. This extra effect, he argues, is what makes the velocity correlation so much better than the brightness correlation. To have an extra-heavy black hole, a galaxy or bulge must, in addition to being massive, also be unusually small and dense.

In other words, the mass of the hole is determined by whatever fixes the size of the bulge. The size is thought to be set during bulge formation; once gas has condensed into stars, the size is locked in forever. So it seems that neither hole nor bulge came first; they developed together, limited by the amount of available material and by the gravity of the dark matter that ultimately calls the shots in this universe. The new correlation thus supports theories that quasars are black holes and bulges in the throes of growth. Such episodes, though not always so intense, may be a natural part of the life cycle of most galaxies, triggered by interactions or mergers with other galaxies. The quasar phase overlaps with the appearance of the first stars in the bulge. Some galaxies then acquire a flattened disk, such as the one the sun lives in. Other galaxies begin as a disk and later develop a bulge with a black hole.

Richstone, however, worries that this scenario may contradict the available historical evidence, which suggests a gap in time between the onset of quasar activity and that of star formation. Kormendy disagrees but admits that the scenario is still sketchy. Nobody quite knows what created the "seed" hole that then grew to massive proportions, how material steered itself into the hole or why the beast decided to stop eating. "The art of doing science is to try to get the right answer with imperfect data," he says.

To improve the data, Hubble is testing the correlation between black holes and bulges at its weakest points—namely, the heaviest and lightest galaxies. Meanwhile the Chandra x-ray satellite is probing black holes in galaxies near and far, and ground-based instruments continue to inspect the black hole at the heart of the Milky Way. Over the coming decade new satellites—the Space Infrared Telescope Facility and the Next Generation Space Telescope—will scrutinize the very earliest galaxies. By then, black holes may have lost their reputation as quintessentially bizarre objects. They are already beginning to seem rather mundane. —*George Musser*

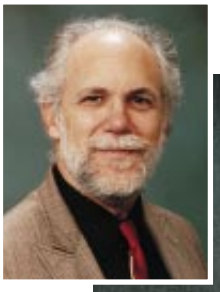
Science on the Canvas

Ron Miller has made a career of unveiling the cosmos. His gorgeous renderings of planets, stars and other heavenly objects have appeared in scores of books and magazines. With astronomer William K. Hartmann, he co-authored *The Grand Tour,* a classic pictorial guide to the solar system. He has also been a production illustrator for science fiction movies such as *Dune* and *Total Recall*. More recently he has turned to writing fantasy novels; his latest effort, *Bradamant* (Timberwolf Press, 2000), is based on the 16th-century epic poem *Orlando Furioso*. Here Miller discusses the nebulous boundary between science and art. —*Mark Alpert*

Now that people can watch space shuttle flights on television, is it more difficult to create exciting space art?

Yes, people are more jaded. The golden era for space illustrators was from 1945 to 1960, when manned flights were still a thing of the imagination. The great master then was Chesley Bonestell [see "Chesley Bonestell's Astronomical Visions," by Ron Miller; Scientific American, May 1994]. Have you ever seen his painting *Saturn from Titan*? It's often been called the painting that launched a thousand careers, because it inspired so many young people to go into science. We know now that it's not accurate—it shows a blue sky above Titan, for one thing. But it's a beautiful piece of art. No one had ever seen anything like it before.

Nowadays, images of space are much more commonplace. But there's still a need for good space art. When most people look at the recent images from the Mars Global Surveyor spacecraft, they don't know what to make of them. But a space artist could give you a view of the same geological formations as if you were standing on Mars's surface. It provides the kind of drama that a layperson can't get from a satellite photo. And remember, those people are the ones who are paying for the space missions.

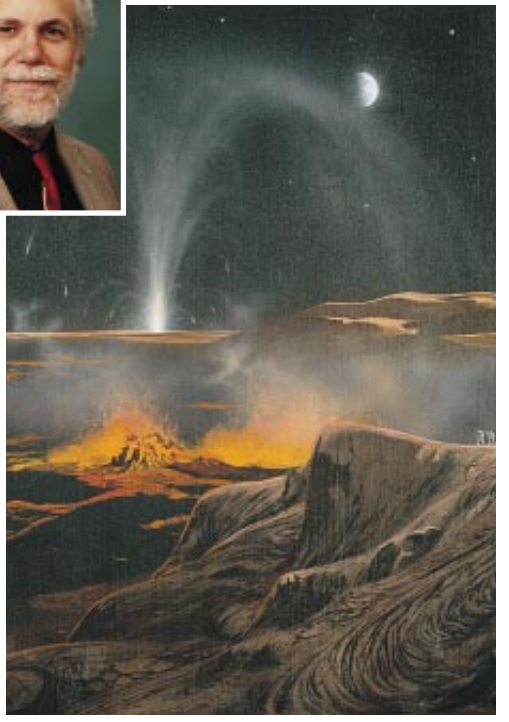


Which current artists are following in Bonestell's footsteps?

Don Davis and Michael Carroll are doing outstanding work in astronomical art. For spacecraft hardware, I like Robert McCall. He's been doing space art for 50 years, and he's always been ahead of his time. I also like the work of Syd Mead. His specialty is designing future technologies. Not only does he show you what the technologies look like, he shows you how they work.

What about Hollywood? Why aren't there more good space movies?

The people in Hollywood are way behind the times. They keep putting out remakes of *Star Wars* and *Alien*. Sometimes they make an effort to get the science right, like in *Mission to Mars*. But that movie was appallingly derivative. The only good space film to come out in the past few years was *Apollo 13*. It's not that hard to make good space movies, and yet they're very rare.

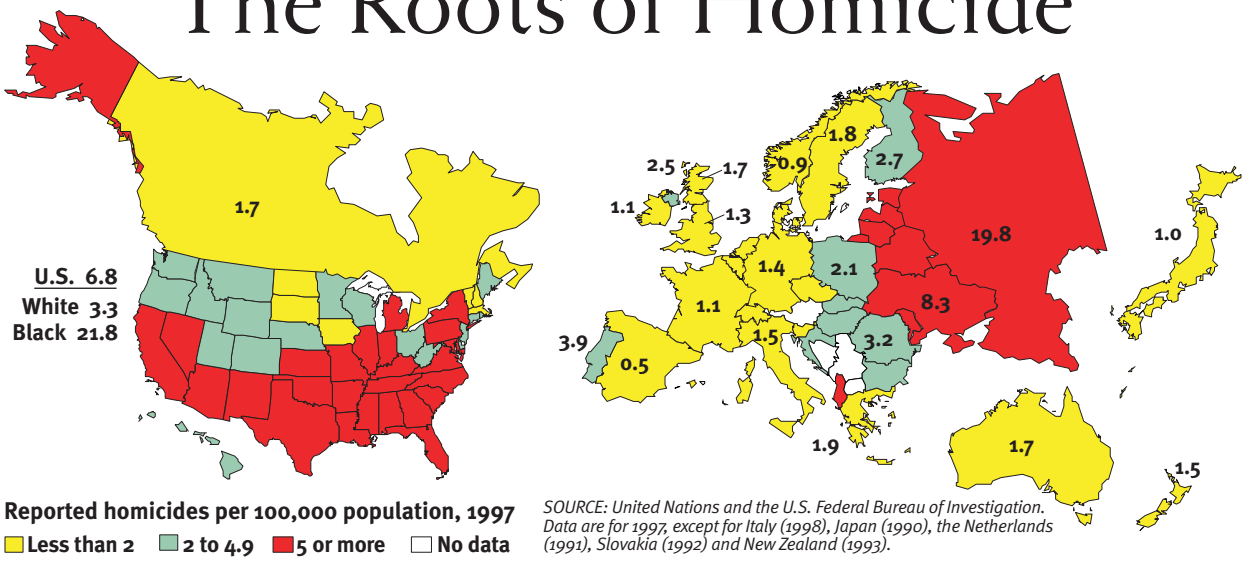


VOLCANO ON IO, **by Ron Miller.**

RON MILLER; BLACK CAT STUDIOS (*inset*)

MORTALITY_CRIME

The Roots of Homicide



SOURCE: United Nations and the U.S. Federal Bureau of Investigation. Data are for 1997, except for Italy (1998), Japan (1990), the Netherlands (1991), Slovakia (1992) and New Zealand (1993).

The U.S. property crime rate matches those of most other industrialized countries, but its homicide rate exceeds western Europe's by 4 to 1 and Japan's by 7 to 1. The historical roots of this disparity may lie not in the Western frontier, as many believe, but in the institution of slavery and the unusual history of firearms in America.

In the antebellum South, whites used the threat of violence to intimidate blacks and encourage deference. In the view of historian Roger Lane of Haverford College, the respect demanded of slaves fostered a "culture of honor," in which a man's personal worth was measured by how others behaved toward him. Trivial slights had to be answered immediately and with physical force, if necessary. Homicide resulting from quarrels did not usually result in a conviction. The Southern culture of honor spread to poor whites and to the slaves themselves, who eventually brought it to the inner cities of the North. Disrespect for the law was reinforced by the tendency of authorities to ignore murders of blacks by blacks. Current high homicide rates in the former Confederate states and in many large cities trace largely to the attitudes developed during slavery, according to Lane. He also says that high rates in the Southwest reflect in part attitudes among Mexican-Americans, many of whom also practice a culture of honor tracing to the region's historical circumstances.

The American attitude on firearms is rooted in British North America, where all freemen, except in Quaker Pennsylvania, were required to carry arms for protection against the Indians, the French and others. The colonial era's long guns and dueling pistols were expensive and hard to manipulate and thus were not often used in disputes. But then in the 1840s came the more efficient, cheaper and easily concealed Colt revolvers and with them, an increase in white homicide rates. More than 80 percent of gun murders today involve a handgun.

Among Western industrialized nations, gun ownership correlates with homicide: in England and Wales, where virtually no one owns a gun, the homicide rate in 1997 was only 1.3 per 100,000 population, whereas in Finland, which has the highest gun ownership level, the homicide rate was 2.7. If gun ownership were the only determinant of homicide, the U.S. rate would fall into the intermediate category shown on the map. It is the *combination* of easy access to guns and an extraordinary readiness to use them that helps make the U.S. homicide rate so high. According to Franklin Zimring and Gordon Hawkins of the University of California at Berkeley, up to half the difference in homicide rates between the U.S. and Europe is explained by greater gun use by Americans.

The U.S. has seen several waves of homicide, including one that peaked before the Civil War, a possible second wave that crested in the 1920s, and the current wave, which peaked in 1980. The ascending phase of this wave, which began in about 1960, more or less coincided with several trends that have been proposed as contributors to homicide: the decline of union manufacturing jobs; the breakup of families with the rise in divorce; the increase in births to unwed mothers; and the growth of illegal drug use. The decline in rates since 1991 coincided with the waning of the crack cocaine epidemic that started in 1985. Other developments, including greater police efforts to prevent gun carrying and the recent economic expansion, which provided more jobs, have played a role. The proportion of young men, always the most violent group in society, fell in the 1990s and so also contributed to the decline in homicides.

One of the most hopeful developments of recent years is detailed by Richard Curtis of the John Jay College of Criminal Justice, who found that many disadvantaged Puerto Rican and black youths in New York City became deeply disenchanted with the drug use of parents and older siblings and are now attempting to reestablish their lives and their communities. Curtis believes that similar developments are happening in other cities across the country. Still, no one knows how the next generation of young men will feel and act, and no one can predict what devastating new drug might be concocted or how the fast-changing U.S. economy will affect the murder rate.

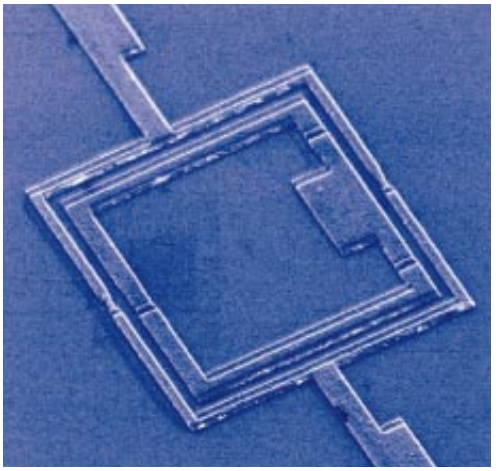
—Rodger Doyle (rdoyle2@aol.com)

RODGER DOYLE

Schrödinger's SQUID

In superconducting loops, electric current flows both ways at once

Schrödinger's cat turned 65 this year, but instead of thinking about retirement, the quantum feline is making increasingly bolder appearances. Recently two independent groups have demonstrated the largest examples of Schrödinger's cat states by using superconducting loops. In the original thought experiment, quantum effects and a Rube Goldberg–like poison apparatus rendered the cat simultaneously alive and dead inside its sealed torture chamber. In the new experiments, an electric current stood in for the cat and flowed both ways around a loop at the same time. Tony Leggett of the University of Illinois at Urbana-Champaign, one who suggested in the 1980s that such large quantum mechanical systems could be demonstrated, calls the research "a milestone in experimental quantum physics."



CASPAR VAN DER WAL *Delft University of Technology*

QUANTUM CAT BOX: **Outer SQUID detects superposition of currents flowing in inner SQUID.**

The key phenomenon at work is superposition of waves—similar to the way different individual sound waves from people chatting at a party overlap and add up to a total sound wave that goes into our ears. In quantum mechanics, matter itself behaves like a wave: electrons and other particles can exist in superpositions of different states.

The problem, as Erwin Schrödinger pointed out in 1935, is to understand why "quite ridiculous" superpositions like that of his cat are never seen in reality, despite there being no prohibition of them in unadulterated quantum mechanics. Today theorists have a much better understanding of how tiny disturbances from the environment tend to upset quantum superpositions and turn them into the unambiguous reality that we see around us every day—a process known as decoherence. Conversely, in the past decade experimenters have created and scrutinized coherent quantum states with a degree of control only dreamed of in idealized textbook descriptions. Experiments have superposed small numbers of particles and put individual atoms in two places at once.

The two new experiments take things to a substantially more macroscopic level. They were conducted by Jonathan Friedman, James Lukens and their co-workers at the State University of New York at Stony Brook and by Caspar van der Wal, Johan E. Mooij and their co-workers at the Delft University of Technology in the Netherlands. Both groups used SQUIDs—superconducting quantum interference devices. Quantum effects permit only certain discrete amounts of magnetic flux to thread through such a superconducting loop. If a field is applied that lies between the allowed values, an electric current flows around the loop, generating just the right additional field to round off the total flux to an allowed value.

Things get interesting when the applied flux is midway between two allowed values. That makes the SQUID equally inclined to produce a clockwise or a counterclockwise current—to round up or round down the incommensurate flux—and conditions are most favorable for producing a superposition of these two alternatives. For the Stony Brook SQUID, these currents amounted to flows of billions of electrons, totaling microamps, traveling around a 140-micron-square loop, large enough to encircle a human hair—gargantuan by quantum standards

and truly macroscopic. The Delft design was smaller, 1/30 the size.

The superposition state does not correspond to a billion electrons flowing one way and a billion others flowing the other way. Superconducting electrons move en masse. *All* the superconducting electrons in the SQUID flow both ways around the loop at once when they are in the Schrödinger's cat state.

Important differences remain, however, between these devices and the canonical thought experiment. In the imaginary scenario, the superposition of alive and dead cat inside the box is static, from the time it is created until the lid is opened and the experimenter sees one outcome or the other. Two idealizations are at work here. One is that the interior of the box is so well isolated that the superposition remains undisturbed until the lid is opened. In the present SQUID experiments nearby devices spoil this isolation, and the superpositions decohere rapidly, possibly within a few nanoseconds.

The second idealization relates to there being two closely related superpositions: one can have the alive state plus the dead state, or the alive state minus the dead state. So far as the cat is concerned, this mathematical distinction is immaterial. Either case amounts to a 50 percent mortality rate. In the SQUID experiments, in contrast, quantum mechanics predicts that the two alternatives will have slightly different energies. Detecting this energy difference is how the two experiments inferred that the cat state was achieved.

The existence of two different superposition states with unequal energy also implies that an oscillating state should be observable: in this case, the probability of detecting the clockwise current, say, would oscillate between zero and 100 percent, depending on the time delay from the preparation of the state to the measurement. Although it may sound less bizarre than a static superposition, consider what it would mean for the cat: while the box remained sealed, the cat would be oscillating from 100 percent alive to 100 percent dead and back!

Leggett considers such oscillation experiments (on SQUIDs, not cats) to be the crucial goal in testing the quantum mechanical predictions. Then physicists will be examining just what happens inside Schrödinger's sealed cat box. Both groups are working on developing just that capability. —*Graham P. Collins*

GENETICS_FETAL DEVELOPMENT

Womb Wars

New evidence that a mother's and father's "imprinted genes" battle to determine a baby's size

LONDON—Mammalian parents are caught in a fierce conflict over developing embryos. A father wants bigger babies, because they are usually healthier and fitter. But a mother prefers them to be small, because a fetus that grows too large may drain her resources and jeopardize her ability to sustain additional pregnancies. Scientists have long suspected that special, chemically marked genes, called imprinted genes, are the weapons wielded in this parental warfare over size. Exactly how genes fight such a battle, though, has remained a mystery. Now biologists have discovered a clever strategy employed by the father's genes to gain the upper hand.

The father's genes tend to aim for large, nutrient-guzzling babies to maximize the chances of being passed on. Because few male mammals are monogamous, the father can be fairly sure that the next litter will not be his, and so his genes strive for copious nourishment to be pumped into his developing offspring. "Paternal genes are greedy and suck out precious nutrients from the mother," explains Wolf Reik, a molecular biologist at the Babraham Institute near Cambridge, England. But the fetus "can kill the mother if [it grows] too big," he says.

BIGGER IS BETTER from the point of view of the father's genes; maternal genes strive for smaller, less nutritionally demanding fetuses.

G. MOSCOSO SPL/Photo Researchers, Inc.

Researchers believe that imprinted genes may have evolved from this tussle between the sexes. Embryos receive two copies of every gene: one from Mom and one from Dad. Most of the time, the genes in a pair behave in exactly the same way and therefore cannot be distinguished from each other. But imprinted genes are different. They are tagged with methyl groups (one carbon and three hydrogen atoms) that instruct them to become active or to remain silent, depending on which parent they come from.

At first, many investigators found it hard to believe that genes could behave differently depending on their parentage. But it is now widely accepted that some genes work only when inherited from the egg and are silenced if coming from the sperm, or the other way around. To date, researchers have found 40 or so imprinted genes in mice and humans, although their true numbers may well be in the hundreds.

"The argument runs that paternally derived genes and maternally derived genes have different evolutionary interests," says Laurence D. Hurst, an evolutionary geneti-

cist from the University of Bath. A "parental conflict" theory has been put forward to explain the genetic battle that ensues as soon as egg meets sperm. "It sounds waffly," Hurst admits, "but it has a very sound basis." In fact, the first imprinted gene to be discovered—a gene for a growth hormone called *Igf2,* for insulin-like growth factor 2—slots neatly into this hypothesis. *Igf2* leads to bigger offspring in both mice and humans and is active only when it comes from the father.

"If the balance between [male and female] imprinted genes is okay, then you have normal-size babies," Reik explains. But what happens if you knock *Igf2* out of the equation? The parental conflict theory predicts that the equilibrium should shift in favor of the mother, who will then give birth to smaller offspring. To find out whether this is true, Miguel Constancia, from Reik's team, genetically engineered mice that lack a section of *Igf2* DNA expressed exclusively in the placenta. Reik told researchers at a recent Novartis Foundation symposium in London that when this region of paternal *Igf2* is missing, litters are born 30 percent smaller than normal. Reassuringly, these mutant newborns soon catch up with the controls—evidence that the pups are normal despite their stunted growth in the uterus.

These mutant mice have thus revealed one of *Igf2*'s tactics. The gene operates in a vital area of the placenta: the labyrinthine trophoblasts, where maternal and fetal blood mix and nutrients are exchanged. From this prime location, this paternal gene encourages the influx of nutrients to the fetus. "It's beautiful work, and the placental effects certainly support the conflict model," Hurst comments on Reik's findings, which have been submitted for publication.

But do the placental results confirm paternal genes as the winners? They do not, Reik says. "Neither of them really wins. The maternal and paternal genomes are both being quite inventive. When one comes up with a new thing, the other has to invent a counterweapon." Yet despite these elegant findings, much about imprinted genes remains a mystery, and researchers are still unsure whether genetic warfare explains all aspects of imprinting. —*Lisa Melton*

LISA MELTON is a science writer based in London. She recently gave birth and, based on the baby's size, believes that the father's genes fought a hard battle.

BIOLOGY_CELL DEATH

One-Hit Wonder

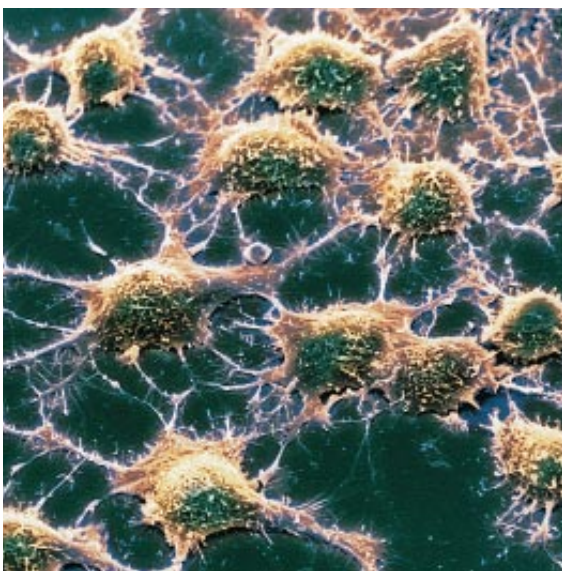
Upending the model of how neurons die in degenerative brain diseases

Although therapies exist that mitigate the devastating effects of Alzheimer's, Huntington's and other inherited neurodegenerative diseases, nothing can stop the slow but relentless progression of neuronal death. Now scientists examining the way in which the cells die have discovered that widely held beliefs regarding the pathways culminating in cell death may be inaccurate and that, despite the widely varying symptoms and pathogenesis of neurodegenerative diseases, all sick neurons die in an analogous fashion. If correct, the new model could lead to life-sustaining drugs for neurons.

Everyone loses brain cells over time. In inherited neurodegenerative diseases, neurons die much more frequently than on average. The prevailing view of neuronal death is called the cumulative damage hypothesis. Cells suffering from neurodegenerative diseases become sensitive to biochemical stress, which can lead to proteins that stick to one another and form damaging clumps called fibrils and to deleterious molecules such as oxygen free radicals. Low levels may be tolerated, but eventually the concentration of the fibrils and free radicals builds up and overwhelms the cells, killing them.

In this scenario, the probability of a neuron dying increases over time as neurons yield to the damaging effects of stress. Accordingly, the status of a given neuron is an indicator of the health of all the other neurons: if any single neuron has become sick enough to die, other neurons are most likely about to die as well. In a way, this situation resembles human demographics in that few young people die, but the probability of death increases dramatically in older populations. The cumulative death hypothesis suggests that the change in the percent of healthy neurons over time stays relatively flat and then plummets dramatically (mathematically, it follows a so-called sigmoidal curve).

Recently a group led by Roderick R. McInnes of the University of Toronto decided to investigate this hypothesis. The team obtained data from more than 10 previous studies covering different neuron subtypes associated with illnesses such as vision impairment, Parkinson's disease and Huntington's disease. To the surprise of the researchers, a mathematical analysis revealed that none of the neurons died in a pattern consistent with the cumulative death hypothesis: there was no dramatic, sudden loss of the neuronal population. Instead the neurons died at a fairly constant rate regardless of the stage of the disease. The neurons in a person with an inherited neurodegenerative disease function much like people who have high cholesterol, McInnes explains. "By all accounts they function



ALZHEIMER'S DISEASE leaves damaging plaques of protein on neurons. Sick neural cells seem to die in the same way for a wide range of neurodegenerative diseases, suggesting that a universal therapy is possible.

SIMON FRASER SPL/Photo Researchers, Inc.

pretty well," he says. "They can run 10 miles a day. They feel healthy, but they're at increased risk of dropping dead from a heart attack."

To account for these results, which appear in the July 13 *Nature,* the researchers devised a new scenario, which they refer to as the one-hit, or catastrophic event, model. In this theory, "at any given interval the same percent of neurons would die," says Harvard University geneticist Thaddeus Dryja, who is familiar with the team's work. Cell death, rather than the result of cumulative damage, is a random process—the status of one extremely sick neuron has no implications for the health of the other neurons. So why do symptoms appear only late in life? According to Rockefeller University biologist Nathaniel Heintz, "A late-onset disease may not significantly disturb cell homeostasis. As a result, there is a low probability of cell death and a long period until enough cells die to cause noticeable dysfunction."

McInnes and his colleagues also attempt to explain the molecular events that may cause cell death to occur randomly. The genetic mutations associated with most neurodegenerative diseases lead to proteins that misbehave. The Toronto team proposes that cells carrying these mutations are in a "mutant steady state" because of the presence of "mutant response proteins." These proteins could be enzymes that help produce a compound needed for the cell to function normally. In neurodegenerative diseases these enzymes consistently make higher-than-normal levels of this compound—and may occasionally produce deadly amounts. The proteins, however, have yet to be identified. "We've proposed that these mutant response genes or proteins exist, so the challenge is to find them," remarks McInnes, who acknowledges that the idea is just the beginning of a complicated story.

Because the one-hit model applies to many neurodegenerative disorders, it has several implications for treatment. Therapies could halt the progress of disease by keeping the putative mutant response proteins below threshold levels. Such an approach would therefore limit cell death, unlike currently available treatments. Even more promising, the one-hit model suggests that the probability of rescuing neurons does not decrease with age. No matter how far along in their disease, patients would benefit from the treatment. —*Rebecca Lipsitz*

SCIENTIFIC AMERICAN

Help Desk

For your convenience, we have summarized the answers to often-asked customer service questions and how to get in touch with us.

Scientific American Explorations
Our new quarterly publication dedicated to making it both easy and fun for adults and children to enjoy science and technology together. Annual subscriptions: In U.S. $15.80. Elsewhere $19.80. Call (800)285-5264, or visit the *Explorations* Web site at www.explorations.org

Scientific American Frontiers
The magazine's PBS television series, is hosted by actor and life long science buff, Alan Alda. Beginning with the Fall 2000 broadcast season, there will be 10 one-hour episodes each season, with *Frontiers* becoming an integral part of a new PBS science programming initiative. Visit the *Scientific American Frontiers* Web site at www.pbs.org/saf

www.sciam.com
Our award-winning Internet resource, updated weekly and thoroughly linked to related sites. Here you will find timely and interesting current news, recent illustrated scientific articles, current issue highlights, ask the editors, our on-line store, subscriptions, e-mail, and other features not found in *Scientific American*, and much more.

Scientific American subscription rates
In U.S., *Scientific American* is $34.97. Elsewhere $55. Your 12-issue subscription includes all regular issues with the in-depth special reports. Our 2000 report subjects: May: Data Storage Technology, July: The Human Genome Business, October: The Wireless Web.

Subscription inquiries
Call us: In the U.S. and Canada: (800) 333-1199. In other countries: (515) 247-7631.
E-mail: subscriptions@sciam.com. Or write: Subscription Manager; *Scientific American*; PO Box 3187; Harlan, IA 51537. The date of the last issue of your subscription appears on the mailing label. For change of address, please notify us at least four weeks in advance, and include both old and new addresses.

Back issues
In U.S. $9.95 each. Elsewhere $12.95 each. Many issues are available. Fax your order with your Visa, MasterCard or AMEX information to: (212) 355-0408.

Reprints
$4 each, minimum order 10 copies, prepaid. Limited availability. Write to: Reprint Department; *Scientific American*; 415 Madison Avenue; New York, NY 10017-1111; fax: (212) 355-0408 or tel: (212) 451-8877.

PHYSICS

The Tau of Neutrinos

On July 21 physicists announced that they had directly detected the tau neutrino, the third such species (after the electron and muon neutrinos), completing the third and presumably final generation of elementary particles of matter in the Standard Model of particle physics (see www.fnal.gov/pub/standardmodel.html). The experiment, conducted in 1997 at the Fermi National Accelerator Laboratory in Batavia, Ill., slammed protons into tungsten, producing a thick stream of particles. Powerful deflecting magnets and 15 meters of steel shielding removed all but the neutrinos—almost massless, barely interacting particles. An estimated one in a trillion of the beam's tau neutrinos then produced its distinctive partner, a tau lepton, in the experiment's detector. Three years of painstaking data analysis, conducted at Nagoya University in Japan, turned up four tau neutrino events among nearly seven million candidate interactions.

—Graham P. Collins

ECOLOGY

Sea of Troubles

Threat to shrimp?

MONTY GRAHAM *Dauphin Island Sea Lab*

Giant aliens have been spotted in the northern Gulf of Mexico. An Australian native normally six to eight inches across, the spotted jellyfish (*Phyllorhiza punctata*) is reaching diameters of two feet, apparently by feeding in the algae-rich waters of the Mississippi Sound. They may threaten the local shrimp industry—the second most valuable fishery in the U.S., behind that for Alaskan salmon. The crucial questions are whether the creatures will survive the winter and if they'll turn their attention to shrimp eggs and larvae. If they do, "their effect on the Gulf's environment and commercial fisheries could be one of the area's biggest problems next year," says Monty Graham of the Mississippi-Alabama Sea Grant Consortium, a group made up of eight local universities and research facilities.

—Steve Mirsky

CONSERVATION

Roadkill Genes

DNA blocker

FRED BRUEMMER *Peter Arnold, Inc.*

In ecological terms, motor vehicles do more than just flatten adventurous animals. Large highways act as effective genetic barriers, much the way rivers and mountains do, according to researchers at the University of Konstanz in Germany. Reporting in the August *Conservation Biology,* they note that vole populations living on each side of a four-lane highway are genetically distinct. Such limited DNA transfer makes small populations vulnerable to extinction. Roads also play a role in the worldwide decline in reptiles, according to researchers with the Savannah River Ecology Laboratory of the University of Georgia. Although the drop in amphibians is well documented, reptiles may be suffering more: according to World Conservation Union (IUCN) figures, 3.82 percent of the approximately 7,150 reptile species are extinct, endangered or vulnerable, compared with 2.75 percent of 4,680 amphibian species. The investigators, reporting in the August *Bioscience,* identify several factors but zoom in on human activities, such as habitat destruction and commercial trading.

—Philip Yam

"We may look alike, but because I'm a clone I'm actually someone's intellectual property."

SIDNEY HARRIS

PSYCHOLOGY

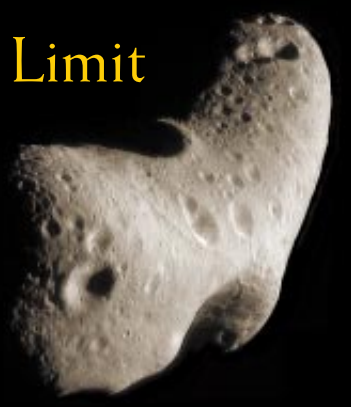
Down with E-reading?

Prefer reading *Scientific American* in print rather than at our Web site? Perhaps you find print versions more interesting and comprehensible. Those were the feelings of test subjects—even computer-savvy undergraduates—who read essays from another magazine, *Time*. On-line readers also found authors less credible and arguments less persuasive. The reason may be that students need to learn different processing abilities when they are attempting to read text on computer screens, speculates Ohio State University's P. Karen Murphy, co-author of the study. The findings were presented at a recent American Psychological Association meeting. —*S.M.*

DOUG MARTIN *Photo Researchers, Inc.*

DATA POINTS

The Sky's the Limit



Estimated minimum diameter of asteroid that would have global consequences if it hit the earth: 1 kilometer

Explosive yield: 100,000 megatons

Estimated number of such asteroids near the earth: 900

Percent so far detected: 40

Approximate size of object that leveled Tunguska forest in Siberia in 1908: 100 meters

That killed the dinosaurs: 10–15 kilometers

Largest identified terrestrial impact craters:

Crater Name	Location	Crater Diameter (kilometers)	Impact (millions of years ago)
Vredefort	South Africa	300	2,023
Sudbury	Canada	250	1,850
Chicxulub	Mexico	170	65
Manicougan	Canada	100	214
Popigai	Russia	100	35

SOURCES: NASA Ames; Science, June 23, 2000; National Weather Service; National Space Science Data Center

BIOLOGY

Stressed for Life

It's not news that early childhood trauma can exert lifelong emotional influences. But researchers from Emory University may now have found a physiological basis. In the August 2 *Journal of the American Medical Association,* they report that early-life abuse makes the brain hyperreactive to stress in later life. They studied four groups of women—depressed and nondepressed women with and without a history of childhood abuse—and then stressed them by having them speak and solve a mental math exercise in front of an audience. Levels of ACTH, a stress hormone, were higher in the abused women, most strikingly in those who were also depressed. Stress hypersensitivity due to childhood abuse may increase the risk for mood and anxiety disorders, and drugs interfering with the hormonal stress response might help prevent them. —*Julia Karow*

ATMOSPHERIC SCIENCE

Dingy Skies

The northern Pacific Ocean, once the only atmospheric area in the Northern Hemisphere untarnished by pollutants, now reportedly contains toxic aerosols originating from Asia. These results, announced at a July meeting, come from the University of California Pacific Rim Aerosol Network. The group sampled the air to detect aerosols containing pollutants, including minerals and radioactive lead isotopes, which can point to the country of origin. Asian countries are heavily dependent on coal-burning industries, and they emit more aerosols than any other region in the world. The minerals in the aerosols could interfere with the Pacific Ocean's heating and cooling cycles, severely perturbing weather patterns. —*Rebecca Lipsitz*


UNIVERSITY OF CALIFORNIA, DAVIS

GETTING THE DIRT: *Thomas Cahill of the University of California at Davis holds two filters: one clean and one exposed to air in Kyoto, Japan.*

ASTRONOMY

Mr. Spock, Phone Home

Researchers announced at the August gathering of the International Astronomical Union the discovery of nine new planets outside our solar system and several candidate planets, bringing the total to at least 50 since the first was discovered five years ago. One new find orbits Epsilon Eridani, only 10.5 light-years away—and one of the possible stars around which the home planet of *Star Trek*'s Mr. Spock circles, according to Treksperts. Some of the other new bodies belong to a family of planets—there's one new extrasolar system, with two low-mass planets, and five others are suspected. Extrasolar systems may therefore be common; previously, astronomers knew of only one, around Upsilon Andromedae. See http://obswww.unige.ch/~udry/planet/planet.html for more information. No word yet on the whereabouts of the Klingon sun. —*P.Y.*

BIOLOGIST_PAUL R. EHRLICH

Six Billion and Counting

Overpopulation hasn't brought humanity to its knees, but that doesn't mean people aren't overburdening the earth

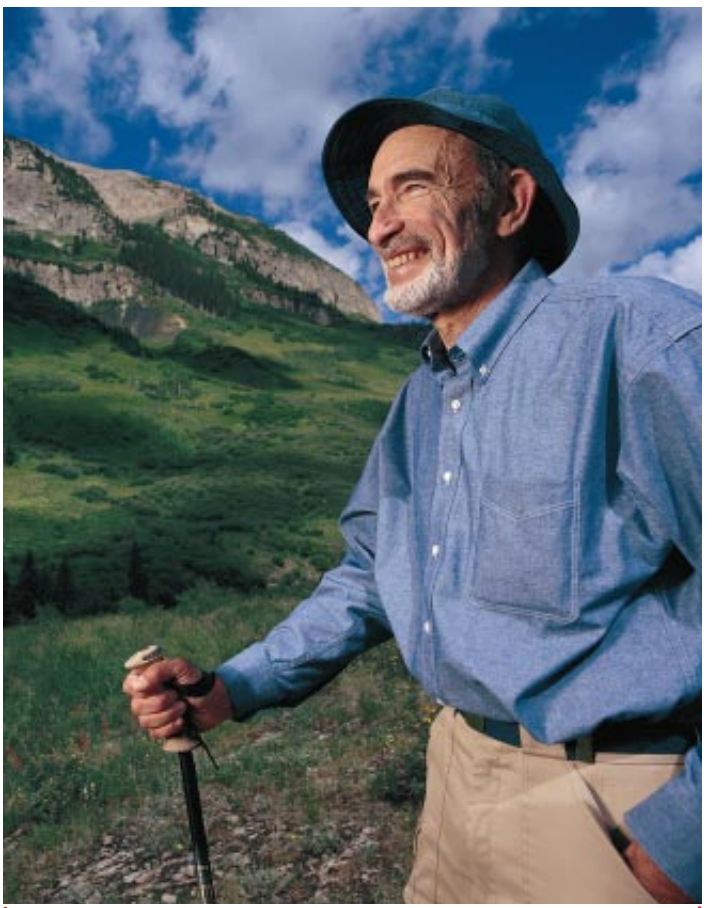
COPPER CREEK VALLEY, Colo.—One moment Paul R. Ehrlich and I are standing in the sunshine in a Rocky Mountain meadow strewn with wildflowers, discussing the mating habits of butterflies. The next we are hustling down a mountainside in a drenching downpour, dodging lightning strikes far too close for comfort. Inky black clouds had crept up on us without Ehrlich realizing. When the loquacious professor is deep in conversation, not much else matters, and besides, Ehrlich is used to being in the center of a storm.

For decades, the outspoken Stanford University biologist's warnings of imminent environmental collapse have landed him squarely in the middle of controversy. To his fiercely loyal friends, he is a bold crusader who has used his considerable charm and formidable verbal agility to bring an important if unpalatable message to the public. To his equally fierce enemies, he is a media hound and a Jeremiah who has been proved spectacularly wrong all too often yet refuses to admit his mistakes. His most trenchant critic, the late Julian L. Simon, a fervent proponent of the earth's capacity for limitless growth and a fellow at the conservative Cato Institute, repeatedly accused Ehrlich of leading a "juggernaut of environmentalist hysteria." Not known for pulling his punches, Ehrlich once said of Simon, "The one thing we'll never run out of is imbeciles."

Ehrlich crashed into the public consciousness in 1968 with a slim volume titled *The Population Bomb*. It predicted that overpopulation would lead to famine on a massive scale—and soon. Designed to influence the then upcoming presidential election (a notion that today makes Ehrlich roll his eyes at his naïveté), the Malthusian manifesto included possible scenarios for the coming chaos. One, envisioning a mystery viral pandemic of animal origins emerging from the shrinking jungles of overpopulated Africa, sounds eerily prescient. Another, foreseeing starvation in Japan and food rationing in the U.S., does not.

At the time, the book struck a loud chord and sold some three million copies. Ehrlich appeared on *The Tonight Show with Johnny Carson* and went public about his own vasectomy. His proselytizing helped galvanize international action on overpopulation and hunger and made him one of the world's most well known scientists. "He was the one who brought this problem to the attention of the world," says John Bongaarts, vice president of policy research at the Population Council in New York City. But, as Bongaarts acknowledges, "Paul Ehrlich's predictions of mass famines have not panned out."

Indeed, today, although too many people continue to go hungry, there are fewer now than when *The Population Bomb* was written. The Food and Agriculture Organization of the United Nations estimates that the number of chronically undernourished people in developing countries fell from 900 million in 1969–1971 to 800 million in 1988–1990. The figure hovers around 790 million in the latest estimates. While the world's population has boomed, hitting six billion, the number of malnourished as a percentage of the total has declined. The impassioned professor underestimated



STEVE STARR *SABA*

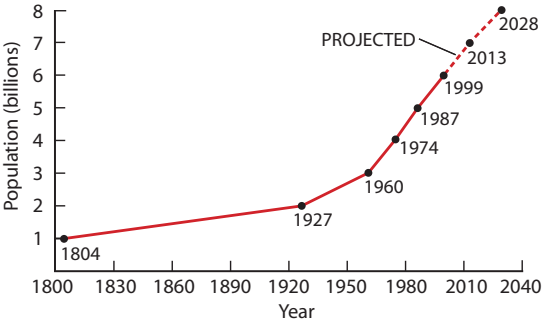
PAUL EHRLICH: FAST FACTS

- **Married to wife and colleague Anne for 46 years; one daughter, Lisa, and three grandchildren**
- **Calls his 93-year-old mother at least once a week**
- **Speaks Spanish and some French, German and Aivilikmiut Eskimo**
- **Became an atheist at age six**
- **Turned down medical school**
- **Serious wine buff; adores chocolate; devoted bird-watcher**

the impact of the "green revolution," says Per Pinstrup-Andersen of the International Food Policy Research Institute: "I don't think Paul Ehrlich understood—and I'm not sure that anybody understood—the potential of agricultural research in expanding productivity in agriculture."

That did not stop Ehrlich from moving beyond population growth to speak out on a broad range of environmental concerns—nuclear proliferation, biodiversity, pollution and global warming. His laissez-faire critics followed, however, and in 1980 he made himself an easy target. He and his like-minded friends accepted a now famous bet with free-market economist Simon that the price of five metals would increase by 1990 as a result of shortages. Instead they went down, and Simon won handily. Ehrlich argued in a feisty 1998 book, *Betrayal of Science and Reason,* that a recession in the first half of the decade had depressed prices, but the *New York Times,* in a seminal article on the wager, credited "entrepreneurship and technological improvements." Plastics replaced metals. Satellites and fiber-optic cables replaced copper wiring. Aluminum replaced tin.

Now, after 30 years of public life and strife, Ehrlich is unrepentant. Sure, his famous book had its faults, he acknowledges, but he counters, "Show me a scientist old enough to write something in 1968 who would still write the same thing today, and I'll show you an idiot." The bomb is still ticking, he says. "We're now in the middle of an explosion." And the deceleration in the growth rate doesn't give him much comfort. "Explosions don't accelerate forever—we seem to be past the peak but still growing very rapidly in historic terms," he notes. If speculation that population may hit a maximum at around 10 billion in 2050 is borne out, population will still be too high. "If that happens, we'll add many more people to the planet between now and then than existed when *The Population Bomb* was written, twice as many additional people as occupied the earth when I was born. And the average added person will be a much bigger threat to our life-support systems, a bigger consumer, than the average person in 1932."



SOURCE: United Nations Population Division

BOOM TIME: Some estimates peg the 10-billion mark as occurring around the year 2050 rather than in 2183.

Environmental alarm bells first rang for Ehrlich when, as a butterfly-collecting teenager in New Jersey, he witnessed developers destroying habitat and pesticides interfering with his efforts to raise caterpillars. While working on his doctorate at the University of Kansas, he studied evolutionary biology and the selection process whereby insects develop resistance to DDT. In 1959 he joined the faculty at Stanford, where he studied the ecology and evolution of local checkerspot butterfly populations and spent a decade trying to stop the university from subdividing the butterfly habitat for housing.

Butterflies also led Ehrlich to make significant contributions to evolutionary biology, particularly the theory of coevolution, which he developed with his friend and colleague Peter H. Raven, now director of the Missouri Botanical Garden in St. Louis. They studied the interrelationships between butterflies and the plants their caterpillars eat, finding that plants evolve to foil their predators and that butterflies evolve in turn to avoid plant defenses. This theory is now crucial to understanding the increasing resistance of insects to pesticides. His achievements include a host of prizes, election to the National Academy of Sciences and authorship or co-authorship of nearly 40 books.

Still, he is best known for his conservation efforts and their attendant controversies, although today the crusader seems tired of the pitched public battle and speaks of it as if it were a chess match he's grown bored with. "You think you have hammered a certain piece of nonsense into the ground, and then it just pops up again," he says. "You get a little weary of it, because it is just not interesting anymore."

At the age of 68, Ehrlich remains an unbowed six feet, two inches. The deep, sonorous voice that must have delighted *Tonight Show* producers still rumbles out from his rangy frame. Here at his summer workplace, the Rocky Mountain Biological Laboratory in Gothic, Colo., in the heart of mountains some 120 miles southwest of Denver, he's grown a beard that makes him resemble one of the Victorian-era miners who once worked the area. His large hands firmly grip the steering wheel of his Jeep as we climb the steep, grooved mud track to the cramped four-room cabin he has shared with wife, Anne, for the past 40 summers.

Nevertheless, age is catching up with Ehrlich, and he doesn't like it. "I can't beat the grad students up the mountain anymore. They have to wait for me. That's painful," he grimaces. His only concession to age as we walk up the hill behind his cabin are two hiking poles to stop him from slipping, and we set off at a steady pace. His take-no-prisoners wit and tidbits about the local environment spice the conversation. (The Western Fuels Association becomes the "Western Fools Association," his longstanding antagonist the Cato Institute is a "thoughtless tank," and presidential hopeful George W. Bush is dismissed as "George Shrub … this guy who is running for some office in our country.")

His latest work reflects a shifting interest from protesting against the damage humanity does to its environment, to asking why. In *Human Natures: Genes, Cultures, & the Human Prospect* Ehrlich returns to his academic training as an evolutionary biologist to argue against "extreme hereditary determinism," which he accuses evolutionary psychologists and popular writers of promulgating. Our cultural evolution plays a far greater role than our genes in determining behavior, Ehrlich writes, and we need to consciously direct that evolutionary process to our benefit. "I want to understand how we can change human cultural evolution to deal with our environmental problems," he explains.

In the book, Ehrlich alludes to his own fear of death. On a comfy couch back at the cabin after our wet walk, he confides that it's not so much that he fears death as that he regrets leaving his friends behind. And he won't be renouncing his long-held atheism. Offered the choice of "being tortured for eternity [in hell] or bored for eternity" in heaven, the rebel chooses neither. "No organized religion has ever presented me with anything that is remotely attractive as a reward," he maintains. To the end, Paul Ehrlich will stick to his convictions. *—Julie Lewis*

JULIE LEWIS is a freelance journalist based in Washington, D.C.

JENNIFER JOHANSEN

INTERNET_ANONYMITY

Speech without Accountability

New software makes it nearly impossible to remove illegal material from the Web—or to find out who put it there

SAN FRANCISCO—In the centuries-long struggle to decide what people may say without fear of prosecution, almost all the big decisions have been made by constitution writers, judges and politicians. When things work properly, these players balance one another out and change the limits of free speech only slowly and after much debate. Inventors have played an occasional starring role, too, Gutenberg being the archetype. But with the rise of the Internet, a certain class of inventors—computer scientists—has asserted its own special power to determine the boundaries of permissible speech. Unlike the leaders of governments, programmers release the new methods that they devise for sharing information globally, quickly and often with little thought to the consequences.

Consider Publius, a censor-resistant Web publishing system described in mid-August at a computer security conference in Denver. Engineers at the conference greeted the invention warmly, presenting to its creators—Marc E. Waldman, a Ph.D. student at **New York University**, and Aviel D. Rubin and Lorrie F. Cranor of **AT&T Labs-Research**—the award for best paper. Publius is indeed an impressive technical achievement: a tiny little program that, once widely installed, allows almost any computer user to publish a document on the Web in such a way that for all practical purposes it cannot be altered or removed without the author's consent, even by an incensed government. In fact, authors can post files to Publius that even they themselves cannot delete. Yet it is quite simple for any Web surfer anywhere to view files published this way.

The details of its design give Publius another important property. If publishers use an inexpensive anonymizing service (such as **Anonymizer.com, Rewebber.de** or **Freedom.net**) or a public Internet terminal to cover any tracks, then they can upload computer files—not just Web pages but also software and digital recordings—irremovably and with almost no risk of identification. The **Federal Bureau of Investigation** would not comment on how it might track down those who use Publius to put illegal material on-line.

CENSOR BEATERS: **Marc E. Waldman (*left*), Lorrie F. Cranor (*center*) and Aviel D. Rubin are the creators of Publius, an impressive tamperproof publishing system for the Web.**

Publius thus appears to allow speech without accountability, and that is something fundamentally new. Deep Throat was anonymous, for example, but the *Washington Post* still had to defend its Watergate story in court. When antiabortionists made up a list of doctors who performed abortions and posted it on-line, striking through the names of those physicians who had been murdered, they were hauled before a jury, which fined them $109 million in civil damages.

Every nation outlaws some kinds of communications: libel, piracy, conspiracy, treason. Some nations go much further, of course. "Governments are working on international moral censorship schemes," observes Michael Sims of the **Censorware Project**. "Companies are working on international economic censorship schemes." Publius, Sims says, is a response to this trend. "Many, many people don't want the Internet to end up looking like TV. When the censorship crosses their individual moral thresholds, some of them start to act in response."

But is it an appropriate response for a small number of computer scientists to create software that subverts the efforts of governments, who must answer to citizens, and of companies, who must answer to both governments and customers? Publius has many obviously good uses, Rubin argues. "A whistle-blower could use it to expose illegal dumping by his employer. You could set up a Web site supporting a political candidate that your boss hates. Or companies may want to back up their sensitive data—encrypted—on Publius so that it isn't destroyed in a disaster."

All true. But "there are much more direct ways to protect whistle-blowers, using laws instead of technology," says Joan E. Bertin, executive director of the **National Coalition Against Censorship**. The same is true for anonymous political Web sites and sensitive corporate data—and even for that list of abortion providers, which the judge did not pull from the Net. (The Supreme Court has ruled that threats are protected speech unless they are likely to cause "imminent lawless action.") If Pub-

PETER MURPHY

lius is used to commit crimes with impunity, governments may try to ban the system. Indeed, encryption laws in the U.K. already appear to forbid its installation there. Courts may uphold a ban, Bertin suggests, unless Publius clearly enables legal speech that cannot be protected in a more innocuous way. In any case, so long as Publius servers are running anywhere on the Internet and U.S. citizens can surf anonymously, any ban would have little effect.

Appropriate or not, Publius is now a part of the Internet. Even before he presented his paper in Denver, Waldman posted the source code to the program on his Web site, so that other experts can check it for holes. Within a week almost 40 servers in five countries (including the U.K.) were running the system. "But we haven't had any responses yet from countries under authoritarian regimes," Rubin says. Testing is expected to continue through October.

Ironically, Publius may be ineffective in the very places where censorship is most oppressive. Bennett Haselton of the Censorware Project points out that "it only protects against censorship on the publishing end. In a country like China, where the main problem is censorship on the receiving end (all inbound traffic is filtered through the 'Great Firewall of China'), it is trivial for the censors to detect when someone is accessing a Publius document." So Publius seems to work only for those who are already guaranteed a right to speak anonymously and read what they like. To them, it extends the ability, if not the right, to disregard what the politicians, judges and constitution writers have decided is out of bounds.

—W. Wayt Gibbs

How Publius Thwarts Censors

The Net interprets censorship as damage and routes around it," John Gilmore, a co-founder of the Electronic Frontier Foundation, famously quipped in 1993. But in truth Gilmore's aphorism has been more of an ambition than a description. Although there are many ways to place information on the Internet anonymously, every Web page and computer file resides on a server. Every server has an owner. And with a court order and armed agents, a government can find that owner.

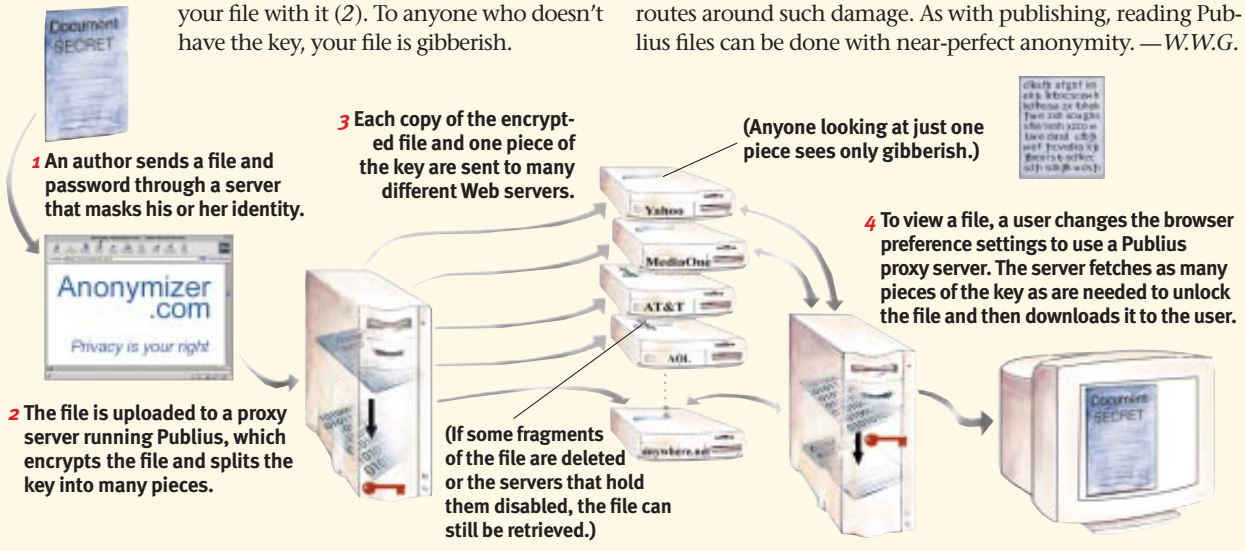
Publius makes Gilmore's idea practical by changing the way the Web works in three ways. First, it breaks that critical one-to-one correspondence between a Web file and a computer. Second, it denies a server owner (and government investigators) the ability to read every file on his machine. Third, it encodes each file so that any alteration to it can be detected.

The first step is the trickiest. Currently, a Web page is copied directly to the computer that will serve it up to the public. Publius adds a step in the middle. The software runs as a so-called proxy, either on the author's machine or on an intermediary computer. Suitably cloaked by an anonymizing service, you can upload a file (legal or not) to the proxy along with a password (*1*). The program then generates a random sequence of characters—a key—and encrypts your file with it (*2*). To anyone who doesn't have the key, your file is gibberish.

Now, using a 30-year-old bit of mathematical magic called secret sharing, the program breaks the key into lots of different pieces—say, 100 of them (but you could choose 500 or 1,000). It does this in such a way that with a fixed number of the pieces—say, 20—you can reconstruct the whole key and read the file. But 19 pieces offer no hints as to what the key might be. Any 20 of the 100 pieces will work.

Next, the proxy sends 100 copies of the encrypted file to 100 different servers, each accompanied by a different piece of the key (*3*). Because the file is encrypted and none of the server owners have the key, they can all plausibly deny knowledge of what files lie on their machines. Even if G-men comb the Web and shut down 80 of the servers on which your offending file rests (and good luck to them), it will remain available on the Web.

Publius also makes it easy for Web surfers to view the encrypted files (*4*). They simply change their browser to send "http:" requests through a Publius proxy. That computer intercepts requests for Publius files, fetches as many pieces of the key as are needed to unlock the file, and then decrypts and downloads it to the Web browser. With the key in hand, the proxy can detect any changes made to the encrypted file without the author's password. The system automatically routes around such damage. As with publishing, reading Publius files can be done with near-perfect anonymity. *—W.W.G.*



JENNIFER JOHANSEN

Octothorp Standard

Is a battle brewing between Sun's Java and Microsoft's C#?

LONDON—Just when you think the computer industry has neatly resolved itself into black hats (Microsoft, the Evil Empire) and white hats (everyone else), someone comes along and confuses things. I guess that makes sense, when you consider that some people get uncomfortable when their computers work too well—it means they're using old technology or something.

The confusion in this case began when Sun Microsystems announced that, after four years of touting Java as an open standard—that is, anyone can create applications with it without having to pay licensing fees—the firm was withdrawing the platform from the process of having it accepted as an international standard. Then, at the end of June, Microsoft announced that as part of its Microsoft .NET ("dot net") initiative—a next generation of Internet software and services—it would be launching a new computer language and submitting it to the standards-acceptance process. Sun: demanding control. Microsoft: giving up control. Has the world gone mad?

Probably not, although the folks at Slashdot, an on-line hacker community, did have a lot to say about Microsoft naming its computing language C#—everyone will argue about how to say "#" ("pound"? "hash"? "octothorp"?), except for musicians, who will pronounce it correctly at first sight as "sharp." The really dumb part of this name is that search engines don't recognize # as a character, so information about the language is nearly impossible to extract from the millions of hits you get on the letter C. This from a company that's supposed to have gotten into the Net in a big way.

But these gripes are mere mayflies. Java began life in 1995 as a hugely hyped bit of technology that was going to take over the Web. It is not just a computing language. Java is actually four separate things: a programming language; a virtual "machine" that runs Java programs ("applets") without need of additional hardware; a set of libraries that Java programs can draw on for frequently used routines; and a Java compiler. Java applets are often slow, badly written and buggy. Nevertheless, in many circumstances Java is extremely useful, and although it hasn't sidelined Windows, the number of Java applications keeps growing.

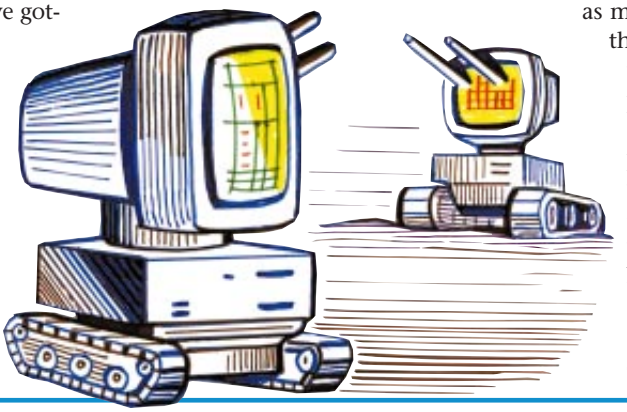
Microsoft responded by creating its own Active-X technology and by incorporating support for Java into its products. By October 1997 Sun was suing Microsoft, claiming that it was trying to splinter Java. Microsoft denied the charges, and the rest, as should be drearily familiar, is still pending in court. Meanwhile Sun declared its intention to get Java accepted by the International Standards Organization (ISO), first via a technical advisory committee known as JTC 1 and then, starting in April 1999, via ECMA (originally for European Computer Manufacturers' Association). Either body provides a route to ISO approval. Then, suddenly, in December, Sun announced that it was withdrawing Java from the standards-acceptance effort entirely, complaining that ECMA wanted to control both the standards process and the copyright of Java standards.

You can read this as evidence that Sun is just another evil company and never really meant to give away anything. Or you can read it as Sun protecting Java from a slow death by committee. Techies have done both, especially after Microsoft stated that it would be submitting C# to ECMA for standards acceptance. Of course, that's another action that can be read both ways. You could say Microsoft is trying to get itself some good publicity for a change. Or you could decide that Microsoft, after making its fortune all these years on proprietary standards like Windows, suddenly wants to make nice, perhaps in the hope that the U.S. Department of Justice is watching.

If the latter is the case, Microsoft probably shouldn't have announced .NET just days after Judge Thomas Penfield Jackson granted a stay against his order to break up the company while the appeal goes through. Of course, anything can happen as technology gets designed and deployed, but at first glance the initiative seems like the kind of Microsoft-owning-everything-everywhere world that the Justice Department was trying to prevent in its antitrust case. Java runs on all kinds of platforms; C# and .NET will run only on Windows. Java, since 1995, has been developed using the so-called Java Community Process, overseen by representatives from IBM, Hewlett-Packard, BEA Systems, Caldera, Apple and other companies. Java is a "de facto" standard, as Sun says, because it runs on pretty much everything. No amount of anointing by ECMA will make C# run on Linux.

For the paranoid, Microsoft .NET, whose announcement prattles cheerfully about renting software and storing all your personal information across the Net in a single location, sounds like the world that many feared when Microsoft began giving away Internet Explorer—a world where other browsers would be shut out because the main servers were all Microsoft beasts that spoke only or best to other Microsoft products. Would you give all your life's details to Microsoft? Would you like to rely on their servers to write a letter? Do you want an Internet as walled off as America Online?

Standards are the key to the widespread adoption of a new technology; that's why people fight over them so hard and so long. But the reasons something becomes a standard are many and varied and have as much to do with who is promoting the standard as with which committee accepts it, whether it meets people's needs, who supports it with ancillary products and when it appears. If a standards war erupts, pitting Java against C#, it will boil down to which company the techies trust more or distrust less: Sun or Microsoft. —*Wendy M. Grossman*



DAVID SUTER

WENDY M. GROSSMAN is a frequent contributor to this column.

TEENS IN JAPAN connect to the Internet with their iMode phones.

KAKU KURITA *Liaison Agency*

The Wireless Web

The Internet has been the subject of more self-parodying hype than anything since 500-channel cable. "The new economy offers synergies of disintermediation for the enterprise in the fast-paced B2B market." Whatever. Teasing out the truth can be tough. And it is getting tougher as the marketing machines gear up for the latest and greatest technomarvel, the wireless Web.

As this special report describes, the basic concept is simple enough: allow people to access the Internet from cellular phones, handheld computers and other portable gadgets. But it is neither an especially new idea—even in the U.S., which has been slower than other nations to catch on, such devices have been available since the mid-1990s—nor a sure sell. The data networks have been hobbled by incompatible standards, awkward user interfaces, punishingly high service charges and problems with spectrum allocation. They could yet go the way of, say, satellite telephone networks, the most famous of which, Iridium, has begun to fall to earth (literally).

That said, even a skeptic would give the wireless Web a good chance of success. Overheated sales rhetoric should not detract from the real engineering achievements. Spraying data through the airwaves isn't easy. Morse code used to be the basis of radio communication for good reason: sometimes only the simplest of signals can wade through the sea of static. Over the decades, partly through the efforts of amateur radio hobbyists, the reliability and speed of data transmissions have slowly but steadily improved. Current wireless networks send data at a rate of about 10 kilobits per second, about the speed of a regular modem circa 1990, and carriers are promising a 100-fold improvement in the next few years.

In the meantime, wireless enthusiasts (including a few *Scientific American* editors) can attest to the usefulness of checking movie times from the local bar and reading e-mail in airport waiting areas. Whether or not the breathless predictions come to pass, the wireless Web is slowly weaving itself into everyday life. —*Mark Alpert and George Musser, staff editors*

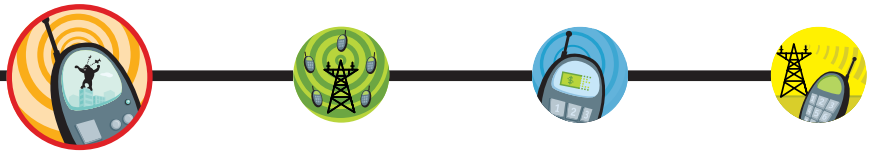
The Internet in Your Hands PAGE 40

The Promise and Perils of WAP PAGE 46

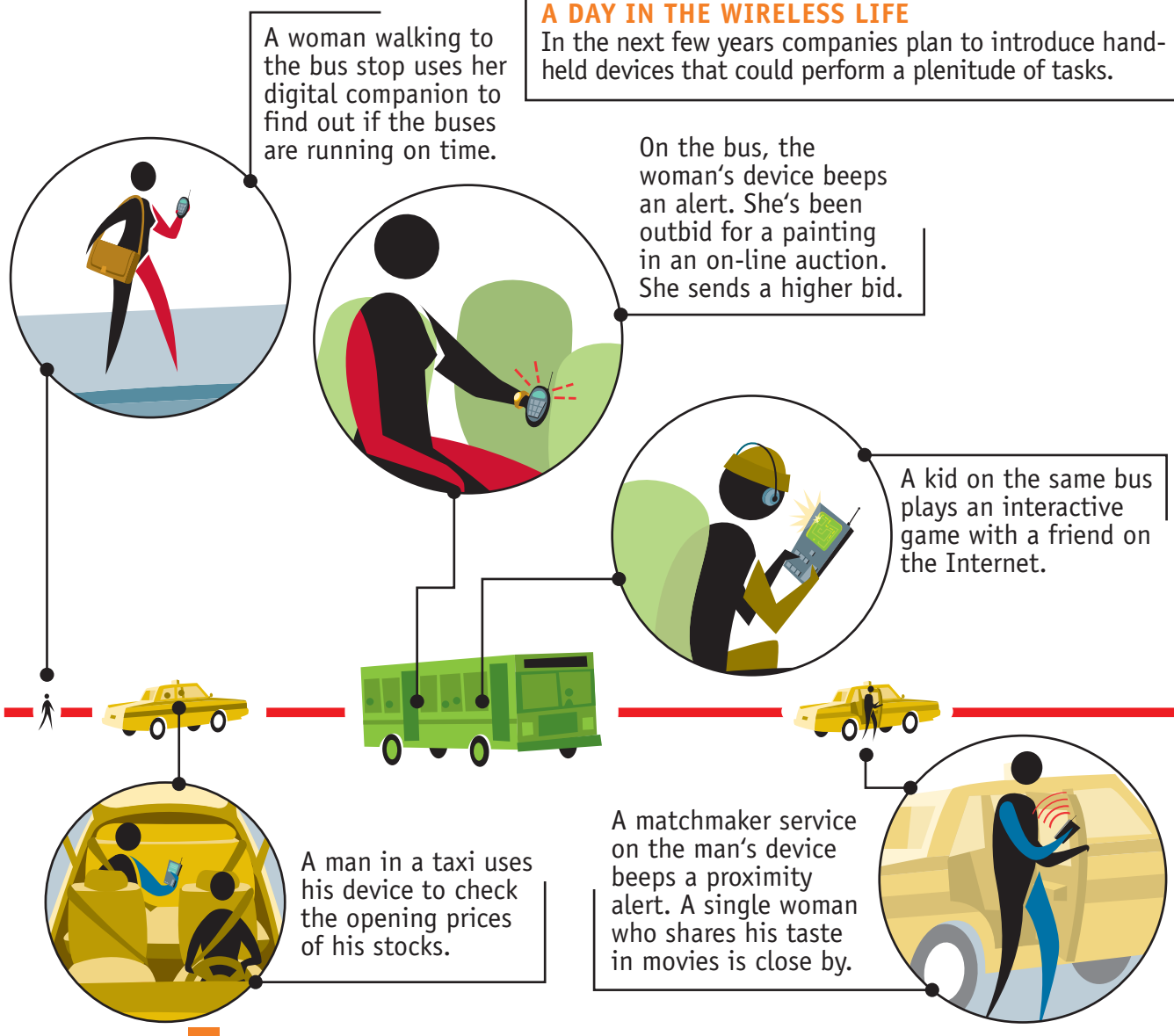
The Future Is Here. Or Is It? PAGE 50

The Third-Generation Gap PAGE 54

ILLUSTRATIONS BY XPLANE



A DAY IN THE WIRELESS LIFE
In the next few years companies plan to introduce handheld devices that could perform a plenitude of tasks.

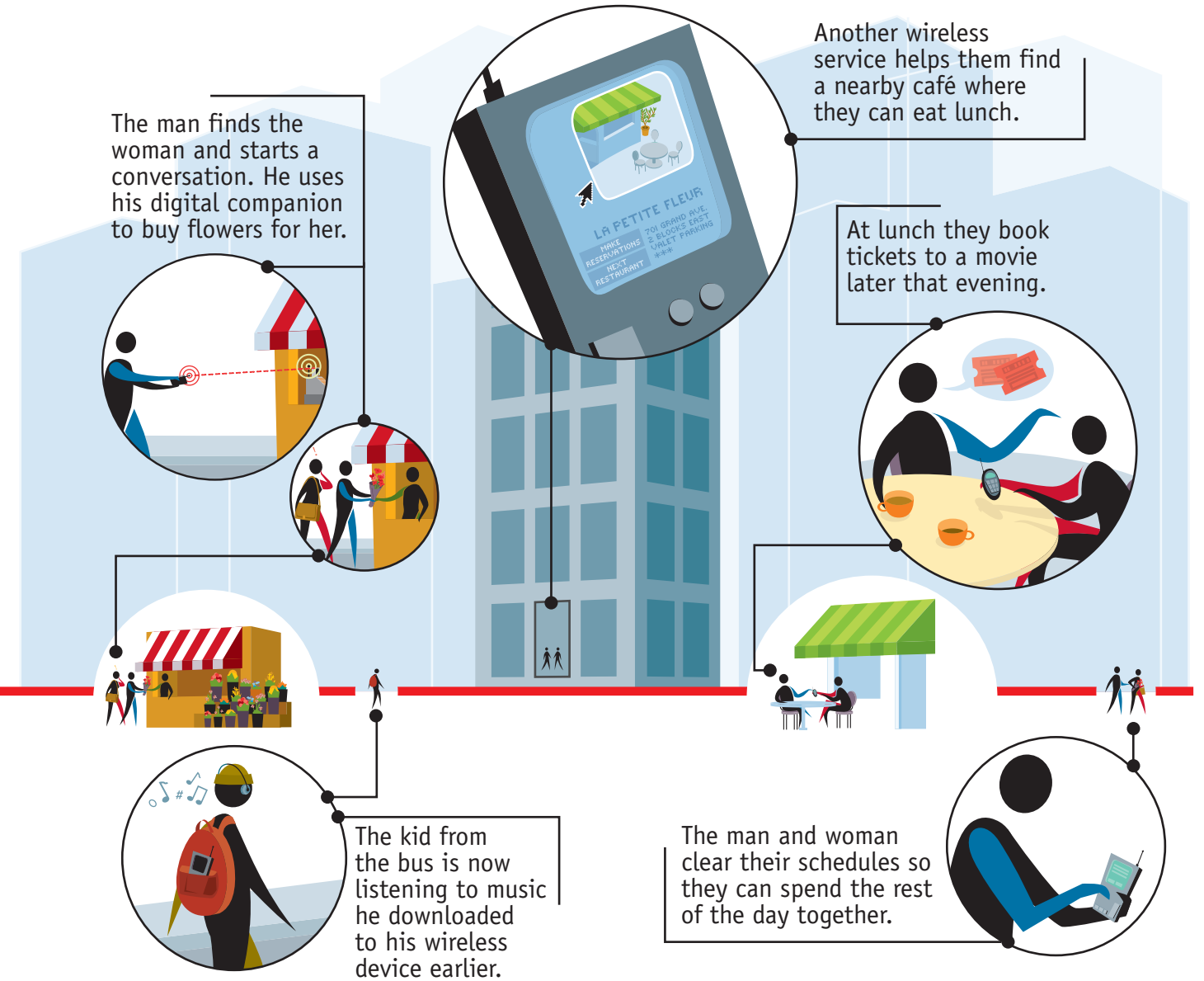


The Internet in Your Hands

To spur the growth of the wireless Web, companies are developing networks that can handle huge amounts of data and handheld devices that can tap into all the Internet's resources

by Fiona Harvey

ILLUSTRATION BY XPLANE



Imagine it's the year 2005, and you're in New York City on a business trip. Strolling past the garish stores on Fifth Avenue, you suddenly hear the distinctive beep of your digital companion. When you remove the device from the pocket of your coat, you find a distressing sight on its video screen: a news bulletin reporting a plunge in your company's stock price. Standing there on the sidewalk, you bark the words "Access personal finances!" into the digital companion's microphone. The gadget instantly calls up the Web page of your favorite on-line stock-trading site. As the other pedestrians stare at you curiously, you start selling shares to reduce your exposure. Then you instruct the digital companion to access an airline reservation service and buy a ticket on the next flight back to the company headquarters. You use the same device to call your wife and tell her about your change of plans. Finally, you order the machine to give you directions to the nearest bar. You're going to need a stiff drink before you head for the airport.

Although this scenario may seem fanciful, many of the biggest telecommunications companies are spending billions of dollars to make it real. What's motivating them is the enormous number of cell phone customers. In the U.S., some 95 million people use mobile phones, or about 34 percent of the population. In many countries in western Europe, the penetration of cell phones is above 50 percent; in Finland, the most cell phone–crazed nation on the planet, it is 71 percent. Even if just a small portion of these customers also subscribed to data services, it would represent a tremendous market.

The infrastructure for a wireless Web is already being constructed, although progress has been more rapid in Europe and Japan than in the U.S. In Finland, cell phone users can send text messages to their friends, pay their bills, get traffic reports and buy a cup of coffee at a Helsinki café simply by

Check It Out!
www.ou.edu/engineering/emc/standard.html

This site offers technical details on the frequency bands and modulation methods used by wireless Web networks in the U.S., Europe and Japan.

dialing the right numbers. In Japan, hip teenagers have gone wild for the Internet-connected iMode phone, which has attracted more than 10 million users. American companies have been slower to offer such services, primarily because the competing wireless networks in the U.S. employ different technologies for transmitting their signals. But in the past year, wireless carriers have launched aggressive campaigns to convince their customers to buy cell phones that can access the World Wide Web. In addition, new handheld computers such as the Palm VII have built-in antennas for wirelessly accessing the Web, and other models can do so by attaching to a mobile phone or a modem with an antenna.

Currently all these devices are restrained by the slow speeds of wireless data transmissions, which average about 10 kilobits per second—less than one fifth the data rate of a typical PC modem over a fixed telephone line. These sluggish speeds may be fine for simple requests such as sending text messages, but they can make surfing the Web a frustrating experience, and they practically rule out the wireless downloading of video, audio and other data-intensive files into a handheld device. What is more, many of the wireless data networks are now limited to major cities, making it impossible for users to go on the Internet when they travel outside the coverage area.

New wireless technologies, however, promise to remove some of these stumbling blocks. The first generation of cellular phones were analog devices that transmitted radio waves modulated to match the sound waves of the users' voices. Today's digital cell phones are the second generation: they convert the sound information into bits of data carried by modulated radio or microwaves. Telecommunications companies are now working on third-generation (3G) wireless networks that could increase the data rates of mobile devices to as much as two megabits per second—high enough to allow users to download songs and movies from the Web. In particular, mobile phone operators across the globe are developing a technology called Wideband Code Division Multiple Access, which raises transmission rates by spreading each wireless signal over a wide band of frequencies [see "The Third-Generation Gap," on page 54].

Licenses to use the sections of the radio spectrum necessary for 3G networks have already been allotted in the U.K., where the auction for the frequency bands raised nearly $35 billion for the government. The U.S. government is working on plans for a similar allocation. In the U.K., wireless carriers such as British Telecommunications, Vodafone and France Telecom are promising that 3G networks will be in place as early as 2002, but it is difficult to predict when the speedy data services will actually become widespread. People in rural communities may have to wait a longer time than big-city dwellers will. Forrester Research, a consulting firm in Cambridge, Mass., forecasts that 3G services will not be available outside major metropolitan areas before 2010.

In the meantime, some wireless carriers will roll out a variety of intermediate technologies, the so-called 2.5G networks. In the U.S., for example, Sprint PCS plans to raise the top speed of its Web applications to 144 kilobits per second by implementing packet-switching techniques. Almost all of today's wireless networks are based on circuit switching: when you make a connection, you are assigned a specific frequency channel until the end of the call. With packet switching, though, the signal is chopped into packets of data. Each packet is tagged with the address of the destination and then

Wireless Devices of the Future

TOMORROW'S GADGETS may look quite different from today's handheld devices. Lernout & Hauspie, a Belgian software company, has demonstrated a prototype called NAK (*left*) that allows users to dictate e-mails and issue voice commands. Nokia, a Finnish cell phone manufacturer, has developed concepts for phones that could display Web images and video (*right* and *opposite page*).

sent swirling through the network to find its own way on any available channel. At the receiving end, the packets are reassembled in the correct order. This system yields higher transmission rates than circuit-switching systems because data from many different sources can be intermingled on a single channel. Customers of packet-switched networks could talk on their cell phones and access the Internet at the same time. And because the signals would occupy a network's channels only when the phones are sending or receiving data, customers would pay for the volume of data that they downloaded rather than the time they spent on the network.

Unfortunately, this approach has a big drawback. Gathering and reassembling the data packets at their destination points takes a little time, and some packets get lost and have to be resent. This time lag is known as latency. It is not an issue with most data transmissions, where a delay of a few seconds can be tolerated, but losing bits of someone's voice can make conversation rather tiresome. Telecommunications companies have tried to minimize the problem with methods such as silence suppression, which removes the noiseless moments of voice communications. (About 30 percent of any phone conversation is silence.) The packets of data that do not contain voices are simply not sent, reducing traffic on the network and thus decreasing latency.

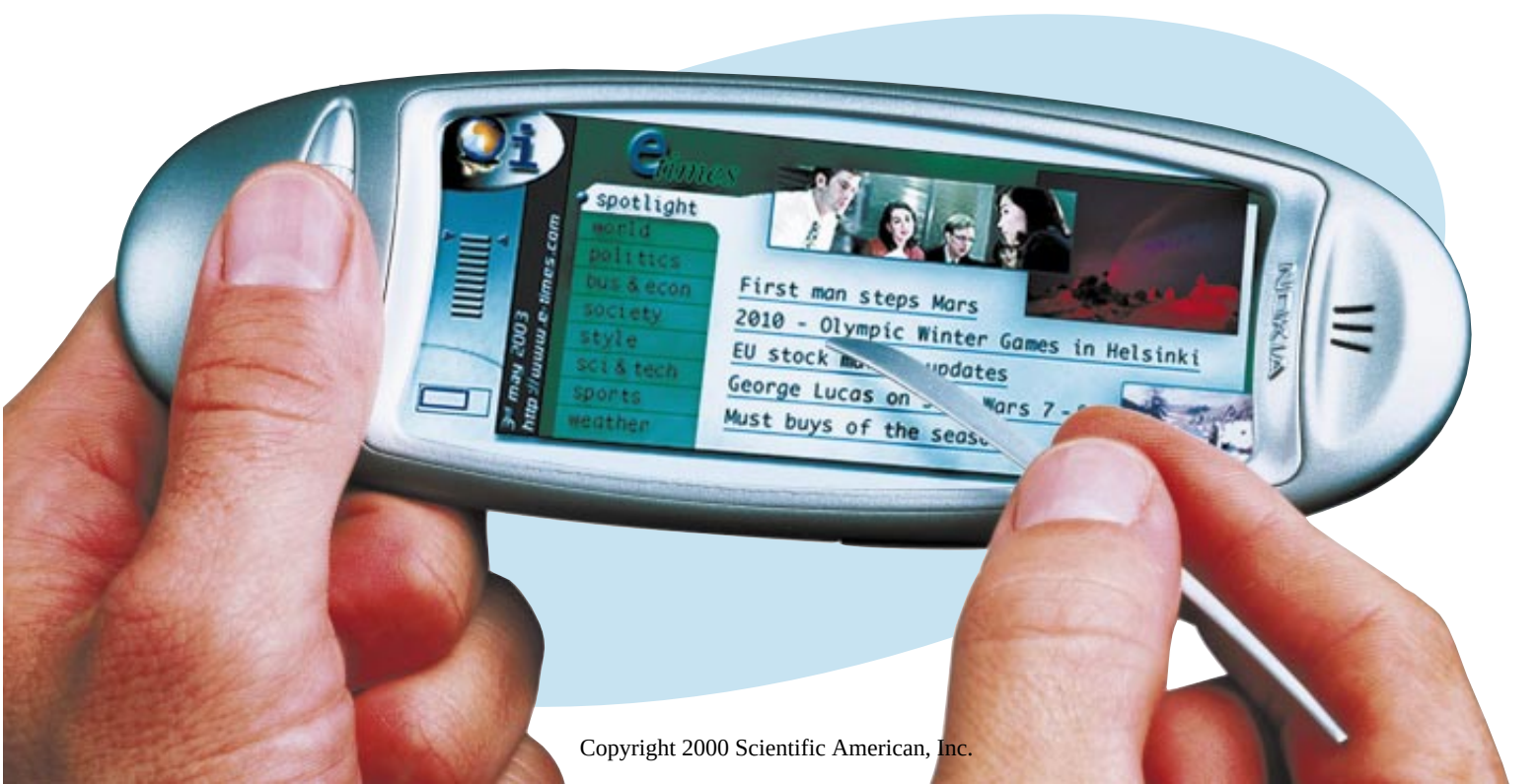
Super Phones

Cell phone manufacturers are experimenting with several different designs for the handheld devices that will be linked to the enhanced wireless networks of the future. If these machines are really destined to become digital companions, they will have to be versatile, adaptable and fashionable. At the high end of the product line, companies such as Nokia, Ericsson and Motorola are working on 3G "super phones" that will look nothing like existing cell phones. In fact, calling them phones seems absurd. They will have color screens, several inches square, for the presentation of high-resolution graphics and video. Some will have keyboards and miniature mice for data input, but most will use touch-sensitive screens and styluses, like those employed now by the Palm machines and other handheld computers.

In addition to carrying voice communications, the super phones will also be able play music files that are circulated on the Web in the popular MP3 format (or in whatever format may replace it). Indeed, this application could open up a huge new market for the devices. "I believe music will be the killer application for the mobile Internet," says Robert Madge, founder of Madge Networks, a British networking and Internet services firm. "The next generation of mobile appliances will take over from the Walkman." PC users who have spent hours trying to download music from the Web may scoff at this prediction. But a high-speed access service that is solely dedicated to music distribution could conceivably send highly compressed audio files to mobile devices when they are not being otherwise used (at night, say). The music files would then be stored in memory, perhaps on cards like Sony's Memory Sticks, until the user wanted them.

For customers who are inclined more toward the written word, there will be handheld devices with bigger screens for reading electronic books. Some companies have even developed prototypes equipped with tiny video cameras and projectors for showing movies downloaded from the Web. "If you're on a train with one of these, you could project a film on a fellow passenger who's wearing a white shirt," quips Lars Bergendahl, a program director for Ericsson. "After asking permission, of course."

Almost all the prototypes have earpieces that are separate from the body of the device, and many have headsets that hold both the earpiece and a microphone, so that users can speak, listen and see the device's screen at the same time. To eliminate the need for wires, many prototypes use low-power radio waves to transmit signals between the headset and the body of the device. Several companies have already incorporated short-range radio-wave devices—based on an industry standard called Bluetooth—into their existing product lines. This same technology could also transmit data between the digital companion and any nearby device, including cash registers at supermarkets and other stores, thus making it possi-



COURTESY OF LERNOUT & HAUSPIE AND NOKIA

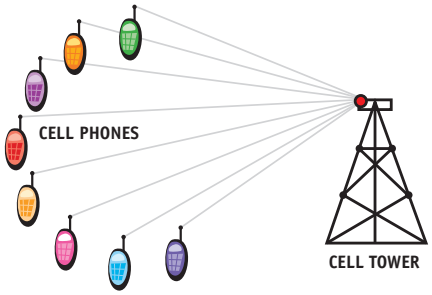
THE ABCs OF WIRELESS

THIS GLOSSARY DEFINES THE MOST COMMON ACRONYMS USED TO IDENTIFY WIRELESS TECHNOLOGIES. THE TERMS ARE GROUPED INTO THREE CATEGORIES REPRESENTING STAGES IN THE DEVELOPMENT OF THE NETWORKS.

FIRST GENERATION: ANALOG

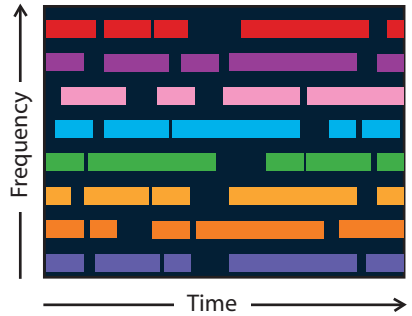
(Early 1980s–present)

The earliest cellular phones used analog technology, which modulates the phones' radio signals—varying their frequency in a continuous manner—so that they can convey the sounds of the users' voices.



AMPS (Advanced Mobile Phone System)

The most common type of analog cell phone network. Because analog phones are suited for voice rather than data communications, AMPS networks are rapidly being supplanted by digital wireless networks. Operates in the 800-megahertz frequency band and uses FDMA technology.



FREQUENCY DIVISION MULTIPLE ACCESS allows many cell phones to communicate with one base station (*left*). Each caller is assigned a different frequency channel (*above*). Breaks in the colored bars indicate that the channels are not in constant use.

FDMA (Frequency Division Multiple Access)

Technique that allows many cell phone users to communicate with one base station by assigning each user a different frequency channel [*see illustration at left*]. An AMPS network, for example, has 832 channels, spaced about 30 kilohertz apart. In digital networks, FDMA is used in conjunction with CDMA or TDMA.

SECOND GENERATION: DIGITAL

(Early 1990s–present)

Currently the most advanced cell phones employ digital technology, which converts the sounds of users' voices into streams of bits (1's and 0's) that are then used to modulate the wireless signals. Digital networks can also be used for data communications.

CDMA (Code Division Multiple Access)

Digital technique that enables cell phone users to share a frequency channel. It splits each wireless signal into many "chips" of data, each of which is tagged with the cell phone user's code [*see bottom illustration on opposite page*]. During transmission, the chips are spread over a band of frequencies, then reassembled at the receiving end.

ble for users to conduct transactions without cash, checks or credit cards.

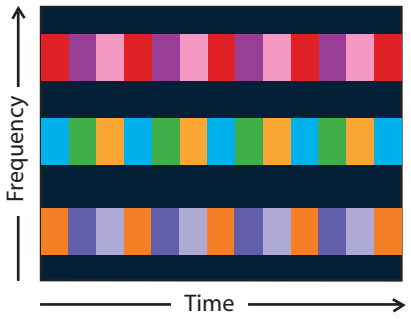
But for all the millions of dollars that are being poured into designing these super phones—a prototype can cost $2.5 million to build—analysts predict that the top-of-the-line devices will occupy only a small niche of the market. "Most people will not need full-motion video on the move," observes Lars Godell, an analyst at Forrester Research. He believes that simpler "smart phones" will be much more common. These devices will look more like today's cell phones but with slightly larger screens and better input options than today's alphanumeric keys. The capabilities of smart phones will most likely be limited to sending and receiving e-mail, accessing the Internet (probably through specially adapted portals like those used by today's Web-browsing cell phones) and perhaps playing downloaded music.

Phone companies are not the only ones eyeing the market. The makers of handheld computers, which have until now been used primarily for scheduling appointments and compiling address lists, are also introducing devices that can access the Internet. Some manufacturers are planning to add voice communications to the machines as well so that they can compete directly with the super phones. Even the makers of video-game consoles are joining the fray: Nintendo and Sony are expected to equip their GameBoy and Playstation consoles with Internet access in a year or two, according to Godell. And as more and more consumers get their music from the Web rather than from tapes or CDs, the manufacturers of portable music players will start to build models that can wirelessly download the MP3 files on the Internet.

Another interesting idea is the Internet watch. Hewlett-Packard has formed a collaboration with Swatch, the world's biggest manufacturer of watch components, to create a wireless device that will appeal to the fashion-conscious. Although it will look like a normal timepiece, the watch will contain a tiny radio transmitter that will enable the wearer to download sports scores, e-mail and music from the Web.

Software Struggles

While the hardware manufacturers battle it out, a similar competition will focus on the software that will control all these devices. One popular contender is the operating system developed by Palm for its handheld computers and also used in Handspring's Visor machines. Another is the EPOC operating system created by Psion, the British manufacturer of handheld computers, and also used in Web-browsing cell phones made by Nokia and Ericsson. The third player is Microsoft, which has realized the importance of portable platforms as shipments of desktop PCs have flattened out. In 1996 the company invented Windows CE, a pared-down version of its operating system for PCs, which is used in handheld computers made by Hewlett-Packard, Compaq and Casio. Microsoft is now developing software specifically designed for smart phones and has joined with manufacturers such as Ericsson to create mobile devices that will be customized for use in particular businesses—for instance, digital companions with special appli-



TIME DIVISION MULTIPLE ACCESS allows many users to share a frequency channel. Each wireless call is assigned a repeating time slot (*colored bands*) in the channel.

GSM (Global System for Mobile Communications)
The European standard for digital networks, which guarantees the compatibility of wireless devices—a German cell phone can be used on a French network, and vice versa. Uses TDMA technology and operates in the 900-megahertz frequency band.

PCS (Personal Communications Service)
A type of digital wireless network in North America that operates in the 1,900-megahertz frequency band. Example: Sprint PCS, which uses CDMA technology.

TDMA (Time Division Multiple Access)
Another digital technique for channel sharing, which assigns each cell phone user a repeating time slot in a frequency channel [*see illustration at left*]. Because the data from each user always appears in the same time slot, the receiver can separate the signals.

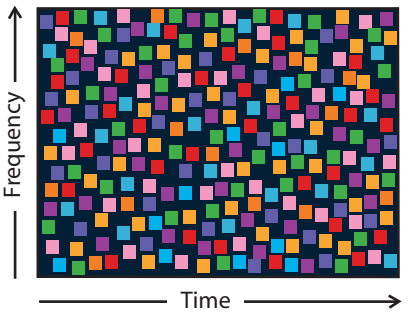
THIRD GENERATION: BROADBAND DIGITAL

(Expected in early 2000s)
Wireless companies are promising to revamp their networks so they can handle broadband data communications, with transmission rates much higher than the current average of 10 kilobits per second.

EDGE (Enhanced Data Rates for Global Evolution)
An upgrade of TDMA-based networks that could achieve data transmission rates as high as 384 kilobits per second. Currently being pursued by GSM networks in Europe and AT&T Wireless in the U.S.

GPRS (General Packet Radio Service)
A planned improvement for GSM networks that implements packet switching for data communications. Instead of sending data on dedicated circuits, a packet-switching network divides the information into packets and transmits them on any of the network's available channels.

W-CDMA (Wideband Code Division Multiple Access)
A refinement of CDMA technology that could raise data transmission rates up to two megabits per second. W-CDMA spreads the chips of the wireless signal over a much wider band of frequencies than CDMA does.



CODE DIVISION MULTIPLE ACCESS breaks each wireless signal into many chips of data (*colored squares*). The chips are transmitted over a band of frequencies and then reassembled at the receiving end.

XPLANE (*illustration*); JENNIFER JOHANSEN (*graphs*)

cations for lawyers, doctors or children. Because of the diversity of devices, most industry watchers agree that there is unlikely to be a single winner in this software war, in contrast with Microsoft's domination of PC operating systems.

As technological advances continue to miniaturize batteries and electronics, the size of handheld devices will come to depend primarily on their screens and input systems, which must be large enough to be seen and manipulated by users. But what if the user could interface with the device by voice command instead of by screen or keyboard? To make this possible, the machines would need software that could understand and replicate human speech. Microsoft has set a goal of developing an operating system that can cope with voice commands, and IBM and Philips are pursuing similar projects. Lernout & Hauspie, a Belgian voice-recognition software company, recently acquired two rival firms, Dragon and Dictaphone, to accelerate its own development efforts.

Unfortunately, progress in this area has been disappointingly slow. Although the quality of dictation software has climbed impressively—some systems can accurately transcribe between 95 and 99 percent of the words dictated to them—even the most advanced computers still struggle with basic language tasks. The problem is that the software for voice command is not yet sophisticated enough to catch the nuances of human speech [see "Talking with Your Computer," by Victor Zue; Scientific American, August 1999].

Lernout & Hauspie, however, has developed an intermediate technology that could allow a limited number of voice commands. The software is code-named NAK (the name comes from the Hawaiian word *nakulu,* meaning "echo"), and it is designed for mobile devices running on Intel's StrongARM chips, a class of powerful and energy-efficient microprocessors. A NAK device would use a speech synthesizer to read the text of an incoming e-mail aloud. The user could then dictate a response into the device's microphone and ship it off with a voice command such as "Send e-mail." Lernout & Hauspie claims that NAK will have a vocabulary of more than 30,000 words, a wide range of possible commands and the capability of understanding more than one language. Products based on the technology are expected to cost about $500 and will ship in volume early next year.

It is becoming clear that today's panoply of portable electronics gadgets will eventually give way to a new order of mobile devices that will combine the functions of their forerunners. Deluxe all-in-one digital companions will provide full access to all the resources on the Web, including graphics, video and music. Other wireless devices will be more specialized, such as Walkman-like music players that can download songs from the Internet. The most generally useful functions—accessing the Web, probably through customized portals, and sending e-mails—will be built into most portable devices, because it will be cheap, easy and expedient to do so. Otherwise, users would have to carry a Walkman *and* a mobile phone. And in the near future that will seem as quaint as wearing a monocle.

FIONA HARVEY, a journalist based in London, frequently writes about wireless technologies and the telecommunications industry.

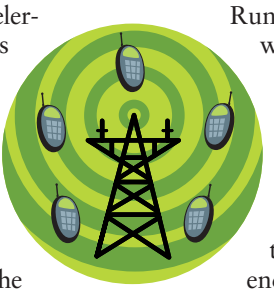
The Promise and Perils of WAP

The Wireless Application Protocol allows cell phone users to connect to the Internet, but the technology has serious limitations

by Karen J. Bannan

Every great journey begins with a first step. Many leaders in the telecommunications industry are convinced that the first step toward a wireless Web is a set of technical specifications called the Wireless Application Protocol, or WAP.

In recent years the leading wireless companies have introduced data networks that allow cell phone users to pull information from the World Wide Web and display it on the phones' tiny screens. The WAP specifications would essentially standardize how the networks transmit Web documents to cell phones, pagers and other handheld devices. Wireless carriers in Europe are currently implementing the WAP standards in their data networks, and some U.S. carriers—including Sprint PCS and Nextel—promise to do so in the near future. WAP's proponents say that the transition will accelerate the growth of the wireless Web. But critics counter that the WAP specifications are inadequate because they severely limit the user's access to the Internet. It is still unclear whether WAP will truly usher in a golden age of wireless data or turn out to be just a false start.



The Wireless Application Protocol evolved from technologies developed by Phone.com, a software company based in Redwood City, Calif. The great obstacle to wireless Web access is the fact that cellular phones and their networks are not robust enough to handle HyperText Markup Language (HTML), the lingua franca of the Internet. The publishers of Web sites use HTML to weave text and graphics into simple, easy-to-navigate documents. Current cell phone networks, however, have a low bandwidth: they relay data much more slowly than fixed-line networks do, making it difficult if not impossible to transmit Web pictures wirelessly. And today's cell phones do not have the processing power or the display screens needed to show complex images.

To get around these problems, the software engineers at Phone.com created Handheld Device Markup Language (HDML), which was specifically designed for wireless networks. HDML allows the text portions of Web pages to be transmitted to cell phones and other mobile devices. Many of the wireless data networks in the U.S. incorporated Phone.com's software into their phones and network servers. What is more, Phone.com allied with three cell phone manufacturers—Motorola, Nokia and Ericsson—to devise a standardized language based on HDML. The result was Wireless Markup Language (WML), which became the core of the WAP specifications. The companies also formed an industry group called the WAP Forum to promote the new standards. The group now has more than 530 members.

Here's how a WAP network taps into the Web: first, the owner of a WAP-enabled phone uses the device's microbrowser—a pared-down version of the browser software used by PCs to navigate the Internet—to request a specific Web page [*see illustration on opposite page*]. The request goes over the airwaves to a cellular transmission tower and is then routed to a server operated by the wireless carrier. Running on the server is the WAP gateway: software that serves as a filter between the wireless network and the Internet. The gateway finds the Web page requested by the cell phone user. If the page is written in HTML (as most Web pages are), coding software converts the document to WML, stripping away the page's graphics and any specialized formatting of the text, such as elegant fonts. The WAP gateway encodes the WML translation so that it can be transmitted wirelessly, and then the document is sent to the user's cell phone, where it appears on the device's small screen.

Conversion from HTML to WML, however, is usually not trouble-free. If a Web page uses pictures as links to other pages, for example, the page has to be rewritten to provide text links instead. In practice, this conversion frequently limits the amount of accessible information and occasionally makes the Web page completely unreadable. For this reason, many Web publishers have created separate versions of their pages written in WML and expressly tailored for WAP devices. For example, MapQuest, MSNBC.com and Go2Online.com have created WAP versions of their sites. WAP users can find a list of more than 5,000 WAP-friendly sites by visiting www.cellmania.com.

With the help of these portals, cell phone users can call up

ILLUSTRATIONS BY XPLANE

sports scores, airline flight schedules or the latest book bargains from Amazon.com. But the range of services and information that can be accessed by cellular phones is only a small fraction of the wealth of data available on the Internet. Some analysts believe that the WAP standards will swiftly become obsolete as cell phone technology advances. "WAP was designed for the low-bandwidth cellular networks of the 1990s," says Rich Luhr, an analyst with Herschel Shostech, a technology consulting firm in Wheaton, Md. "And it was designed for phones that had no graphics capabilities and screens with only two to four lines of text." According to Luhr, as wireless networks and devices improve, WAP's raison d'être will disappear.

The companies that are supporting WAP, however, say that the specifications can be revised to suit future technologies such as the high-bandwidth third-generation networks now being developed [see "The Third-Generation Gap," on page 54]. Furthermore, they emphasize that WAP-compatible phones are not intended as substitutes for Web-browsing PCs. "WAP isn't about browsing the Internet," says Skip Speaks, general manager of Ericsson's Network Operators Group. "It's about delivering unique content that's optimized for wireless devices." Adds Scott Goldman, the chief executive officer of the WAP Forum: "Using or accessing the Internet from a wireless device is a different experience than accessing it from a PC. I use an analogy that accessing the Internet from a PC is like going to an all-you-can-eat buffet. You see a broad range of foods, pick and choose what you want, and move quickly from one food to another, putting as much as you want on your plate. WAP is more like room service. You see a menu, order what you want, and it's delivered to you."

Spurring Internet companies to produce more content for WAP phones is a high priority for the WAP Forum. But according to some analysts, creating a WAP-compatible Web page is more difficult than setting up a typical Internet page, because WML is harder to learn than HTML. The WAP Forum's Goldman says that WML developer tools similar to those used for HTML will become more prevalent as WAP gains in popularity. Even so, the doubts about WAP's future have made many Web developers leery: they do not want to invest the effort in creating separate pages for cell phone users if the standards are likely to become outdated. "No one wants to jump in and commit to anything until things settle down," says Herb Williams of Spyglass, a software company based in Naperville, Ill. "There's a lot of negativism about WAP, and that limits the amount of content available."

Check It Out!
www.wapforum.org

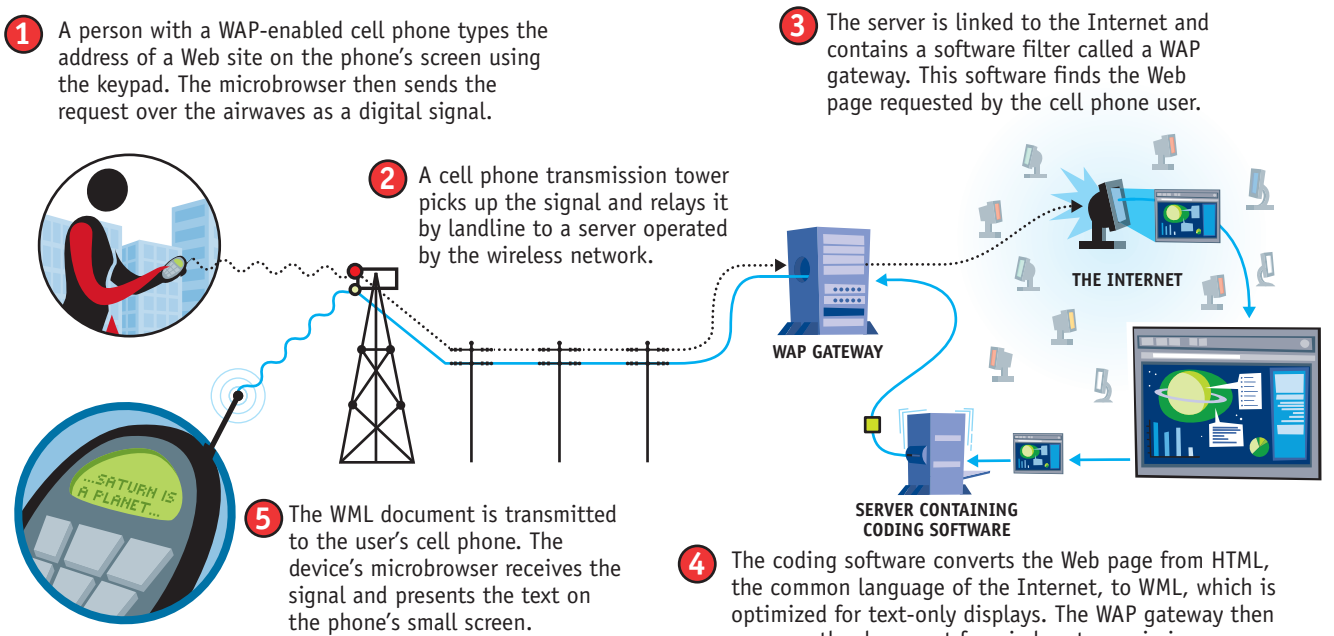
The WAP Forum's Web site details the technical specifications of the Wireless Application Protocol.

WAP's Weak Point

WAP has another hurdle to overcome: security. The current version of the standards includes a set of provisions called Wireless Transport Layer Security (WTLS), which specifies how to encrypt wireless data while they are in transit from the cell phone to the network operator. The WTLS techniques require less power and memory than the Secure Sockets Layer technology that is used to protect credit-card numbers and other sensitive information on the Internet. The system's weak point, however, is the server that runs the network's WAP gateway, where the data must be decrypted from the wireless coding and reencrypted using

How WAP Gets You on the Web

The Wireless Application Protocol is a set of standards specifying how cell phone users can access the World Wide Web.



Lost in the Translation

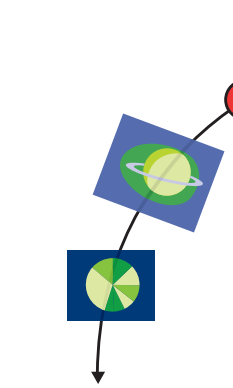
WAP networks have to strip down Web pages to make their content suitable for the small screens of wireless devices.

1 A typical Web page written in HTML may include text, graphics, animations and audio files. Pictures are often used as links to other pages, and the text may be presented in elegant fonts. The page may also be broken up into frames to make it easier for users to navigate the site.

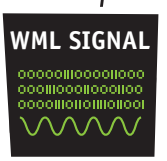
HTML WEB PAGE

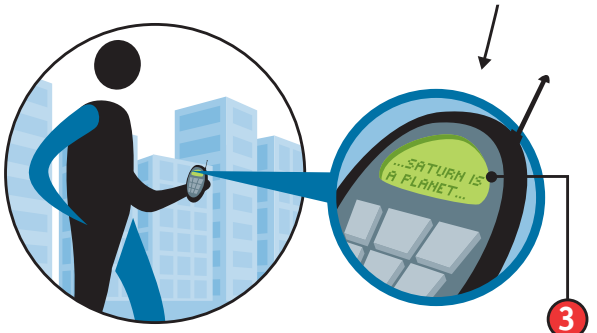
SERVER CONTAINING CODING SOFTWARE

2 When the coding software converts the Web page from HTML to WML, it removes everything that cannot be presented on a cell phone's screen: graphics, animations, audio files, frames and so on.



WML SIGNAL





3 After the cell phone receives the WML document, the Web page's text is displayed on the phone's screen. The words are presented in a standard font, and the Web page's picture links appear as text links.

Internet coding. For a fraction of a second (the exact time depends on a network's latency and speed) the cell phone user's private information is unencrypted. Granted, this moment of vulnerability occurs on servers that are closely guarded by the wireless carriers. But critics insist that even a split second of exposure is too much. Goldman asserts that WAP's security problem is a nonissue. "It's like a one-inch hole in a 100-foot wall," he says.

As if all these problems were not enough, a patent dispute also clouds WAP's prospects. Geoworks, a company based in Alameda, Calif., that develops software for wireless communications, holds a patent for the user interface incorporated into WAP phones and is seeking a $20,000 annual licensing fee from any large company that uses the technology. Phone.com filed a suit against Geoworks this past April, challenging the validity of the patent. Two months later Geoworks filed a countersuit, saying that Phone.com's activities infringe on Geoworks's patent. Although the dispute may seem like mere posturing from two rival companies, the lawsuits and the uncertainty over licensing fees may have a chilling effect, says Eddie Hold of Current Analysis, a technology research firm in Sterling, Va. "The development of WAP is at the same stage as the development of the Internet in the early 1990s," Hold says. "If the companies that developed content for the Internet had been required to pay a licensing fee, it's quite possible there would not be an Internet today."

Furthermore, WAP is not the only game in town. Some wireless networks in the U.S. may decide to stick with their current HDML-based microbrowsers and gateways rather than switch to the standardized WAP software. Technologically, there is not much difference between WAP and HDML. WAP's advantage is that it is an open standard rather than a proprietary system; the operator of a WAP network can buy its microbrowsers and gateways from a variety of vendors instead of relying solely on Phone.com. Switching from HDML to WAP, however, has its costs. A network operator must replace any equipment that is not compatible with the WAP standards, including older cell phones that work only with HDML.

Some analysts are certain that U.S. carriers will eventually rally around WAP as the European carriers have done. But the bigger question is whether wireless customers will actually use the data services that WAP provides [see "The Future Is Here. Or Is It?" on page 50]. Many observers believe that WAP networks will not become popular unless the wireless carriers charge no more for data services than they do for voice communications. "WAP content should be used as a point of leverage to garner and retain voice customers," Luhr says.

In the end, WAP's future may hinge on the attitudes of its promoters. The current wireless networks are not as open as the Internet; the carriers and hardware manufacturers are controlling, to a large degree, the kinds of data available on their phones. But the explosive growth of the Internet was a direct result of its openness. The medium became so popular because any business could create its own site on the Web with a minimum of effort and expense. Many analysts believe that the wireless data networks must follow this model to succeed. Says Luhr: "Today's WAP is about control, so the wireless industry can tell people what to do and what to look at. WAP must evolve beyond that."

KAREN J. BANNAN is a freelancer who writes and edits for the New York Times, *the* Wall Street Journal, Internet World *and* PC World, *among other publications.*

U.S. Allocation of Radio Spectrum

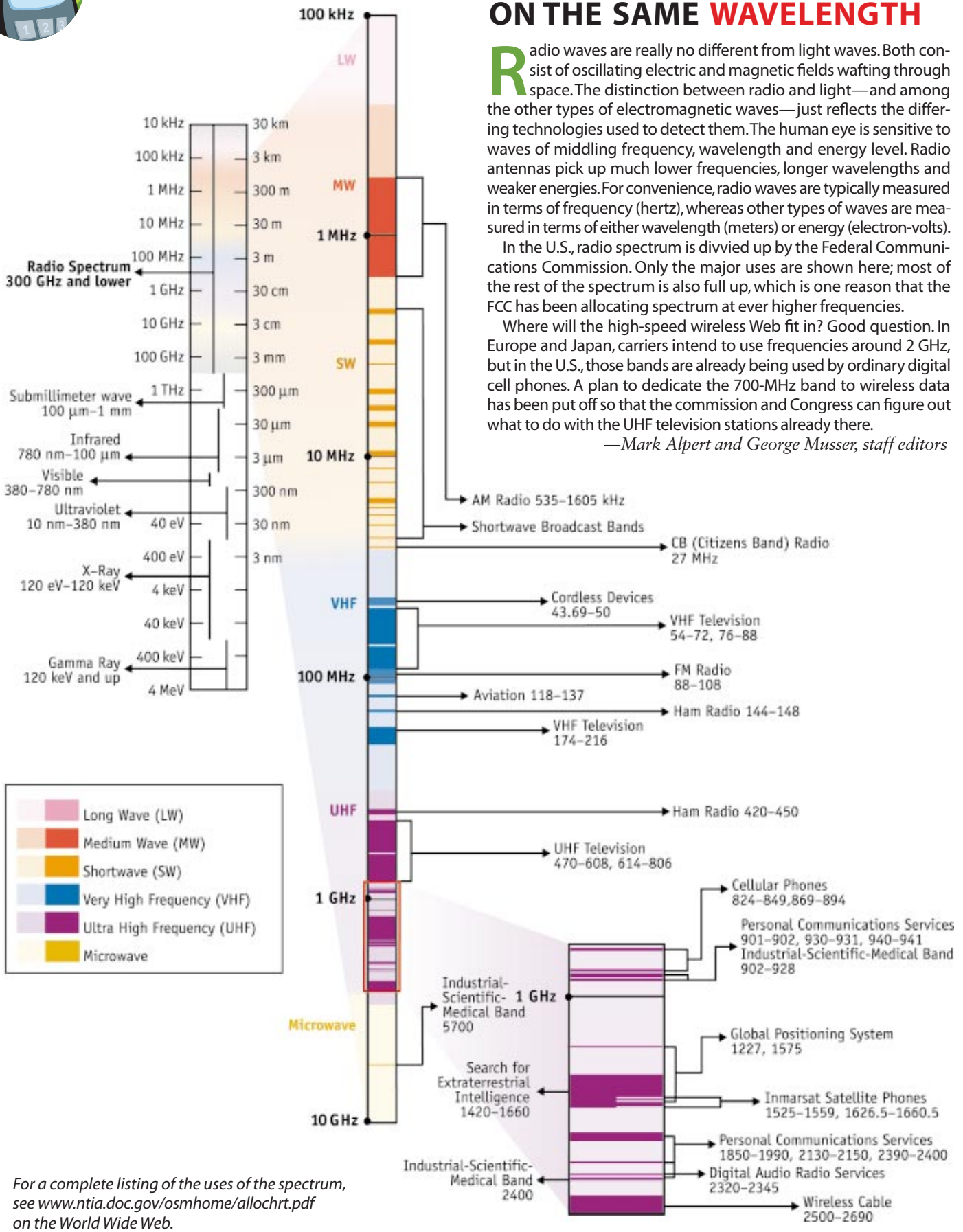
ON THE SAME WAVELENGTH

Radio waves are really no different from light waves. Both consist of oscillating electric and magnetic fields wafting through space. The distinction between radio and light—and among the other types of electromagnetic waves—just reflects the differing technologies used to detect them. The human eye is sensitive to waves of middling frequency, wavelength and energy level. Radio antennas pick up much lower frequencies, longer wavelengths and weaker energies. For convenience, radio waves are typically measured in terms of frequency (hertz), whereas other types of waves are measured in terms of either wavelength (meters) or energy (electron-volts).

In the U.S., radio spectrum is divvied up by the Federal Communications Commission. Only the major uses are shown here; most of the rest of the spectrum is also full up, which is one reason that the FCC has been allocating spectrum at ever higher frequencies.

Where will the high-speed wireless Web fit in? Good question. In Europe and Japan, carriers intend to use frequencies around 2 GHz, but in the U.S., those bands are already being used by ordinary digital cell phones. A plan to dedicate the 700-MHz band to wireless data has been put off so that the commission and Congress can figure out what to do with the UHF television stations already there.

—Mark Alpert and George Musser, staff editors



For a complete listing of the uses of the spectrum, see www.ntia.doc.gov/osmhome/allochrt.pdf on the World Wide Web.

ILLUSTRATION BY JENNIFER JOHANSEN

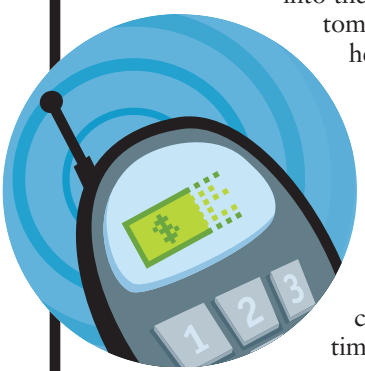
The Future Is Here. Or Is It?

How will Web phones ever become popular if it takes 10 minutes and costs $4 to send one e-mail?

by David Wilson

If you want to see the future, watch a teenager in Japan. For young Japanese, the cell phone call—that phenomenon of modern living—is already going the way of 45-rpm vinyl. Phones aren't just for calling; they're for sending e-mail. Since its introduction in February 1999 the Internet-ready iMode phone has been taken up by some 10 million Japanese. It has proved so popular that the carrier, NTT DoCoMo, is now Japan's largest Internet service provider. "All my friends have them," says 18-year-old Aya Shimizu. "We use them all the time to stay in touch."

In the U.S. the acceptance of Web phones is growing rather more slowly. Cell companies have built "microbrowsers" into their latest phones, allowing their customers to send e-mail and check news headlines. Those technophiles who use Web phones swear by them. When Edward Learned, marketing director for an Internet service provider in Minneapolis, gets lost, he simply enters his location and the phone gives him directions. Thumb twiddling is a thing of his past. When he's standing in line, he checks his e-mail, looks up movie times and tracks his portfolio. "I can even pull up a stock chart," he marvels.



But it's still not clear that Americans will embrace Web phones with the same enthusiasm as the Japanese or the Scandinavians. Industry analysts are divided over the long-term viability of the Wireless Application Protocol, in particular. Typing out e-mails using the keypad code is a real chore; newer phones are incorporating ways to make it easier, such as guessing the word you're trying to spell, but it's still much slower than a full keyboard. Worse, customers quickly find that small screens and meager services can cost big bucks.

"The truth is, what you're getting isn't worth what it costs," says Simon Buckingham of Mobile Lifestreams LLP, a British consulting group. "The real problem with the technology right now is that it doesn't reflect the rhetoric. There is a huge gap between what people have been promised and what you can actually do today."

Thy Web Is So Great, My Phone So Small

For many customers, the first hurdle is simply the idea that you can navigate a Web site on that teensy cell-phone screen. Unconvinced? You can try it yourself using your desktop computer to simulate the Web phone experience. Of course, of the vast ocean of information available on the Internet, only a trickle is available through a Web phone. The sites you use most often when sitting at your desk will, as likely as not, overwhelm a phone.

The ones that do work must be specially tailored. Although a few companies such as Google have developed technology that converts a standard Web page into a Web phone page on the fly, most sites prefer to build a kind of parallel Web site that Web phone users can access. The number of such sites is increasing by the week. Pinpoint.com in Durham, N.C., runs an Internet search engine that seeks out and tracks phone-friendly Web pages. This past April it found 150,000 of them; by June, 2.4 million. Company president Jud Bowman estimates that 25,000 of the five million Internet domains—each of which corresponds to an individual Web site—offer Web-phone-optimized pages. The vast majority are in Europe and Japan.

In short, a Web phone is no surfboard. It can be used to get snippets of information from certain sites such as MapQuest, Amazon and CNN, but it is not a general-purpose browser.

ILLUSTRATIONS BY XPLANE

Check It Out! www.gelon.net/#wapalizer

Visit this site to see what the Web would look like from a cell phone.

Almost all carriers restrict users to a small subset of the phone-optimized Web pages. Is that enough? For many consumers, yes. Those who want bigger screens, easier text entry and broader access, at the price of carrying an extra device, can turn to handheld or laptop computers. The Palm VII model comes equipped with wireless Internet access, and radio modems such as OmniSky's Minstrel can be clipped onto other handhelds. For laptops, Metricom has just rolled out a new version of its Ricochet wireless modem, promising speeds twice as fast as the best ordinary modems and eight times faster than cell phone modems.

But even the most patient early adopters get frustrated with the spotty coverage, glacial speeds and long latency. If anything, the service in many areas has been deteriorating over time. "Some of the services [that] people have set up are not robust, not scalable," Buckingham explains. "They work fine for a few dozen people, but if a few hundred people try to use them they break or work very slowly." The much-touted third-generation networks may also run into such difficulties [see "The Third-Generation Gap," on page 54].

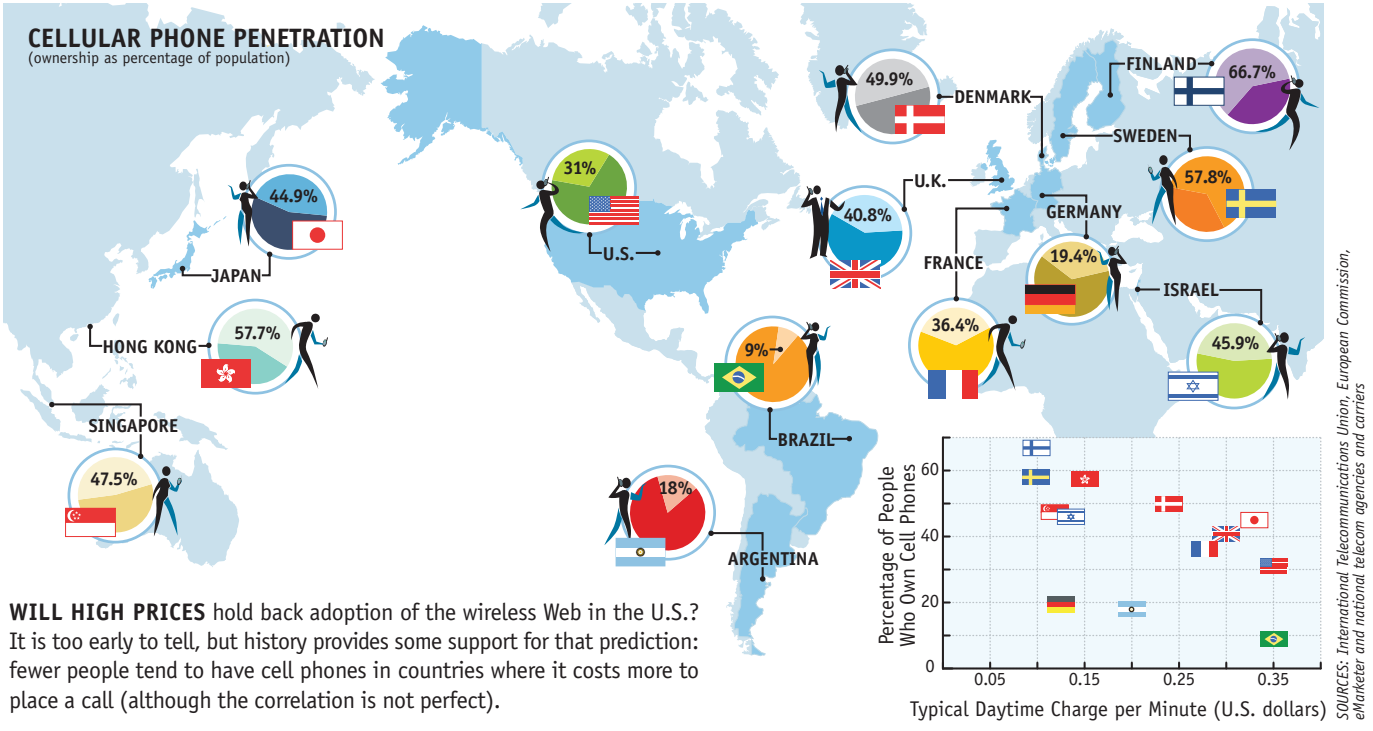
When delay strikes, the mismatch of cost and capability becomes seriously irksome. U.S. carriers charge as much as 39 cents per minute, and European carriers are similarly pricey. The clock ticks even when your fingers cramp up or the World Wide Web turns into the World Wide Wait. It's disconcertingly easy to lose track of how much money you're spending. "We've been hearing stories about bill shock," Buckingham says. "Somebody looks up a sports score and checks the headlines and then finds out that cost them 10 bucks." Although some carriers now offer flat-rate plans, these, too, tend to be more expensive than equivalent Internet access over an ordinary phone line.

In Japan the pricing model is less punitive. Making a brief call on a cell phone runs about 20 cents; sending an e-mail message, about a penny. NTT DoCoMo makes its money in other ways. The company estimates that the average iMode user spends an average of about $13 a month on Internet-related services in addition to telephone time. The most popular feature is a monthly download of cartoon characters from a company called Bandai. The files cost about $1 apiece and are used as screen savers. But even in Japan, open-ended costs frighten some consumers away. "I think it would be very useful, but I'm worried about how much using all those features would cost," says Ida Kohei, a former television executive who is, admittedly, 72 years old.

To be sure, these are probably just the early teething pains of the industry. Businesses everywhere see the potential of the technology; connected consumers will be big, mobile wallets. The plan is that you'll use your phone to spend money everywhere, all the time. Riding the elevator and hear a song you love? Whip out the phone and order the album. Found the perfect sweater but not sure about the price? Use your phone to see if you can get a better deal across the street or on the Internet. Thirsty? Access a list of nearby coffeehouses. But until cell phone companies get their act together, the dream will remain just a dream.

DAVID WILSON, who writes for the San Jose Mercury News, *has covered the Internet since 1991. His motto: He who laughs last thinks slowest.*

Cellular Usage vs. Prices



WILL HIGH PRICES hold back adoption of the wireless Web in the U.S.? It is too early to tell, but history provides some support for that prediction: fewer people tend to have cell phones in countries where it costs more to place a call (although the correlation is not perfect).

SOURCES: International Telecommunications Union, European Commission, eMarketer and national telecom agencies and carriers

The Third-Generation Gap

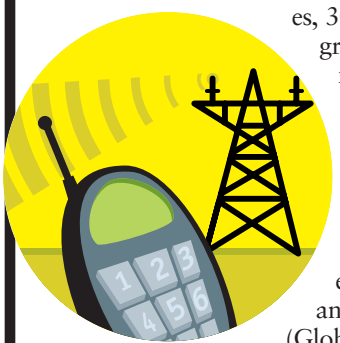
A revolution needs a plan. Which technology will provide it?

by Leander Kahney

Just wait until cell phone networks go high-speed. It will start to happen later this year, as carriers in Japan begin to deploy so-called 3G, or third-generation wireless cell phone systems. Spreading from east to west, the nimble networks should arrive in Europe in 2002 and the U.S. in 2003. Unlike the previous two generations of cellular networks, 3G systems have been designed from the get-go to carry data as well as voice. Carriers promise downloads approaching 2.4 megabits per second (Mbps)—twice as fast as wired broadband services, and fast enough to bombard cell phones, handhelds and laptops with video, music and games.

Or so they say. But there is a growing chorus warning that 3G will not be all it's cracked up to be.

3G is not a single standard or technology but an umbrella term for a variety of approaches to bringing high-speed Internet services to cell phone networks. In most cases, 3G will come from updates and upgrades to current systems, which differ from continent to continent and from country to country. Most 3G networks will start off as hybrids, with new capabilities added gradually as demand dictates.



The result is an alphabet soup worthy of a convocation of rocket scientists. In general, Europe and Asia will convert from GSM (Global Standard for Mobile communications), whose widespread adoption has given them the lead in wireless technology, to W-CDMA (Wideband Code Division Multiple Access). In North America, CDMA (Code Division Multiple Access) networks, such as Sprint's and GTE's, will also migrate to W-CDMA. But TDMA (Time Division Multiple Access) systems, such as AT&T's and Southwestern Bell's, plan to go to EDGE (Enhanced Data rates for Global Evolution).

These systems are still mostly in an experimental or testing stage, and each has its advantages and disadvantages. EDGE requires relatively minor infrastructure upgrades, but its theoretical maximum data rate of 384 kilobits per second (kbps) pales when compared with W-CDMA's much faster 2 Mbps.

W-CDMA and CDMA are based on a technology known as spread spectrum. Older cellular technologies such as GSM and TDMA use a variant of the approach taken by ordinary radio stations—namely, they divide the radio spectrum into narrow frequency bands. To add capacity, these networks can interleave several phone calls on each frequency channel, but there is a tight limit to how many users can share a channel before the signal quality suffers. CDMA, on the other hand, assigns each phone call a particular code. Multiple radio signals can then share a fairly wide range of radio frequencies. Each phone will pick up the transmissions intended for it by watching for its code. In some implementations of spread spectrum, the transmitter and receiver hopscotch among frequencies in a prearranged sequence [see "Spread-Spectrum Radio," by David R. Hughes and Dewayne Hendricks; Scientific American, April 1998].

Although spread-spectrum systems have their inefficiencies—with all the overhead to determine which messages are going to which phone, they tend to use a lot more bandwidth than the signals alone require—they are very tolerant of noise and are difficult to intercept or interfere with. CDMA uses channels 1.25 megahertz (MHz) wide in the 800-MHz or 1.9-gigahertz (Ghz) bands. W-CDMA channels are 5, 10, 15 or 20 MHz wide in several bands located around 2 GHz, which allows for faster data rates and more users.

In the Clouds

Yet these are not the only available technologies—or even, critics say, the best ones. One of the most vocal naysayers is Martin Cooper, who is widely credited with inventing the cell phone for Motorola in the early 1970s. The 3G networks, he says, will offer just over 1 Mbps when all the overhead is taken into account. But that's not 1 Mbps per user: the bandwidth will be shared among everyone in a particular cell (the geographical area covered by a single cell tower), which could be dozens of people at a time on each channel. Cooper says users should expect 64 kbps from 3G networks at best, a privilege for which they will pay a handsome premium. Although quite an improvement on current wireless networks, it is only marginally faster than an ordinary modem and hardly enough to justify all the futuristic claims made for the networks.

For Cooper, 3G is a baby step toward real high-speed, inexpensive wireless communications. He is now at ArrayComm, a

ILLUSTRATIONS BY XPLANE

BLUE IN TOOTH AND CLAW

THEY'VE FINALLY COME UP WITH A WAY TO GET RID OF THE TANGLE OF CABLES

Harald Bluetooth was a fierce Viking king who made history by uniting Denmark and Norway by force of arms in the 10th century. The king has now lent his name to a technology that may make history in the world of electronics: a short-range wireless system that spells a long-overdue death for computer wires and cables.

Created by a consortium of mobile-phone manufacturers and silicon chipmakers, Bluetooth is a "wireless wire" that allows computing and telecommunication devices to be connected without cables. It was originally conceived as a way of making wireless headsets for cell phones but has been rapidly adopted across the electronics industry. Although the first Bluetooth devices are only just appearing, it is already becoming the de facto standard, with more than 1,000 manufacturers committed to making Bluetooth-enabled devices, from laptops and cell phones to toys and refrigerators.

The technology has a maximum range of about 10 meters (30 feet) and operates in the 2.4–2.5-gigahertz industrial-scientific-medical (ISM) band, in which low-power radio transmitters are allowed to operate without first getting a government license. To avoid interfering with other devices, Bluetooth hops around frequencies at a rate of 1,600 times per second.

As well as replacing cables, Bluetooth will make it easy to set up wireless networks at home or in the office and in public places such as airports and libraries. Bluetooth will also allow devices to swap information with each other whenever they come within range, allowing conference attendees to trade business cards or workers to synchronize handheld organizers as they walk into the office. Many cell phones, handhelds and laptops have infrared ports for these purposes, but their range is typically limited to one meter or so.

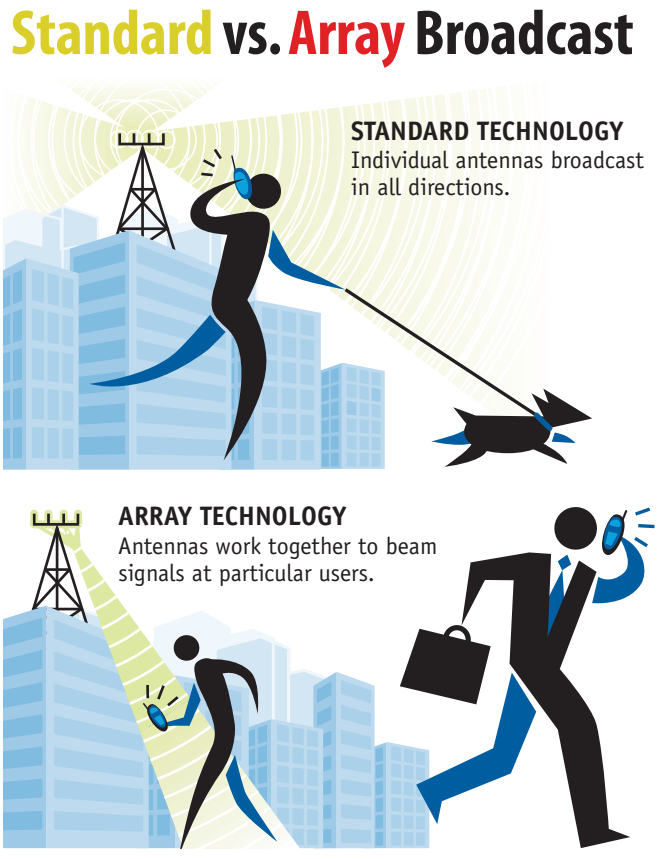
The industry expects Bluetooth to become ubiquitous, but its rate of adoption will depend on how often electronic devices are replaced. Although people may buy new cell phones and computers every year or two, washing machines and microwaves are on a much longer replacement cycle. —*L.K.*

San Jose, Calif., start-up working on "smart antennas," which, he claims, could provide 1 Mbps for each of up to 40 concurrent users. The technology makes better use of the arrays of antennas found in cellular base stations. As you may have noticed when driving by a cell tower, each station contains a forest of up to a dozen antennas. Currently they are used to broadcast omnidirectionally—that is, with equal strength in all directions.

But many communications and radar systems have long used similar arrays to aim their signals in particular directions. The transmissions from individual antennas interact with one another, preventing the signals from going in some directions and amplifying them in others. Cooper proposes retrofitting cellular base stations to the same end.

His system is based on digital signal processors originally developed by the U.S. military for spying on foreign radio broadcasts. Such signal processors, when attached to an antenna array, can beam radio signals precisely at individual users. As each user moves around, the smart antennas track them. The result is a kind of cloud of radio signals that follows each user around like the cloud of dust around Pigpen. The system can reuse the same radio frequencies for different users in the same vicinity, without worrying that the transmissions will interfere with one another. The result is very efficient use of the carrier's spectrum, which affords the high data rates.

The antennas are already in place, and most cellular base stations have signal processors with the necessary computational power. So in most cases a software upgrade is all that is required to turn them into smart antennas. The drawback is that high data rates come at the expense of movement. Al-



Check It Out!

www.avren.com/Courses/TX_RX_Architectures_plain.htm Visit this site for more information on how radio works.

though the system is able to track a walking subject, it currently can't keep pace with a fast-moving vehicle. ArrayComm plans to begin wide-scale tests soon and has teamed up with Sony to deliver video, music and games over the airwaves in San Diego.

•• — • — — — •••

A radically different approach is being taken by inventor Larry W. Fullerton, who has spent the past two decades working in obscurity on a potentially revolutionary technology known as UWB (ultrawideband). Most radio transmissions have two components: a carrier wave and a signal. The carrier wave is the vehicle; it is the frequency to which you tune a radio. The signal is the passenger; it comes from a microphone, TV camera or Internet connection and is imprinted onto the carrier wave in a process known as modulation. The most common style of modulation, FM (frequency modulation), causes the carrier to spread out by an amount roughly equal to the data rate of the signal. A 10,000-bps message, for instance, causes the carrier to "smear" by 10 kHz on each side. This is why radio stations have to be spaced apart in frequency.

Spread-spectrum radio, used in the most advanced cell networks today, essentially switches among many different carriers for a given transmission. But UWB, first devised in the 1960s, dispenses with the carrier altogether. It is pure signal. In essence, a switch attached to the antenna turns on and off, which produces a pulse of electromagnetic energy—rather like the pop you hear on the radio when turning on a lamp. In Fullerton's systems, the pulses last less than a billionth of a second each and occur up to 40 million times per second. Like an ultrafast Morse code, the pulses occur in a very particular pattern, which can encode the desired information.

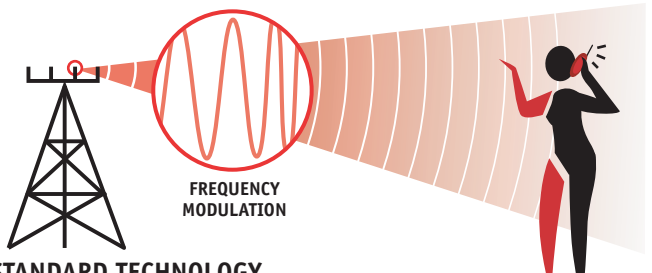
One implication of UWB sounds utterly crazy: rather than take up a small slice of the radio spectrum, as other technologies do, it uses the whole thing. Typically the pulses carry energy from 1 to 3 GHz. Fortunately, that doesn't lock out other radio systems. To most radio receivers, the UWB signals sound like random static and can be filtered out as long as their power remains low. Only receivers that know the pattern of pulses can recognize and decode the signals. Different UWB transmitters can use different patterns, allowing many to operate at once without interfering with one another.

Fullerton is now chief technology officer of Time Domain, a Huntsville, Ala., firm created to commercialize the technology. The firm hopes to push the data rates even higher. "Our engineers—with a straight face—tell me we can get a gigabit per second," says Ralph Petroff, the company's president. Several other firms, such as Multispectral Solutions in Gaithersburg, Md., have also been working on the technology. Until this spring the Federal Communications Commission (FCC) had licensed UWB only for limited experiments, but in May it gave the go-ahead for much wider tests.

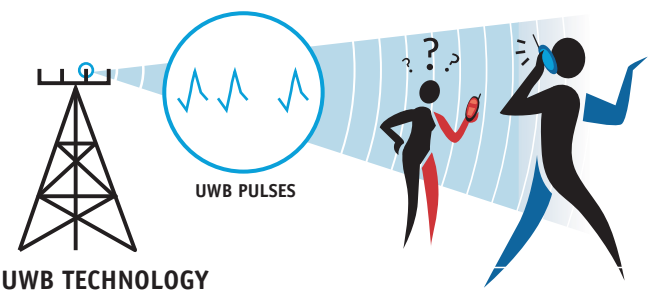
UWB has a wide variety of potential uses, from personal radar systems for detecting collisions to imaging devices that can see through walls. But will it ever provide high-bandwidth wireless communication? To keep the signals from interfering with other radio devices, UWB broadcasts at extraordinarily low power—50 millionths of a watt. Trouble is, low power means low range—just a few meters. The more power, the farther it could reach, but the greater the chance it would interfere with radios, televisions and Global Positioning Satellite receivers. Petroff says UWB will initially be confined to indoor local area networks, a kind of Bluetooth on steroids [*see box on preceding page*], but may one day be used for neighborhoodwide networks.

"I think there's going to be some kind of power restriction from the FCC that will restrict its range," comments Bob Scholtz, a professor of electrical engineering at the University of Southern California. "But we don't know what that will be. It could be hundreds of yards."

Standard vs. UWB



STANDARD TECHNOLOGY
Cellular signals piggyback on a carrier wave of a particular frequency.

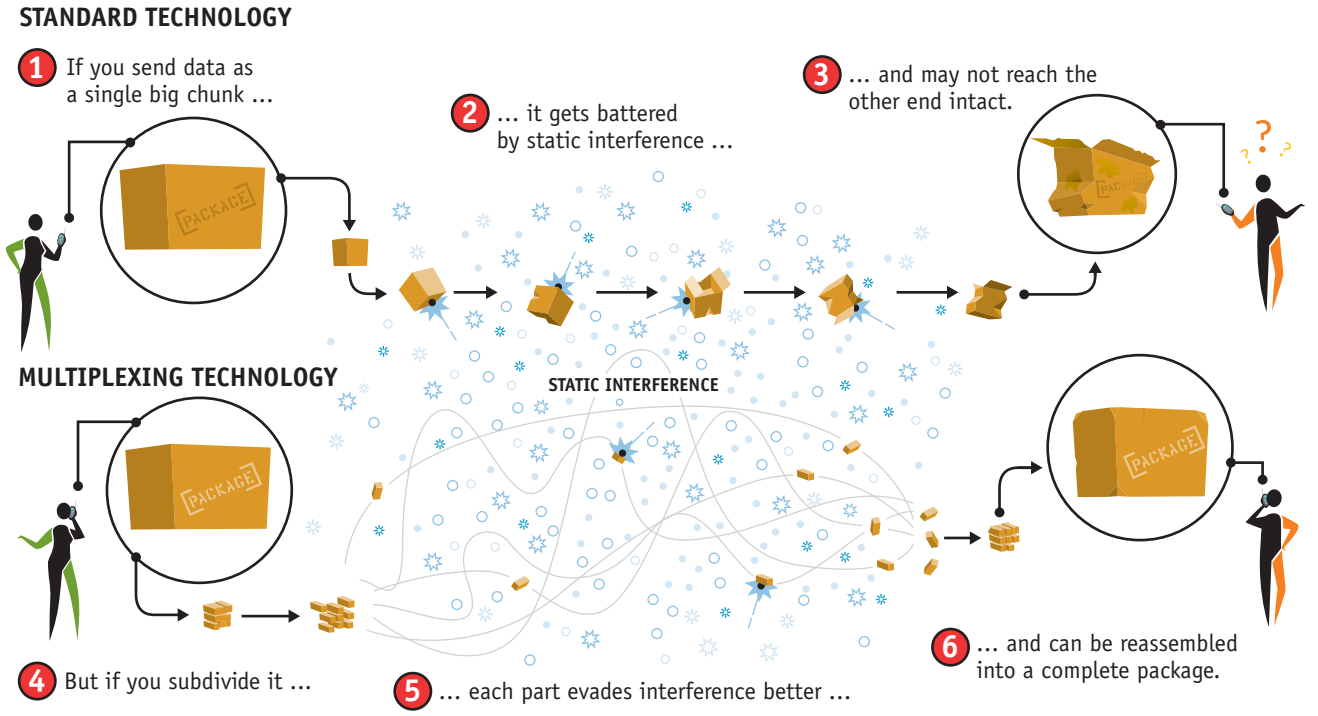


UWB TECHNOLOGY
Small pulses, in a prearranged sequence, encode the signal. Nonusers perceive it as noise and therefore ignore it.

Multiple Channels

Yet another approach—one that has been around in one form or another since the 1950s—is based on a communications technique known as multiplexing, which involves the transmission of more than one signal over the same channel. Multiplexing is commonly used in fiber optics, in

Standard vs. Multiplexing



which a big packet of data is chopped into smaller pieces, transmitted simultaneously on different wavelengths of light and stitched back together at the other end. Exactly the same principle applies in a wireless system, except that the wavelengths used are in the radio part of the electromagnetic spectrum.

To date, wireless multiplexing hasn't been exploited for cellular systems because digital signal processors fast enough to track and combine the different signals have not been available. That may change soon. A Calgary, Alberta–based company called Wi-LAN holds a number of key patents for a multiplexing technology known as wideband orthogonal frequency division multiplexing, or W-OFDM.

According to the company's CEO, Hatim Zaghloul, W-OFDM can deliver very high data rates across a limited range of radio spectrum—approximately 10 MHz in the unlicensed industrial-scientific-medical (ISM) bands at 900 MHz, 2.4 GHz and 5 GHz. The 10 MHz is divided into 10 evenly spaced channels, each of which can carry 1 Mbps of data.

So what, you might ask? After all, those 10 MHz could just be lumped into a single 10-Mbps channel. Multiplexing can't deliver something for nothing: it may divide a high-speed data stream into several low-speed data streams, but the total capacity of the radio spectrum, which is fixed by the laws of physics, must remain the same.

The key is that fast signals are more easily degraded by noise, interference and so-called multipath effects, which are caused by radio signals' bouncing off buildings or other landmarks. Slow signals, on the other hand, can slink through the static. By subdividing the spectrum, then, W-OFDM uses it more efficiently.

In one configuration, Wi-LAN has achieved 32 Mbps. In tests conducted earlier this year, technicians broadcast a stream of video to a car traveling at 70 miles per hour. Zaghloul says that he expects a whopping 155 Mbps by the end of next year as improved signal processors allow for more channels. The technology could be deployed in fixed wireless systems early next year and in mobile systems by 2003. The downside is that W-OFDM would require significant reworking of current cellular networks. Its adoption may have to wait until carriers look past 3G systems to 4G.

Because so many technologies—spread spectrum, antenna arrays, UWB, multiplexing and others—are in the works, many analysts are coming to realize that the biggest obstacle to fast wireless communications is not the engineering but the business model. What resources are carriers willing to put into their systems? What trade-offs will they make between the available bandwidth and the number of users forced to share it? "Deployment is the big issue," says Craig Mathias, an affiliate analyst with market research firm MobileInsights. "3G depends on the carriers. If they want to deliver high-speed data networks, they'll do it. But the business today is voice. The big question is the business plan, not the technology."

Cooper says that today's wireless industry is dominated by telecom monopolies that think in terms of a one-size-fits-all network. Instead, he says, he would like to see a multitude of different networks for different purposes. He predicts that nationwide voice networks will coexist with local data networks, and that low-cost, low-speed networks will rub shoulders with pricier high-speed ones. As for speed, Cooper says that wireless networks will eventually deliver the performance of wired ones.

In many ways, it's only an accident of history that we have wired, rather than unwired, telecommunications. If Guglielmo Marconi and Nikola Tesla had been a few years ahead of Alexander Graham Bell, instead of the other way around, we might have had a very different telecom landscape today.

LEANDER KAHNEY is a staff writer for Wired News.

Operating on a Beating Heart

Coronary bypass surgery can be a lifesaving operation. Two new surgical techniques should make the procedure safer and less expensive

by Cornelius Borst

OWEN FRANKEN *Corbis*

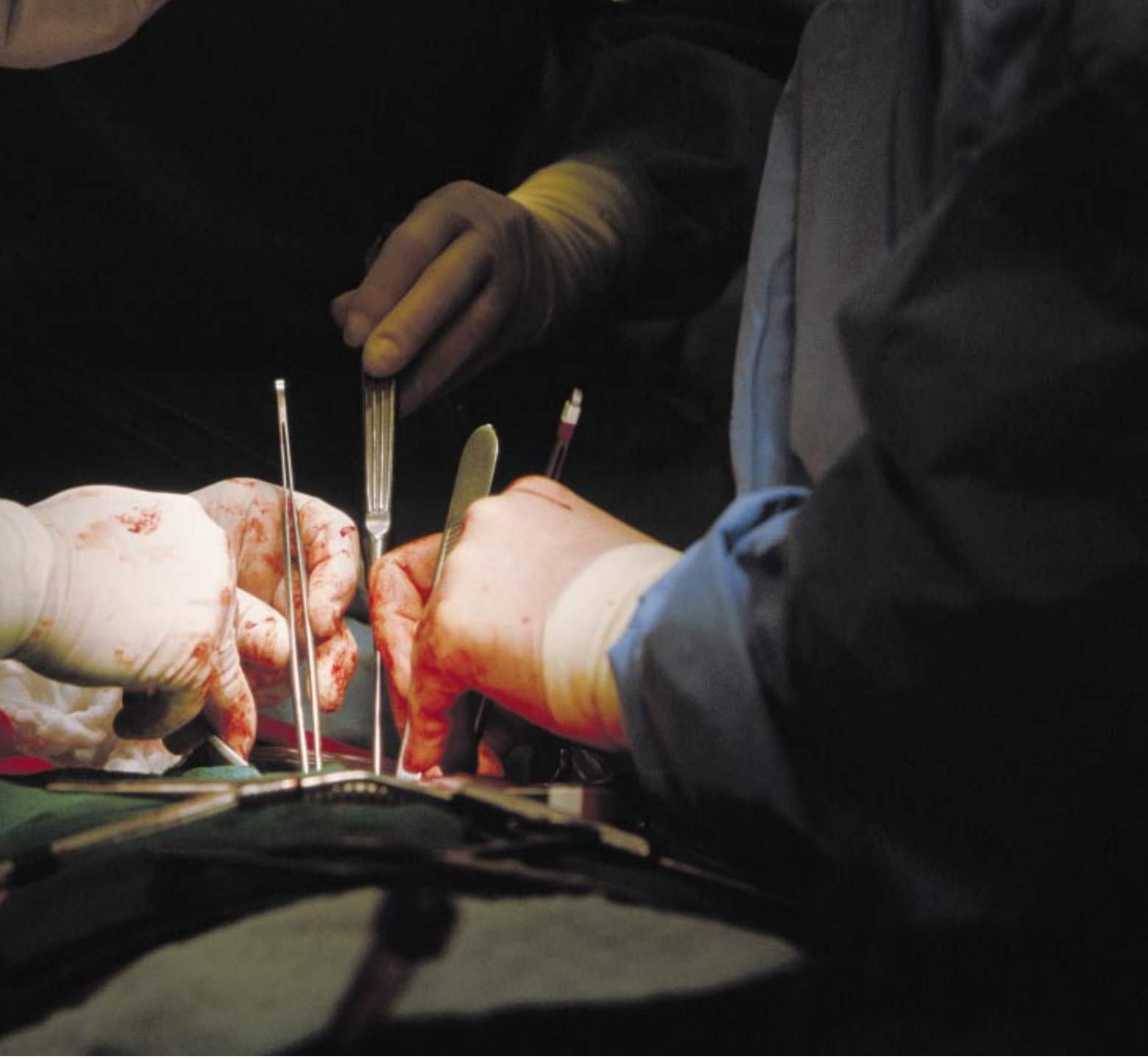
After climbing just one flight of stairs, Mr. Patnaki must rest before he ascends to the next story. He feels as though an elephant has stepped on his chest. Such pain results from blockages in Mr. Patnaki's coronary arteries, the vessels that supply oxygen-rich blood to the muscles of the heart. He needs coronary artery bypass surgery but cannot afford the operation and the lengthy hospital stay required. (In the U.S., for example, the surgery and hospitalization cost around $45,000; in Europe, about half this amount.)

Mrs. Wales is an elderly lady crippled by attacks of chest pain after just the slightest movement. Getting up and putting on her clothes takes at least an hour. She badly needs a coronary bypass. Fortunately, she lives near a cardiac care facility, and her medical insurance will pay for the procedure. Yet Mrs. Wales has lung problems and kidney disease, and she recently suffered a stroke. The cardiac surgeon considers it too dangerous to perform a bypass operation on her.

Mr. Brennick runs his own software business from an office at home. He needs triple bypass surgery but fears that the operation will put him out of business by diminishing his programming skills. Heart operations can sometimes impair a patient's brain function, and Mr. Brennick is not willing to take this chance. (Mr. Patnaki, Mrs. Wales and Mr. Brennick represent composite portraits based on numerous patients.)

Coronary bypass surgery is common—about 800,000 people undergo the procedure every year worldwide. But the operation is expensive and risky. To reroute the flow of blood around blockages in coronary arteries, surgeons must graft other vessels (taken from the patient's chest and leg) onto the diseased vessel, past the obstructions. Before doing so, however, they must open the chest (called "cracking" the chest, because the sternum must be split with a saw and the chest cavity spread open). They must then stop the heart, typically for around an hour. A surgeon simply cannot suture a vessel onto the heart accurately while it is still beating.

During the time the heart is stopped, the patient must be put on a heart-lung machine, which artificially circulates blood and supplies the body's tissues with oxygen until doctors restart the heart. This sophisticated machine ushered in the era of modern cardiac surgery some 40 years ago. Yet to this day,



CORONARY BYPASS SURGERY, as shown above, has typically involved "cracking" open the patient's chest cavity, stopping the heart and relying on a heart-lung machine to circulate blood and oxygen. Recovery from these procedures can take months. Surgeons have recently turned to beating-heart and closed-chest surgery as safer alternatives.

the artificial circulation provided by the heart-lung machine remains associated with serious complications, particularly in elderly or debilitated patients. It is the major cause of the long postoperative hospital stay (typically between six and eight days) and often results in a two- or three-month convalescence period at home. Furthermore, people may recover slowly from having had their chest cracked, and they are susceptible to certain infections, including pneumonia, as they recuperate.

In the mid-1990s two surgical techniques emerged that could signal a revolution in coronary bypass surgery. Researchers, including myself, began examining whether the heart-lung machine could be discarded by having doctors actually operate on a beating heart. Other teams have been investigating methods for performing endoscopic surgery on the heart—an operation that requires little more than a few keyhole-size incisions in the chest. I expect that over the next decade, coronary bypass surgery will become dramatically safer and less expensive thanks to these new technologies.

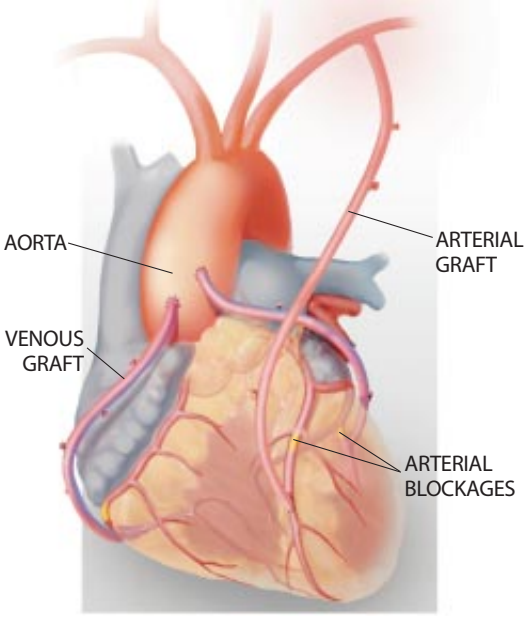
The chest pain experienced by Mr. Patnaki, Mrs. Wales and Mr. Brennick results from atherosclerosis—commonly known as hardening of the arteries—inside the major coronary arteries. Over time, substances such as cholesterol can build up in arterial walls, eventually narrowing these passageways. The disease progresses gradually, but in 19 percent of U.S. men between the ages of 30 and 35, the most impor-

tant coronary artery has already closed by at least 40 percent. By around middle age, people might notice a bit of chest pain when they exert themselves because the coronary blood flow can no longer keep up with the extra amount required during vigorous activity. A clogged vessel may be likened to a garden hose that won't spray after someone has stepped on it.

People are often crippled by the chest pain of atherosclerosis, and millions around the world have been stricken with this devastating disease. Genetic factors play a role in its development, but diet and lifestyle are also important. Although my emphasis—both in this article and in my research—is on improving therapeutic procedures to treat coronary heart disease, I want to stress that its prevention, through encouraging proper diet, exercise and not smoking, must be the medical community's primary focus.

Once a patient's chest pain has been diagnosed as a symptom of atherosclerosis, drugs may be recommended. Other patients opt for angioplasty, a procedure in which a cardiologist inserts a small, sausage-shaped balloon into the obstructed artery; inflating the balloon reopens the vessel by stretching the diseased wall. In addition, the cardiologist might position a tiny metal structure, or stent, inside the vessel to keep it open. But in some cases, when the cardiologist foresees that the artery will renarrow soon after angioplasty, a bypass is the best option for restoring adequate blood flow to the heart. Coronary bypass surgery usually involves grafting between three and five vessels onto the arteries of the heart. For each bypass graft, surgeons must spend up to 20 minutes carefully placing more than a dozen tiny stitches through both the graft vessel and the coronary artery.

The need to use a heart-lung machine is one of the greatest sources of complications during cardiac surgery. To connect a patient to the device, the doctor must insert tubes in the inflow and outflow vessels of the heart, close off the aorta with a clamp and introduce a cardioplegic solution into the coronary arteries, which stops the heart from beating. This complex procedure can dislodge particles of atherosclerotic plaque from the wall of the aorta. Such debris, if it reaches the brain, can cause a stroke. In addition, the heart-lung machine upsets the body's natural defense system, frequently resulting in fever, organ damage and blood loss; after the operation, it can also leave a patient temporarily unable to breathe without the aid of a ventilator. Finally, when the heart does resume beating, it often shows signs of impaired function: a patient may suffer low blood pressure, reduced blood flow through the body and reduced urine production. In rare cases, the patient cannot be weaned from the heart-lung machine without a mechanical pump to maintain acceptable blood pressure.



GRAFTING BYPASSES onto the heart typically involves attaching between three and five vessels to existing arteries so that blood flow through the bypasses will circumvent blockages. Surgeons can use either arterial grafts (arteries redirected from the vicinity of the heart) or venous grafts (vein segments taken from the leg).

Several studies have quantified these hazards. In particular, the likelihood of death soon after coronary bypass surgery increases with age. In the U.S., for example, it rises from a 1.1 percent chance between the ages of 20 and 50 to 7.2 percent between ages 81 and 90. One out of three patients suffers at least one operative complication. A 1997 report on more than 100,000 U.S. health insurance records revealed the dangers posed to bypass patients 65 and older: 4 percent died in the hospital; 4 percent were discharged to a nursing home; and 10 percent were discharged after at least two weeks in the hospital. Memory and attention loss as well as physical weakness and emotional depression often prevent patients from returning to normal activities for at least two or three months.

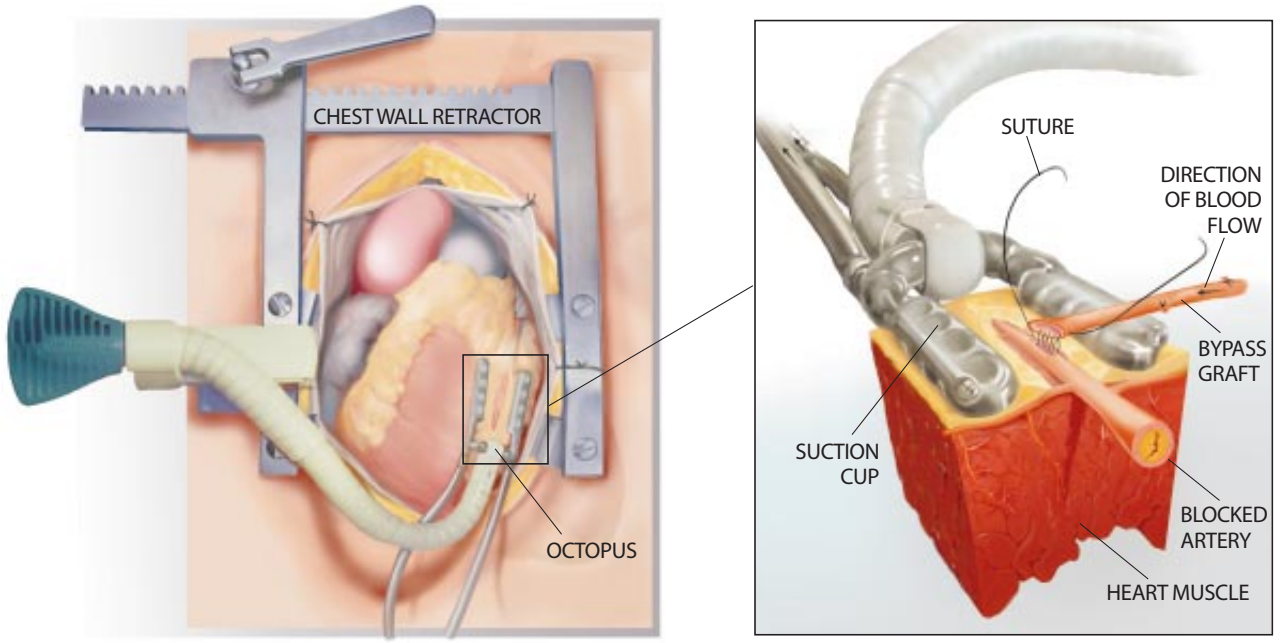
The practical implications of these potential risks vary. The possibility that a patient will require an extended stay in the hospital, perhaps in the intensive care unit on a ventilator, raises the odds that the final bill will be too high for someone like Mr. Patnaki. People who have a history of stroke, for instance, are more likely to have another one during the operation—which is why Mrs. Wales's physician recommended that she avoid bypass surgery. And the specter of possible memory loss scares away candidates like Mr. Brennick.

HEART-LUNG MACHINE is used in conventional bypass surgery, during which doctors must stop the heart. The machine circulates blood and oxygen to the body's tissues. Potential complications resulting from use of this device include an increased risk of stroke, fever, infection and blood loss, as well as the need to keep patients on artificial respirators longer.

For the past 15 years, my research has centered on devising better ways to treat coronary artery disease. By using a mechanical device to stabilize only the clogged vessel, not the entire heart, I believe my colleagues and I may have developed an improved and less expensive surgical therapy for this common disease.

In March 1993 in Palm Coast, Fla., at a workshop for physicians and researchers interested in the use of lasers in medicine and biology, I listened intently to Richard Satava, then a U.S. Army physician. He described a military initiative to design robots that

Operating on a Beating Heart



OCTOPUS HEART STABILIZER immobilizes an area on the surface of the beating heart so that surgeons can accurately suture a bypass graft. The Octopus, invented by the author and his colleagues, uses suction to take hold of a small region of the heart; tightening the blue knob anchors the Octopus to the metal device used to retract the chest wall (*left*). Although the heart continues to beat almost normally, the graft site (*right*) remains virtually still, allowing the surgeon to suture a bypass to the blocked artery.

would be remotely controlled by doctors to perform emergency surgery in the battlefield. Satava's photograph showing a prototype robot prompted me to think of using robots to operate on a beating heart inside a closed chest. While exploring a robotic approach to the surgery, I began to consider the feasibility of operating on a beating heart without such complex and expensive equipment.

The "Octopus"

In the spring of 1994 my colleague at the Heart Lung Institute in Utrecht, cardiac surgeon Erik W. L. Jansen, and I attempted to reproduce an approach to beating-heart surgery developed independently in the 1980s by Federico J. Benetti of the Cardiovascular Surgical Center of Buenos Aires and Enio Buffolo of the Paulista School of Medicine of the Federal University of São Paulo. Benetti and Buffolo had each reported their experiences with human patients; Jansen and I operated first on pigs.

In their work, the two South American doctors immobilized a small region of the heart's surface, which then allowed them to suture the coronary artery bypass successfully. They secured the region of interest with the help of a number of stabilizing sutures placed in tissue adjacent to the bypass site and through the use of pressure, applied by an assistant with a stable hand, who held a large surgical clamp. By restraining only part of the beating heart—just a few square centimeters—they hardly impeded its overall pumping action. Other surgeons, however, found it difficult to master this elegant, simple and cheap approach, and Benetti and Buffolo initially had few followers.

One day in May 1994 in Utrecht, during an experimental operation on a pig, I served as the assistant to the surgeon, charged with holding the clamp steady. Unfortunately, we failed to fully arrest the region of the heart where we wanted to place a bypass graft. But the failure inspired me. Unsteady tissue sutures and the human hand could be replaced by one rigid mechanical gadget to stabilize the heart. Exhilarating weeks followed, in which Jansen was able to construct with ease perfect bypasses on a pig's beating heart with the aid of prototype cardiac stabilizers crafted by technician Rik Mansvelt Beck.

Shortly thereafter my Utrecht colleague Paul Gründeman joined our team, and we invented the Octopus cardiac stabilizer—an instrument that can immobilize any small area on the surface of a beating heart [*see illustration above*]. The name originated from the fact that we use suction cups to attach the instrument to the heart and from "Octopussy," one of the laboratory pigs (all our animals were named after characters in James Bond movies). We first used the Octopus during bypass surgery on a human patient in September 1995. By mid-2000, more than 50,000 people had been treated with the Octopus worldwide (more than 400 of them here in Utrecht; in this select group of patients, the mortality rate, both during the operation and for 30 days afterward, is zero).

As is often the case in medical research, other investigators independently began developing mechanical stabilization devices around this time. In contrast to the Octopus, which holds onto the heart by suction, most of the other devices rely on pressure and friction—they resemble a large surgical clamp pressing on the heart. Currently there are some 13 different types of mechanical stabilizers available to cardiac surgeons. In 1994 fewer than 0.1 percent of coronary operations worldwide were performed without the aid of a heart-lung machine. In 1999 this number was about 10 percent. This year we expect it to rise to around 15 percent and by 2005 to more than 50 percent. At hospitals that lack sophisticated facilities with heart-lung machines—especially those in the developing world—the ability to perform beating-heart surgery will make coronary procedures available to patients for the first time.

Around this same time, Benetti, the surgeon from Argentina, gave the beat-

CAROL DONNER

SURGEON'S-EYE VIEW during robotic heart surgery (*bottom*) reveals two tiny instruments used to perform a coronary bypass operation; additional models are shown above. Surgeons can see inside the chest cavity thanks to camera lenses inserted through a small incision between the ribs; the surgeons' hand motions at a computer console are translated to the surgical instruments.

ing-heart approach another boost. He pioneered an operation involving a limited eight-centimeter incision between the ribs on the left side of the chest, which could be used in patients who needed only one bypass graft to the most important coronary artery on the front of the heart. Although this procedure still requires surgeons to separate adjacent ribs, it is significantly less damaging than cracking open the entire chest.

A number of other surgeons quickly recognized the potential advantages of this technique for beating-heart surgery, notably Valavanur Subramanian at Lenox Hill Hospital in New York City and Michael Mack at Columbia Hospital in Dallas. In November 1994 Subramanian showed a video presentation of his technique at a workshop in Rome; as a result, the limited-incision, beating-heart surgery spread quickly through Europe. In addition, Antonio M. Calafiore at the San Camillo de Lellis Hospital in Chieti, Italy, subsequently reported such good results in a large number of patients that beating-heart surgery began to attract worldwide attention. By the start of the first international workshop on minimally invasive coronary surgery, held in September 1995 in Utrecht, several thousand patients had undergone beating-heart surgery.

For the time being, beating-heart surgery will not fully replace traditional bypass surgery. For many candidates, the conventional operation will remain the better choice. But we continue to refine our method, expanding the types of cases for which it can be used. For example, when someone needs a bypass performed on the back of the heart (a common scenario), beating-heart surgery is often difficult. To reach the back of the heart, the surgeon must lift it partly out of the chest. This maneuver, when performed on an active heart, significantly deforms the organ, reduces the amount of blood it can pump and typically leads to a dangerous drop in blood pressure.

In the past few years, however, researchers have discovered a number of simple measures that can be taken to avoid this hazard. In my laboratory, Gründeman has shown that tilting the operating table 15 to 20 degrees down, so that the head is lower than the chest, helps to prevent a serious drop in blood pressure. At the Real Hospital Português in Recife, Brazil, Ricardo Lima found another elegant way to expose the back of the heart without compromising blood pressure too much. Most surgeons have now adopted his technique of using the pericardial sac surrounding the heart to lift the organ partly out of the chest.

By mid-2000, close to 200,000 patients had undergone beating-heart bypass surgery with the aid of a mechanical stabilizer. The first round of follow-up studies that we and many other centers conducted indicated that these people experienced fewer complications during surgery, required fewer blood transfusions, remained on an artificial respirator or in intensive care for less time, and left the hospital and returned to normal activities sooner than patients who had undergone traditional cardiac surgery. In addition, preliminary reports for single bypass procedures show that the overall cost was lower by about one third. Virtually all these studies, however, involved carefully selected patients. Thus, the results may not represent the general coronary surgery population. My colleagues and I await definitive results on the risks and benefits of beating-heart surgery that will be available once randomized clinical trials end. The Octopus trial in the Netherlands should conclude in late 2001.

Keyhole Surgery

The crucial advantage of beating-heart surgery is that the heart-lung machine can be turned off. Unfortunately, though, the other major drawback to conventional bypass surgery—the need to open the chest widely—remains. But this should not always be the case. In abdominal surgery, for example, physicians can perform entire operations, such as removing the gallbladder, through small, keyhole-size incisions, thanks to endoscopic surgery. In this technique, doctors insert a rigid tube connected to a miniature video camera (the endoscope) through one incision and the required surgical instruments through two other incisions; a video feed from the endoscope guides the surgeons' movements. So why not operate on the heart in a minimally invasive way, through one-centimeter openings between the ribs?

Researchers at Stanford University took just such a leap in 1991. The Stanford initiative led to the founding of the company Heartport, now in Redwood City, Calif., dedicated to performing closed-chest endoscopic cardiac surgery on a stopped heart with the patient hooked up to a heart-lung machine.

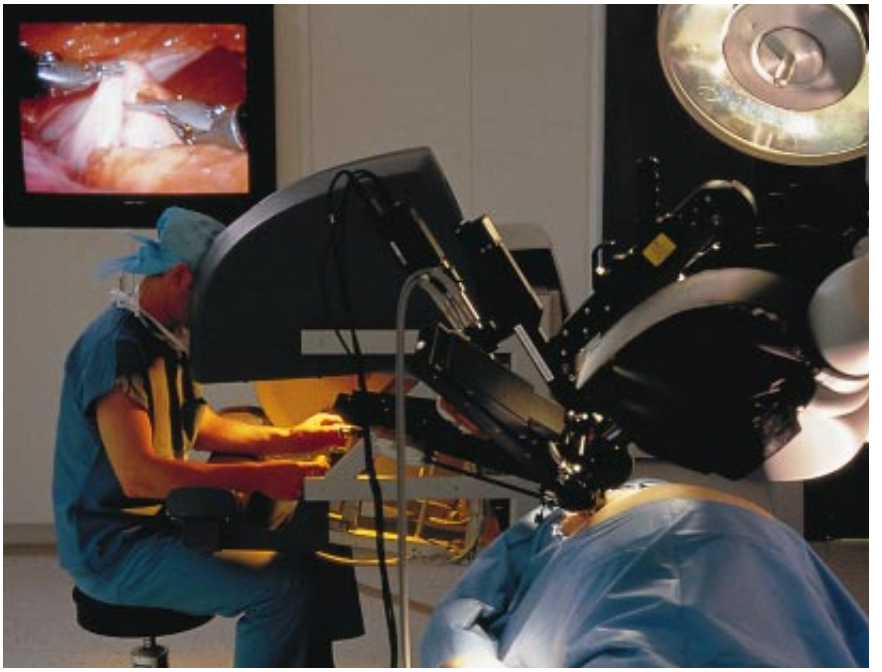
To connect a Heartport patient to the heart-lung machine and to stop the heart

PETER MENZEL (*top*); INTUITIVE SURGICAL (*bottom*)

without opening the chest, various tubes and catheters required for the task had to be manipulated from the groin area. This procedure did not go smoothly in all patients. Furthermore, the actual bypass suturing proved even more demanding. Because of the limitations of conventional endoscopic surgical instruments and the tight maneuvering space in the closed chest, these initial attempts to operate on the heart endoscopically had to be abandoned after just three patients. Only by making larger incisions (between six and nine centimeters) could surgeons reliably suture grafts to the coronary arteries. By mid-2000, more than 6,000 coronary patients had been treated in this manner.

Ideally, cardiac surgeons would like to perform a truly minimally invasive bypass operation: closed-chest, beating-heart coronary surgery. To avoid the restrictions of conventional endoscopic instruments, researchers—proceeding with great caution—have begun to use robotic endoscopic surgery systems for such operations. In these systems, the surgical instruments are not controlled directly by a surgeon's hands but instead by a remotely operated robot. Doctors can see inside the chest cavity in three dimensions, and their hand motions at the computer console are accurately translated to the surgical instruments inside the chest. Indeed, the computer automatically filters these motions to remove natural tremor and thus actually augments precision.

The first surgeons to take advantage of robotic equipment for closed-chest coronary surgery (but with a heart-lung machine) were Friedrich Mohr, Volkmar Falk and Anno Diegeler of the Heart Center of Leipzig University, and Alain Carpentier and Didier Loulmet of the Broussais Hospital in Paris. Working in 1998, in a renewed attempt to apply the original Heartport arrested-heart approach, these doctors combined Heartport with the so-called da Vinci robotic endoscopic surgery system, which was developed by Intuitive Surgical in Mountain View, Calif.



REMOTE CONTROLS used during robotic coronary bypass surgery require the surgeon (*left*) to sit a few feet from the patient. A monitor (*top left*) allows the other members of the surgical team to watch how the operation is proceeding.

In September 1999, at the University of Western Ontario Health Center in London, Ontario, Douglas Boyd utilized the Zeus robotic surgical system, which was developed by Computer Motion in Goleta, Calif., to perform the first computer-assisted, closed-chest, beating-heart surgery. But in contrast to the two hours that a single bypass, limited-incision operation on a beating heart usually requires, this first procedure lasted most of the day. By mid-2000, however, surgeons at five centers—in Munich, Leipzig, Dresden, London, Ontario, and London, England—had reduced operating-room time to between three and five hours for some 25 successful closed-chest, beating-heart, single-bypass operations.

Robotic techniques such as those required for a closed-chest operation are likely to become an integral part of the operating room. As the technology advances, surgical residents might one day be able to practice endoscopic coronary surgery just as pilots practice flying aircraft, and physicians might be able to rehearse upcoming operations. Other innovations may further facilitate the surgical treatment of coronary heart disease. For example, a "snap" connector in development may allow surgeons to attach a bypass rapidly without sutures.

Ultimately, the coronary bypass operation may very well become extinct. In the meantime, however, improving coronary surgery while keeping the cost reasonable remains an important goal—particularly because such advancements could make surgical interventions against coronary heart disease available worldwide to every patient who needs them. But regardless of new developments in surgical techniques, prevention of coronary heart disease must remain at the top of the medical agenda. SA

The Author

CORNELIUS BORST is professor of experimental cardiology at the Utrecht University Medical Center in the Netherlands. After receiving an M.D. and Ph.D. from the University of Amsterdam, he became chairman of the Experimental Cardiology Laboratory in Utrecht in 1981. His other research interests include the mechanisms of atherosclerotic coronary narrowing and renarrowing following angioplasty.

Further Information

MINIMALLY INVASIVE CORONARY ARTERY BYPASS GRAFTING: AN EXPERIMENTAL PERSPECTIVE. Cornelius Borst and Paul F. Gründeman in *Circulation*, Vol. 99, No. 11, pages 1400–1403; March 23, 1999.

MINIMALLY INVASIVE CARDIAC SURGERY. Edited by Robbin G. Cohen et al. Quality Medical Publishing, 1999.

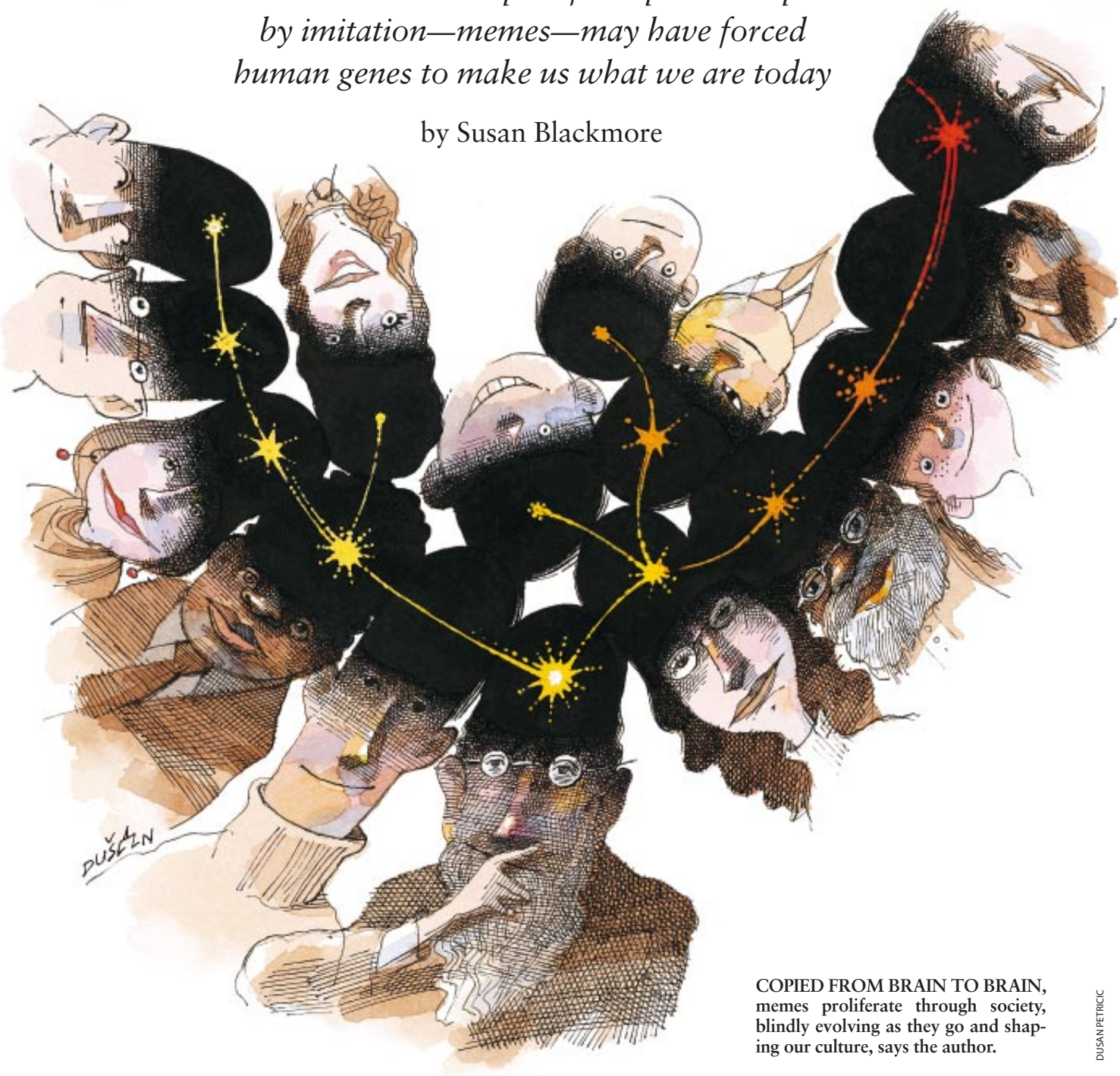
MINIMAL ACCESS CARDIOTHORACIC SURGERY. Edited by Anthony P. C. Yim et al. W. B. Saunders Company, 2000.

PETER MENZEL

THE POWER OF MEMES

Behaviors and ideas copied from person to person by imitation—memes—may have forced human genes to make us what we are today

by Susan Blackmore



COPIED FROM BRAIN TO BRAIN, memes proliferate through society, blindly evolving as they go and shaping our culture, says the author.

DUSAN PETRICIC

Did you know that you spend much of your life copying and transmitting entities called memes? A meme (pronounced "meem") is "an idea, behavior, style or usage that spreads from person to person within a culture." Whenever you shake hands, sing "Happy Birthday" or cast your vote in an election, you are giving life to memes.

So far, no debate. But controversy has erupted over the proposal, presented here by psychologist **Susan Blackmore,** that humans' uncanny ability to imitate, and thus to transmit memes, is what sets us apart from other species. Memes, she argues, have been (and are) a powerful force shaping our cultural—and biological—evolution. To convey the debate, we have included three short counterpoints, written by behavioral ecologist **Lee Alan Dugatkin,** evolutionary anthropologist **Robert Boyd** and population biologist **Peter J. Richerson,** and psychologist **Henry Plotkin.** Enjoy this smorgasbord of competing memes.

Human beings are strange animals. Although evolutionary theory has brilliantly accounted for the features we share with other creatures—from the genetic code that directs the construction of our bodies to the details of how our muscles and neurons work—we still stand out in countless ways. Our brains are exceptionally large, we alone have truly grammatical language, and we alone compose symphonies, drive cars, eat spaghetti with a fork and wonder about the origins of the universe.

The problem is that these abilities seem surplus to requirements, going well beyond what we need to survive. As Steven Pinker of the Massachusetts Institute of Technology points out in *How the Mind Works,* "As far as biological cause and effect are concerned, music is useless." We might say the same of art, chess and pure mathematics.

Classical (Darwinian) evolutionary theory, which focuses on inheritable traits of organisms, cannot directly justify such riches. Expressed in modern terms, this theory holds that genes control the traits of organisms; over the course of many generations, genes that give their bearers a survival advantage and that favor production of many offspring (who will inherit the genes) tend to proliferate at the expense of others. The genes, then, essentially compete against one another, and those that are most proficient at being passed to the next generation gradually prosper.

Few scientists would want to abandon Darwinian theory. But if it does not clarify why we humans have come to apportion so much of our resources to so many abilities that are superfluous to the central biological task of further propagating our genes, where else can we look?

The answer, I suggest, lies in memes. Memes are stories, songs, habits, skills, inventions and ways of doing things that we copy from person to person by imitation. Human nature *can* be explained by evolutionary theory, but only when we consider evolving memes as well as genes.

It is tempting to consider memes as simply "ideas," but more properly memes are a form of information. (Genes, too, are information: instructions, written in DNA, for building proteins.) Thus, the meme for, say, the first eight notes of the *Twilight Zone* theme can be recorded not only in the neurons of a person (who will recognize the notes when she hears them) but also in magnetic patterns on a videocassette or in ink markings on a page of sheet music.

The Birth of Memes

The notion that memes exist and evolve has been around for almost 25 years, but only recently has it gained attention as a powerful force in human evolution. Richard Dawkins of the University of Oxford coined the word in 1976, in his best-selling book *The Selfish Gene.* There he described the basic principle of Darwinian evolution in terms of three general processes—when information is copied again and again, with variations and with selection of some variants over others, you *must* get evolution. That is, over many iterations of this cycle, the population of surviving copies will gradually acquire new properties that tend to make them better suited to succeeding in the ongoing competition to produce progeny. Although the cycle is mindless, it generates design out of chaos.

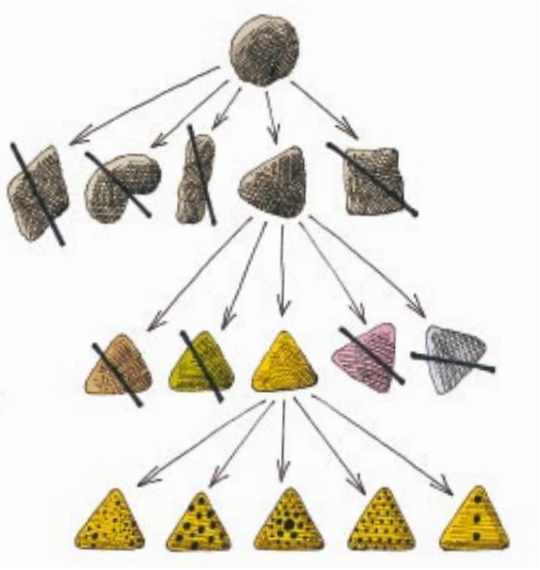
Dawkins called the information that gets copied the "replicator" and pointed out that the most familiar replicator is the gene. But he wanted to emphasize that evolution can be based on any replicator, and so, as an example, he invented the idea of the meme. The copying of memes from one person to another is imperfect, just as the copying of genes from parent to child is sometimes inaccurate. We may embellish a story, forget a word of the song, adapt an old technology or concoct a new theory out of old ideas. Of all these variations, some go on to be copied many times, whereas others die out. Memes are thus true replicators, possessing all three properties—replication, variation, selection—needed to spawn a new Darwinian evolutionary process.

Dawkins says that he had modest intentions for his new term—to prevent his readers from thinking that the gene was the "be-all and end-all of evolution, the fundamental unit of selection"—but in fact his idea is dynamite. If memes are replicators, then they, like genes, compete to get copied *for their own sake.* This conclusion contradicts the assumption, held by most evolutionary psychologists, that the ultimate

function of human culture is to serve the genes by aiding their survival. The founder of sociobiology, E. O. Wilson, famously said that the genes hold culture on a leash. Culture might temporarily develop in some direction that is counterproductive to spreading the genes, but in the long run it is brought back in line by gene-based natural selection, like a straying dog curbed by its owner. In this view, memes would be slaves to the genes that built the brains that copy them, prospering only by helping those genes to proliferate. But if Dawkins is right and memes are replicators, then memes serve their own selfish ends, replicating whenever they can. They sculpt our minds and cultures as they go—whatever their effect on the genes.

The most obvious examples of this phenomenon are "viral" memes. Chain letters (both hard-copy and e-mail) consist of little bits of written information, including a "copy-me" instruction backed up with threats (if you break the chain you will suffer bad luck) or promises (you'll receive money and you can help your friends). It does not matter that the threats and promises are empty and your effort in copying the letters is wasted. These memes have an internal structure that ensures their own propagation.

The same can be said, Dawkins argues, of the great religions of the world. Of all the myriad small cults with charismatic leaders that have sprung up in human history, only a few had what it took to survive—copy-me instructions backed up with threats and promises. In religions the threats are of death or eternal damnation, and the promises are of everlasting bliss. The costs are a proportion of one's income, a lifetime devoted to propagating the word, or resources spent on building magnificent mosques and cathedrals that further promote the memes. The genes may even suffer directly at the hands of the memes—as occurs with a celibate priesthood.



UNIVERSAL DARWINISM produces evolution in any system of replicators that exhibit variation, selection and heredity. Variation arises by recombination and imperfect copying. Selection occurs when limited resources cannot support all the variants. Heredity ensures that good properties are passed on. This mindless algorithm generates highly sophisticated entities.

Of course, not every cult (or chain letter) with the appropriate viral structure will actually succeed. Some threats and promises are more effective, or virulent, than others, and all compete for the limited resource of human attention in the face of experience and skepticism (which, in the viral metaphor, act as a kind of immune system).

Arguably, religions are not entirely viral; for example, they provide comfort and a sense of belonging. In any case, we must not make the mistake of thinking that all memes are viruses. The vast majority make up the very stuff of our lives, including languages, political systems, financial institutions, education, science and technology. All these are memes (or conglomerations of memes), because they are copied from person to person and vie for survival in the limited space of human memories and culture.

Thinking memetically gives rise to a new vision of the world, one that, when you "get" it, transforms everything. From the meme's-eye view, every human is a machine for making more memes—a vehicle for propagation, an opportunity for replication and a resource to compete for. We are neither the slaves of our genes nor rational free agents creating culture, art, science and technology for our own happiness. Instead we are part of a vast evolutionary process in which memes are the evolving replicators and we are the meme machines.

This new vision is stunning and scary: stunning because now one simple theory encompasses all of human culture and creativity as well as biological evolution; scary because it seems to reduce great swathes of our humanity, of our activities and our intellectual lives, to a mindless phenomenon. But is this vision true? Can memetics help us to understand ourselves? Can it lead to testable predictions or do any real scientific work? If it cannot, memetics is worthless.

I believe that the idea of the meme as replicator is what has been missing from our theories of human evolution and that memetics will prove immensely useful for explaining our unique attributes and the rise of our elaborate cultures and societies. We are different from all other animals because we alone, at some time in our far past, became capable of widespread generalized

MEMES AND COMPLEXES OF MEMES

- Stories, urban legends, myths
- Clothing, hairstyles, body piercing
- Cuisine, cigarette smoking
- Applauding, cheering
- Language, accents, catchphrases
- Songs, music, dances
- Belief in UFOs, ghosts, Santa Claus
- Racist slogans, sexist jokes
- Religions
- Inventions, theories, science
- Judicial systems, democracy
- Proust's story of the madeleine cake

NOT MEMES

- Subjective experiences, complex emotions, sensory perceptions
- Eating, breathing, having sex
- Innate behaviors, even if contagious: yawning, coughing, laughing
- Conditioned responses: fear at the sound of a dentist's drill
- Cognitive maps: knowing the way around your neighborhood
- Associations with sounds and smells

Note: Many human behaviors are complicated mixtures of innate, learned and imitated—for example, riding a bicycle.

DUSAN PETRICIC

Animals Imitate, Too

by Lee Alan Dugatkin

BLACKBIRDS can be convinced to fear creatures that are not a threat, because they learn to recognize predators by watching what spooks their companions.

I applaud Susan Blackmore's attempt to infect people's minds with the meme "imitation is important." But I take issue with her view that memes—the imitated entities—influence the evolution of behavior in humans alone. Animals from fish to primates copy one another in making such decisions as what to eat and with whom to mate. That being the case, I will argue that memes may influence the habits of many animals just as they drive human behavior. A close look at blackbirds can help to illustrate that memes are not necessarily unique to humans or even to other primates, such as chimpanzees and other apes. But first I should clarify the definition of the word *imitation*.

Psychologists seem to revel in debating the meaning of imitation, and dozens of papers divide its meaning into an array of subcategories. In a discussion of memes, however, it seems only fair to use Blackmore's own description. Her book *The Meme Machine* presents two different perspectives. The strictest definition states that imitation involves three complex stages: deciding what to imitate, transforming one point of view to another and producing a matched bodily action. Under such strict criteria, no rock-solid cases of animal imitation may exist. It is extraordinarily difficult to decipher whether animals can transform one viewpoint to another and, if so, whether we know what exactly they are choosing to imitate.

Blackmore also promotes a much more liberal idea of imitation when she describes a story being passed from one friend to another. "You have not precisely imitated your friend's every action and word, but something (the gist of the story) has been copied from her to you and then on to someone else," she writes. Surely hundreds of examples of animal imitation fall within this broad definition, and the way blackbirds learn about predators is no exception.

In 1978 Eberhard Curio of Ruhr University of Bochum in Germany and his colleagues created a small theater in which one blackbird could view a second one squawking and flicking its tail in reaction to a nearby predator. The second bird was responding to a true predator—a little owl—but a series of partitions hid the owl from the first blackbird's view. Thanks to some clever manipulations, the observer was made to think that its companion was reacting to a noisy friarbird, which blackbirds do not normally regard as a threat. The researchers then put the observer blackbird near a friarbird, and it, too, reacted with squawks and tail flicks. Curio and his colleagues discovered that the false message "friarbirds are predators" can spread down a chain of at least six other blackbirds.

Yet the simple fact that something is copied does not make it a full-fledged meme. Blackmore argues that a message has to meet three additional criteria: it must be copied accurately, many copies must be made, and the copies must last a long time. The message "friarbirds are predators" was accurately transmitted, and copies of the message spread from individual to individual, thus demonstrating some degree of fecundity. It is impossible to assess the longevity of this meme based on laboratory experiments, but in principle there is no reason that the information wouldn't stick around once established in natural populations.

Memes may influence the habits of many animals just as they drive human behavior.

In my work as a behavioral ecologist I have run across dozens of other examples of animal behavior that fit the definition of a meme, and I would not be surprised if the total number were quite large. Memes may be older and more fundamental to biological evolution than Blackmore or anyone else has argued to date. More specifically, the difference between animal and human memes may be quantitative rather than qualitative. Memeticists may well take hold of the idea that animal memes are real and use this to bolster the claim that memes truly are a universally important force in evolution. But if memes do not separate us from animals, as Blackmore suggests, then they alone cannot explain why human culture is uniquely advanced.

LEE ALAN DUGATKIN is an associate professor in the biology department at the University of Louisville. He has studied imitation in animals for 10 years. His new book on this subject, The Imitation Factor: Evolution Beyond the Gene *(The Free Press), will be published in January.*

DUSAN PETRICIC

RISE OF CULTURE could have begun when our early hominid ancestors learned to imitate one another (*left*). Individuals best able to imitate new survival skills such as fire making would prosper, favoring the genes that made them better adapted for generalized imitation. Later, as humans became more genetically selected to imitate (*center*), the genes would need to evolve strategies to ensure imitation of the most valuable behaviors. Tactics such as "imitate the best imitators" would result in better copying of new

imitation. This let loose new replicators—memes—which then began to propagate, using us as their copying machinery much as genes use the copying machinery inside cells. From then on, this one species has been designed by two replicators, not one. This is why we are different from the millions of other species on the planet. This is how we got our big brains, our language and all our other peculiar "surplus" abilities.

From the meme's-eye view, every human is a machine for making more memes—a resource to compete for.

Big Brains for Memes

Memetics neatly resolves the mystery of the human brain's vastness. The human brain is about as big as the genes can make it—three times bigger, relative to body weight, than the brains of our closest relatives, the great apes. It is expensive to build and maintain, and many mothers and babies die through childbirth complications caused by the size of the head. Why has evolution allowed the brain to grow so hazardously large? Traditional theories look to genetic advantage, in improved hunting or foraging skills or the ability to sustain larger cooperating groups with complex social skills. Memetics provides a completely different explanation.

The critical transition for hominids was the dawn of imitation, perhaps two and a half million years ago, before the advent of stone tools and expanding brains. True imitation means copying a novel behavior or skill from another animal. It is difficult to do, requires a lot of brainpower and is correspondingly rare in the animal kingdom. Although many birds copy songs, and whales and dolphins can imitate sounds and actions, most species cannot. Often animal "imitation," such as learning to respond to a new predator, involves merely the use of an innate behavior in a new situation. Even chimpanzees' imitation is limited to a small range of behaviors, such as methods of fishing for termites. In contrast, generalized imitation of almost any activity seen—as seems to come naturally to humans—is a much more difficult and correspondingly more valuable trick, letting the imitator reap the benefits of someone else's learning or ingenuity as often as possible. For example, in experiments conducted in 1995 at the Yerkes Regional Primate Research Center in Georgia, when the same problems were presented to orangutans and human children, only the humans readily used imitation to solve the problems.

It is easy to imagine that our early ancestors imitated useful new skills in making fire, hunting, and carrying and preparing food. As these early memes spread, the ability to acquire them became increasingly important for survival. In short, people who were better at imitation thrived, and the genes that gave them the bigger brains required for it consequently spread in the gene pool. Everyone got better at imitation, intensifying the pressure to enlarge the

survival skills—but also of extraneous behaviors such as decorating clothing. Good imitators would gain social status, attract mates and have more offspring, further driving the genes to develop bigger brains capable of elaborate imitation. Imitation would become an intrinsic part of human nature, and ever evolving memes would gradually produce entire cultural systems (*right*), complete with activities such as monument building and human sacrifice that have no payoff for the genes but serve to transmit the associated memes.

brain still further in a kind of cerebral arms race.

Once everyone started imitating, the second replicator was let loose on the world, changing human evolution forever. The memes started to take control. Alongside useful skills, such as building fires, people copied less useful ones like fancy body decoration and downright costly ones such as energetic but futile rain dances. The genes faced a problem: how to ensure that their carriers copied only the useful behaviors. Newly arisen memes can spread through a population by imitation in a single generation, faster than genetic evolution can respond. By the time the genes could evolve a hardwired predilection for making fires and an aversion to performing rain dances, completely different fads could arise and hold sway. The genes can develop only broad, long-term strategies to try to make their bearers more discriminating about what they imitate.

A useful general heuristic that the genes could bestow might be a predisposition to copy the best imitators—the people most likely to have accurate versions of currently useful memes. (More familiar terms for "the best imitators" in modern life may be "trendsetters" or "role models.") In addition to their bag of useful tricks for survival, the best imitators would thereby acquire higher social status, further improving their survival chances and helping to propagate the genes that made them talented imitators—the genes that gave them big brains specialized at accurate generalized imitation.

The genes would continue to respond with improvements in people's innate preferences about what to imitate, but the genes' response, requiring generations of people to act on, would always lag far behind the memetic developments. I call the process by which memes control gene selection "memetic drive": memes compete among themselves and evolve rapidly in some direction, and genes must respond by improving selective imitation—increasing brain size and power along the way. Successful memes thus begin dictating which genes will be most successful. The memes take hold of the leash.

In a final twist, it would pay for people to mate with the most proficient imitators, because by and large, good imitators have the best survival skills. Through this effect, sexual selection, guided by memes, could have played a role in creating our big brains. By choosing the best imitator for a mate, women help propagate the genes needed to copy religious rituals, colorful clothes, singing, dancing, painting and so on. By this process, the legacy of past memetic evolution becomes embedded in the structures of our brains and we become musical, artistic and religious creatures. Our big brains are selective imitation devices built by and for the memes as much as for the genes.

Origin of Language

Language could have been another exquisite creation of this same process of meme-gene coevolution. Questions about the origins and function of language have been so contentious that in 1866 the Linguistic Society of Paris banned any more speculation on the issue. Even today scientists have reached no general consensus, but the most popular theories appeal to genetic advantage. For example, evolutionary psychologist Robin Dunbar of the University of Liverpool argues that language is a

DUSAN PETRICIC

Memes are best thought about not by analogy with genes but as new replicators, with their own ways of surviving and getting copied.

substitute for grooming in keeping large social groups together. Evolutionary anthropologist and neuroscientist Terrence Deacon of Boston University proposes that language made symbolic communication possible, which in turn allowed improved hunting skills, tighter social bonds and group defense.

In contrast, the theory of memetic drive explains language by its conferring survival advantages on memes. To understand how this works, we must ask which kinds of memes would have survived best and proliferated in the emerging meme pool of our early ancestors. The general answer for any replicator is those with high fecundity, fidelity and longevity: ones that make many accurate and long-lived copies of themselves.

Sounds are more fecund than gestures, particularly sounds analogous to "hey!" or "look out!" Everyone within earshot can hear a shout, whether they happen to be looking at the speaker or not. Fidelity of spoken memes is higher for those built from discrete units of sound (phonemes) and divided into words—a kind of digitization that reduces errors in copying. As different actions and vocalizations competed in the prehistoric meme pool, such spoken words would prosper and displace less well adapted memes of communication. Then, stringing words together in different orders, and adding prefixes and inflections, would provide fertile niches for new, more sophisticated vocal memes. In sum, the highest-quality replicable sounds would crowd out the poorer ones.

Now consider the effect of this on the genes. Once again the best imitators (the most articulate individuals) would acquire higher status, the best mates and the most offspring. In consequence, genes for the ability to imitate the winning sounds would increase in the gene pool. I suggest that by this process the successful sounds—the foundations of spoken language—gradually drove the genes into creating a brain that was not merely big but especially adept at copying those particular articulations. The result was the remarkable human capacity for language. It was designed by memetic competition and meme-gene coevolution.

The process of memetic driving is an

COUNTERPOINT

Meme Theory Oversimplifies How Culture Changes

by Robert Boyd and Peter J. Richerson

IDEAS are often systematically transformed as they pass from one person to another or from one generation to the next.

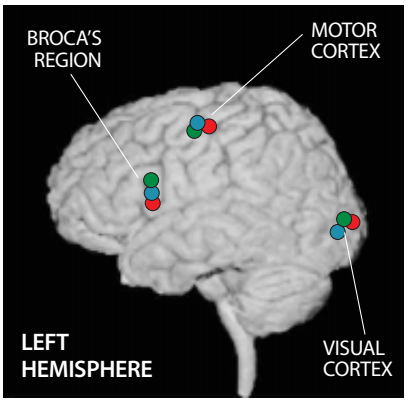
DUSAN PETRICIC

Genes are replicators. They pass faithfully from parent to child and control the machinery of life. This faithful transmission is what enables natural selection to operate: genes that cause their bearers to survive better or reproduce faster than bearers of other genes will spread through the population. Other processes, such as mutation, play crucial roles in evolution, but most adaptation can be explained by asking which genes will replicate at the highest rate. This simple rule has astonishing power, allowing biologists to understand phenomena as diverse as the shape of the human pelvis and the timing of sex changes in hermaphroditic fish.

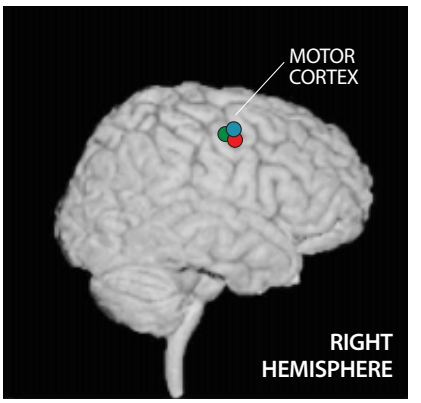
Susan Blackmore argues that beliefs and ideas, which she calls memes, are also replicators. They are copied faithfully from one mind to another and control the behavior of the people who acquire them. That being the case, Blackmore suggests, the evolution of ideas is also shaped by natural selection, and cultural change can be understood by asking which memes replicate most quickly.

We think Blackmore is at least half right. Ideas from biology are certainly useful for studying cultural evolution. Culture does consist of ideas stored in a population of human brains, and mechanisms analogous to natural selection can affect which ideas spread and which ones disappear. But Blackmore is probably wrong in thinking that cultural evolution can be explained in terms of natural selection alone. Instead scientists need to combine research from psychology, anthropology and linguistics to clarify the multiple processes that actually shape human culture.

Unlike genes, ideas usually are not passed intact from one person to another. Information in one person's brain generates a behavior, and then someone else tries to infer the information required to do the same thing. Breakdowns in the accurate transmission of ideas can occur because differences in the genes, culture or personal background of two individuals can cause one person to make a wrong assumption about what motivated the other's behavior. As a result, memes are often systematically transformed



SCANS OF BRAIN ACTIVITY could test whether the human brain has evolved to imitate and spread memes. Here researchers mapped neural activity associated with a specific hand motion. The same areas lit up whether the subjects acted of their own volition (*red*), merely observed someone else (*blue*) or imitated the other person (*green*). Imitation produced the strongest activity. The results suggest that Broca's region in humans controls a "mirror neuron system" specially evolved to imitate actions. Monkeys, however, have a similar system.



BRAIN IMAGES AND DATA FROM NOBUYUKI NISHITANI AND RIITTA HARI *Helsinki University of Technology*

example of replicators (memes) evolving concurrently with their copying machinery (brains). The appearance of memes is not the first time such concurrent evolution has occurred: something similar must have taken place in the earliest stages of life on earth, when the first replicating molecules developed in the primeval soup and evolved into DNA and all its associated cellular replication machinery. As with the evolution of that sophisticated gene-copying apparatus, we might expect better meme-copying machinery to have appeared—and it has. Written language provided a vast leap forward in longevity and fidelity; the printing press enhanced fecundity. From the telegraph to the cell phone, from "snail" mail to e-mail, from phonographs to DVDs and from computers to the Internet, copying machinery has been improving, spreading a growing multitude of memes farther and faster. Today's information explosion is just what we should expect of memetic evolution.

This memetic theory depends on a number of conjectures that can be tested, especially the assumption that imitation requires a lot of brainpower, even though it comes so easily to us. Brain-

during transmission—a process quite unlike natural selection, which depends on one meme spreading more quickly than competing alternatives. Transformation, on the other hand, could cause people in one generation to acquire a different meme than the one held by every person in the previous generation.

David Wilkins of the Max Planck Institute for Psycholinguistics in Nijmegen in the Netherlands discovered a simple example of meme transformation when he found that Americans of different generations varied in their understanding of the word ending *-gate*. People over the age of 40 assumed that *-gate* implied a government scandal in Washington, usually involving a cover-up. These baby boomers experienced Richard Nixon's presidency as adults and interpreted constructions such as *Travelgate* as scandals analogous to Watergate. Younger Americans had heard *-gate* used to refer to a variety of scandals in Washington. But knowing much less about Watergate, they couldn't detect this common thread and instead analyzed *-gate* as a suffix that can be added to any word to indicate a scandal. Notice that this transformation could have occurred without competition among alternative memes. Every meme in every baby boomer brain could specify that *-gate* means a government scandal like Watergate; nonetheless, every younger person could infer *-gate* to mean any scandal.

Cultural evolution cannot be explained in terms of natural selection alone.

As Blackmore notes, genes can also be transformed by spontaneous changes called mutations. But genetic mutations are rare, occurring about once every million replications, and as a result their effect usually can be ignored when thinking about adaptations. If mutations occurred more often—say, every 10 replications—they would have a significant effect on which genes were most common. We think this situation is exactly what occurs with ideas, which can transform rapidly as they spread from one person to the next. If we are right, cultural change will be understood only if the effects of transformation and natural selection are combined.

A number of other nonselective processes may affect the evolution of ideas. For example, a person can learn an idea from someone else and then modify the idea in an effort to improve it. Still other nonselective processes can arise when people synthesize their own beliefs after being exposed to a number of people who behave differently. We think that successful interpretations of cultural change require meticulous attention to the many processes that guide particular instances of cultural evolution. Social scientists have already made some progress on this project. William Labov of the University of Pennsylvania has described psychological and social processes that cause gradual changes in dialect from generation to generation, for instance, and Albert Bandura of Stanford University has studied how imitation shapes the acquisition of ideas.

Over the past century biologists have developed many concepts and mathematical tools that can help clarify what happens when a variety of processes interact to shape the evolution of populations. By combining these ideas with empirical studies, scientists may then be able to understand how culture evolves.

ROBERT BOYD and PETER J. RICHERSON have collaborated for 25 years in studying the evolution of human culture and how cultural and genetic evolution interact. Their work couples mathematical models with empirical work drawn from laboratory and field research. Boyd is an evolutionary anthropologist at the University of California, Los Angeles; Richerson is a population biologist at the University of California, Davis.

People Do More Than Imitate

by Henry Plotkin

I see two main problems with memetics of the Susan Blackmore variety. First is her suggestion that culture is nothing but a collection of memes. She includes everything from a simple action such as using a stone tool to complex institutions such as banks. The second problem is her idea that all memes, and thus all aspects of culture, are spread by imitation. From my perspective as a psychologist, neither assertion is correct.

Early in the 20th century American psychologist Edward Thorndike defined imitation as learning to do an act from seeing it done; this meaning prevails in psychological research today. If the word *imitation* is used in this way, then Blackmore's assertions are trivial, because imitating actions transmits almost nothing of cultural importance. Tying shoelaces and throwing a ball are not in themselves significant in human affairs.

If the word *imitation* is used instead as Blackmore prefers—to mean any and every manner of communication between people, from passing on the gist of a story to remembering the instructions read in a manual a week ago—then the term becomes so vague as to be meaningless. And even this broad definition of imitation cannot account for the existence and evolution of culture, which is much more than the rote repetition of physical actions. Human culture is about the sharing of knowledge, beliefs and values.

At the core of any culture are shared understandings about how the world works, sometimes referred to as schemas. The rules that operate in restaurants form a classic schema: in such places someone prepares your food, brings it to your table and cleans up after you in exchange for money. Children acquire the many schemas that characterize their own culture through a mix of informal guidance from adults and peers and by the complex psychological mechanisms that enable a person to make sense of abstract ideas. Imitation, properly defined, does not come into it.

Human culture is about the sharing of knowledge, beliefs and ideas. Imitation, properly defined, does not come into it.

Shared beliefs and values, also called social constructions, come to us in a similarly complex and ill-understood fashion. In contrast to schemas, which describe tangible entities such as restaurants, social constructions exist only because people agree that they do. Money is a social construction. So, too, is justice. Some of them have physical embodiments, such as paper or coins, but they all go beyond the physical and into mental agreements about what things mean. Without consensus that bills and coins have specific values, money is worthless. Many beliefs and values also regulate social interactions. In much of Western culture, for instance, justice is based on concepts of fairness and ownership. Other cultures define justice through such ideas as service or revenge. In all cases, justice goes beyond courts of law, judges or prisons.

SHARED KNOWLEDGE, such as the rules that operate in restaurants, cannot be imitated.

Scientists have remarkably few detailed studies of how children come to understand and uphold such complex abstractions. Language is obviously involved. Also significant is our ability to realize that other people have intentions and desires, a capacity that psychologists call "theory of the mind." Responsiveness to social force—another psychological trait unique to our species—is an additional potent motive for adhering to shared beliefs and values. Again, imitation does not come into it. We do not and cannot imitate justice. Rather we come to understand it slowly through conversations, formal teaching, reading books, watching films, and the like.

Blackmore argues that this slow accumulation of understanding depends on imitation, but it isn't that simple. Recent neurobiological studies indicate that imitation requires specific messages to be computed in specialized areas of the brain. That means that when as a child I came to understand what a restaurant is, or what justice is, I did so by following a sequence of psychological steps entirely different from those by which I learned to tie my shoelaces.

Schemas and social constructions arise out of the operation of memory and abstraction. They have nothing to do with "learning to do an act from seeing it done." The acceptance and spread of ideas through society—especially an ideology such as justice—are slow, unpredictable and difficult to measure, and certainly do not fit within the restrictive theory of memes. Culture, as a collective of human brains and minds, is the most complex phenomenon on earth. We will never understand it if we approach it in a simpleminded way.

HENRY PLOTKIN is a psychobiologist at University College London, where he has worked since 1972. The author of two books on evolution and cognition, he is currently writing a third, on the evolution of culture.

DUSAN PETRICIC

scan studies might compare people carrying out actions with others copying them. Contrary to common sense, this theory postulates that imitation is the harder part—and also that the evolutionarily newer parts of the brain should be especially implicated in carrying it out. In addition, within any group of related animal species, those with the most ability at imitation should have the largest brains. The scarcity of imitation in animals limits the amount of data available, but species of birds, whales and dolphins could be analyzed and compared with this prediction.

Experimental Tests

If language developed in humans as a result of meme-gene coevolution, linguists should find signs that grammar is optimized for transmitting memes with high fecundity, fidelity and longevity, rather than for conveying information on specific topics such as hunting or for forming social contracts. Social psychology experiments should show that people preferentially copy more articulate people and find them more sexually attractive than less eloquent people.

Other predictions can be tested by mathematical modeling and computer simulations, which many researchers have used to model evolutionary processes. The addition of a second, faster replicator to a system should introduce a dramatic change, analogous to the appearance of memes and the human brain's expansion. The second replicator should also be able to control, and even stop, the evolution of the first. Such models might then be used to understand in greater detail the coevolution of memes and genes. In addition, the idea that language could spontaneously emerge in a population of imitating creatures could be tested with simulations of noisy imitating robots.

EMOTIONS ROUSED BY IDEOLOGICAL SYMBOLS suggest the importance that memes hold for us and the power they have over our behavior.

Memetics is a new science, struggling to find its place and with many critics. Some of these critics have simply failed to grasp the idea of a replicator. We need to remember that memes, like genes, are merely bits of information that either succeed in getting copied or do not. In this sense, but no other, memes can be said to be "selfish" and to have replicator power. Memes are not magical entities or free-floating Platonic ideals but information lodged in specific human memories, actions and artifacts. Nor are all mental contents memes, because not all of them were copied from someone else. If all your memes were removed, you would still have many perceptions, emotions, imaginings and learned skills that are yours alone, that you did not acquire from anyone else and that you can never share with another.

A common objection is that memes are very different from genes. And so they are. They suffer (or benefit) from much greater mutation rates, and they are not locked into a system as rigidly prescribed as DNA replication and protein synthesis. Memes are best thought about not by analogy with genes but as new replicators, with their own ways of surviving and getting copied. Memes can be copied all over the place, from speech to paper to book to computer, and to another person.

Yet many more potential criticisms remain, and much work is still to be done. In the end, memetics deserves to succeed only if it provides better explanations than rival theories and offers valid and testable predictions. Unlike religions, the great meme-complex of science includes methods for throwing out ideas that are vacuous, nonsensical or plain wrong. It is against these criteria that memetics, quite rightly, will be judged. SA

The Author

SUSAN BLACKMORE was infected by the meme meme in 1995 by Daniel Dennett's book *Darwin's Dangerous Idea* and by an essay on memes and consciousness by one of her Ph.D. students. The concept took root: by promising to transform the understanding of the human mind, it caused Blackmore to devote many of her meme-generating resources to further study and propagation of the idea. Her status as a source of scientific memes is embodied in her position as reader in psychology at the University of the West of England, Bristol. The Committee for the Scientific Investigation of Claims of the Paranormal (CSICOP) has awarded her the Distinguished Skeptic Award for her studies of near-death experiences and her suggestion that tales of alien abduction are generated by people trying (with the wrong set of memes) to make sense of a form of sleep paralysis.

Further Information

THE SELFISH GENE. Richard Dawkins. Oxford University Press, 1976. (Revised edition with additional material, 1989.)

CULTURE AND THE EVOLUTIONARY PROCESS. Robert Boyd and Peter J. Richerson. University of Chicago Press, 1985.

DARWIN MACHINES AND THE NATURE OF KNOWLEDGE. Henry C. Plotkin. Penguin, 1993.

DARWIN'S DANGEROUS IDEA. Daniel Dennett. Penguin, 1995.

THE MEME MACHINE. Susan Blackmore. Oxford University Press, 1999.

DARWINIZING CULTURE: THE STATUS OF MEMETICS AS A SCIENCE. Edited by Robert Aunger. Oxford University Press (in press).

MARY ANN CHASTAIN *AP Photo*

NABADA:
The Buried City

Excavations in northern Syria reveal the metropolis of Nabada, founded 4,500 years ago. Its elaborate administration and culture rivaled those of the fabled cities of southern Mesopotamia

by Joachim Bretschneider

Palace ruins

The "tell," or mound, at Nabada

Aerial view of ruins

JOACHIM BRETSCHNEIDER (*left* and *top right*); TEAM DRIESSEN-CUNNINGHAM (*right*)

Since the end of the 19th century, archaeologists have strived to uncover the ancient history of the Near East and to trace the region's biblical roots. They focused on the Fertile Crescent between the rivers Euphrates and Tigris in Iraq, where lie the ruins of the ancient city-states of Assur, Babylon, Ur and Uruk. The architecture, tablets and other artifacts of these cities illustrate a tumultuous history that began more than 5,000 years ago. Scholars assumed that only here, in southern Mesopotamia, were the earliest centers of power and the origins of civilization.

Until quite recently, the steppes of northern Mesopotamia (in present-day Syria) were largely neglected. But when the war between Iraq and Iran closed access to southerly sites in the 1980s, archaeologists were forced to pay more attention to peripheral areas. A research team led by Marc Lebeau of the European Center for Upper Mesopotamian Studies and Antoine Suleiman of the Directorate-General of Antiquities and Museums (Syria) began to excavate Tell Beydar, a large mound—or "tell"—rising out of the flat steppes near the Khābūr River. The team also consists of the universities of Leuven (Karel Van Lerberghe), Venice (Philippe Talon), Brussels (Lucio Milano) and Münster (Bretschneider). In these steppes, a tell indicates a long-buried city; after eight years of intensive research, we can now say that Beydar did not disappoint.

Inside the 28-meter-high circular hill archaeologists found a complex almost as large as the citadel of ancient Troy. The city, known in ancient times as Nabada, evidently enjoyed its greatest prosperity during the early Bronze Age, between 2800 and 2200 B.C., and the excavation concentrates on this period. Our aim is to understand the birth of city-states—the metropolises that ruled the surrounding countryside and, sometimes, other cities—in northern Mesopotamia. Complex administration, as evinced by written tablets and seals, evolved at this time.

In places the Bronze Age strata are right at the surface, whereas at others they lie several meters deep. Almost all of the scientific staff and students work for free, and so the summer holidays at European universities dictate the length of an excavation campaign: from the end of August until the end of October. Strenuous work under a ferocious sun, as well as occasional sandstorms and scorpions, fall to our lot. The day begins at 5 A.M. and continues in a miasma of sweat and dust until 1 P.M., when we retire to the shadows to process, record and analyze the objects we have uncovered. Digging resumes in the late afternoon; when the sun sets quickly at 6 P.M. we have to work by electric light or occasionally by candles. At night a cloak of unearthly silence falls on the ancient site.

Early in this century German archaeologist Max Freiherr von Oppenheim demonstrated that the vast and now abandoned spaces of northern Syria were densely inhabited in ancient times. Today archaeologists know of almost 300 tells rising conspicuously from the flat plains. In one of these, Tell Halaf, von Oppenheim found exquisitely shaped and painted pottery dating back to the fifth millennium B.C. He also surveyed some of the more conspicuous circular tells, which cluster around the upper course of the Khābūr River. Such a tell, which he described as "Kranzhügel," or "wreath hill," is surrounded by a ring, the decomposed mud brick of a circular fortification wall. Von Oppenheim suggested that the lower levels of these towns were all created at around the same time and formed a single political or cultural unit, the so-called Kranzhügel culture.

Another explorer in the Khābūr re-

The Cities of Northern Mesopotamia

The steppes of northern Mesopotamia near the Khābūr River in present-day Syria are home to the ruins of numerous ancient cities, including Nabada. Nabada reached its peak of prosperity in the third millennium B.C., when political and economic power in the region were being concentrated in a few large centers: Kish, Lagash, Umma, Ur and Uruk in modern Iraq; and Chuera, Ebla, Mari, Nagar, Nabada and Tuttul in northern Syria.

Nabada and nearby cities probably served as relay stations for caravans traveling the ancient routes (*red lines*) between Anatolia (present-day Turkey) and Babylonia or between Egypt and Mesopotamia. Archaeologists have uncovered a variety of artifacts at Nabada that reflect the city's role as a cultural and trade center (*photographs at right*). —*J.B.*

PATRICIA J. WYNNE AND BRYAN CHRISTIE

PHOTOGRAPHS BY JOACHIM BRETSCHNEIDER

Ivory furniture inlay, 1400 B.C.

Ceremonial vase, 1400 B.C.

gion, Englishman Max E. L. Mallowan, excavated at Tell Brak, a mountain now known to hide the city of Nagar, and at Tell Chagar Bazar, a much smaller site. In 1958 a small team led by Anton Moortgat of the Free University of Berlin began a systematic study of Tell Chuera, a circular hill with a diameter of almost a kilometer. Later excavations at Tell Chuera revealed an urban complex dating to the first half of the third millennium. Although Chuera is made of sun-dried mud brick, as was common in the metropolises of southern Mesopotamia, its temples were constructed on monumental stone terraces.

Who built Tell Chuera or where these people came from remains a mystery. Curiously, Tell Beydar, the only other circular tell to be systematically investigated, is turning out to be quite different.

In Tell Beydar we discerned three main phases of occupation. Researchers date these phases by a combination of techniques: comparing trends in pottery design; measuring the occurrence of radioactive carbon in ashes and other debris of organic origin; and relating names occurring on tablets with those known from other sources. In my view, the dates are uncertain by only about 50 to 100 years, although other scholars differ.

Wreath City

The first and most significant phase began with the founding of Nabada around 2800 B.C. Apparently following a set plan, the builders constructed a circular settlement with a diameter of 600 meters. They protected it by a wall five meters thick, built on a raised embankment. Four gates, now seen as gaps in the buried outer ring, penetrated this wall. Peasants' dwellings and artisans' quarters clung to its inner side as in Europe's medieval cities. About 20 tombs have so far been excavated near the wall, whose elevated base apparently served as a cemetery.

The tombs, some of which also lie beneath the houses, provide a clue to the people's religious beliefs. In one case, a shaft led to the burial chamber, constructed of mud brick and sealed. The dead man, apparently an important official, lay in a fetal position surrounded by weapons, jewelry and pottery to ease his passage to the afterlife. Another grave was supplied with a bronze ax, ceramic jars filled with wheat and many other objects. Whereas soldiers were buried with their weapons, artisans were interred with their work tools; social stratification is evident in the varying richness of the burial gifts.

Some later Babylonian sources describe a tomb as an entrance to the underworld, a place of damnation from which there is no return. The offerings of food and drink by the relatives appeased the spirits of the dead in this dark and gloomy realm. Evidently the citizens of Nabada adhered to a similar belief.

An inner wall 300 meters in diameter protected the heart of the settlement. It is very likely that traders were allowed to spend the night between the two walls, safe from highway robbers but not themselves posing a danger to the sleeping citizens of Nabada. The double wall may also have let peasants from surrounding regions take refuge in the city in times of trouble. From the gates, radiating streets led to the central mound, on which rose a palace. Lining the city's streets were blocks of houses, filling the space between the palace and the inner city wall. Drainage systems evacuated water from houses and courtyards into channels underlying the paved streets and alleys.

In the eastern section of the city stood a remarkable 27-meter-long building whose walls still rise up to three meters. Wide, arched doorways connected the structure's four rooms, a sign of advanced architecture. Low sockets in the walls show that the rooms were fitted with elevated wooden floors for storing grain or wool; the supplies stayed dry thanks to the ventilation underneath them. Storehouses of such size suggest a complex economy.

For its size, the city inside Tell Beydar had surprisingly few houses: most of it consisted of a palace, rising like a fortress on the central, 20-meter-high acropolis. (That is, at least 20 meters of cultural levels and the ruins of several palaces, all from the early Bronze Age, underlie the present excavation.) The royal complex covered about 50 by 60 meters, comprising almost 50 rooms on the ground floor alone. Many of the mud-brick walls are excellently preserved, being up to four meters high and having intact doorways frequently spanned by a vault. A number of rooms boast a fine, white lime plaster on their walls.

A large central courtyard provided easy access to the palace's many rooms. Friezes with clay rosettes decorated the walls of the main rooms, and stairways led up to a throne room (where the king met his subjects) and ceremonial chambers at a higher level. These rooms, which archaeologists recognize by their altars and other characteristic features, had annexes supplied with terracotta

shafts up to 20 meters deep. The small rooms may well have been used for ritual washing and purification, with the shafts providing drainage.

From this large courtyard a smaller one toward the southeast could be reached. From here other staircases led to a higher level, where the living quarters of the ruler may have been situated. All around the perimeter of the palace are storage rooms, still filled with huge ceramic jars that once brimmed with goods. The southern part of the palace featured elaborate wall niches and altar platforms. The ruler may even have been interred here beneath the floor, following an old Syrian practice.

Last year excavators uncovered an intriguing set of terraces. These suggest that this year we are likely to find giant stairways leading up to the palace from the southern gate of the outer wall. They would have formed a steep ramp or stepped pyramid, lined with temples, the whole probably creating a monumental entrance for visiting dignitaries and the elaborate processions accompanying them.

Accounts of Nabada

In 1993 and 1994 excavators made a surprising discovery: a collection of clay tablets with a meticulous record of the palace's daily accounts. Since then, we have found 170 tablets inscribed with a cuneiform script familiar from southern Mesopotamia. Most of the tablets were part of the floor of a house; they had evidently been discarded and reused as building material. Recently we have come across a heap of trash thrown over the palace wall, including many tablets. These written documents date to 2350 B.C., their age making them an important key to the culture.

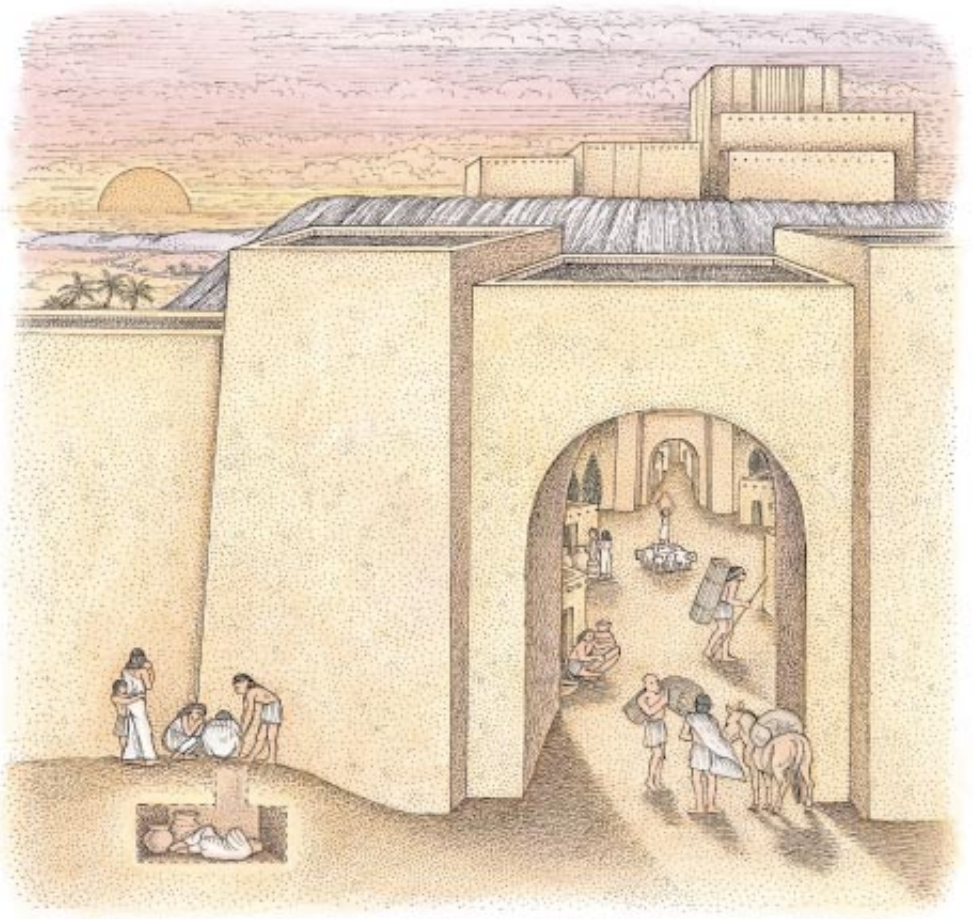
The tablets are curious in one aspect: the script is Sumerian, but the language is Semitic. Philologists assume that Semites migrated into Mesopotamia around the end of the fourth millennium B.C., intermingling with Sumerians and finally dominating Mesopotamian civilization. They adopted the Sumerian script—the only one available—to express their own language. The tablets of Tell Beydar represent the largest collection of Old Semitic texts found in the Khābūr area.

The Semitic royal cities of Mari and Ebla had yielded archives of this period. Mari, discovered in the 1930s, lay halfway between southern and northern Mesopotamia and formed a link between the two cultures. Around 2400 B.C. it ruled much of the region to its north. Ebla, in western Syria, was discovered in 1968 by an Italian team led by Paolo Matthiae of the University of Rome "La Sapienza." Here the archaeologists found an extensive archive of cuneiform tablets, which describe trade relations with Nagar and Mari. Nagar was said to lie on an international trade route between the mountains, which were rich in ores, and southern Mesopotamia, with its major center at Kish.

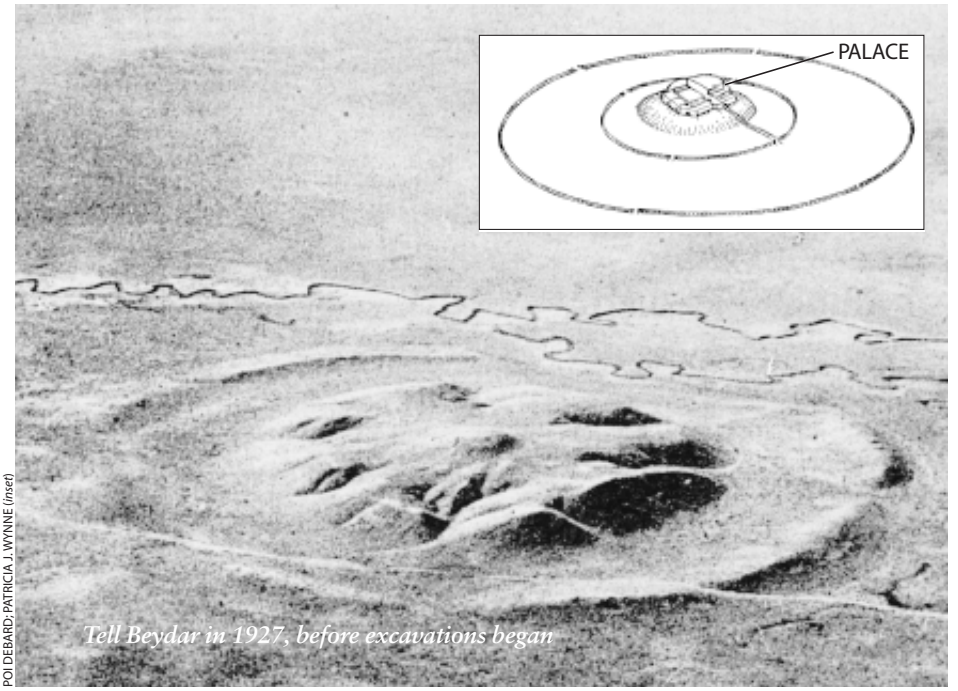
Some of the tablets, deciphered by Walther Sallaberger of the University of Munich, speak of neighboring hamlets administered by Nabada. One group of texts informs us of the care taken of travelers, recording precisely the rations for people and animals. At one time, 11 teams of 44 onagers—a subspecies of the wild ass now found mainly in the north of Iran—had to be taken care of for four days, costing the city a considerable amount of grain. From other texts, Sallaberger concluded that onagers of high quality were bred in the region of Nabada and traded over hundreds

Life inside Nabada's Walls

Nabada was founded around 2800 B.C., with the construction of a circular settlement 600 meters in diameter. The early inhabitants protected their town with a wall five meters thick built on a raised embankment (*illustration below*). Even before the mound, or tell, was excavated, the remains of the city's outer walls and the central mound were evident in the Syrian desert (*top photograph at right*). About 20 tombs have been excavated from this embankment; in one tomb, a shaft led to the burial chamber, which was made of mud brick and then sealed. The dead man, apparently an important official, lay in a fetal position surrounded by gifts such as weapons, jewelry and pottery, which were thought to ease his passage to the afterlife (*bottom photograph*). An inner wall 300 meters in diameter protected the heart of this settlement: rising from a central 20-meter-high mound stood Nabada's palace (*inset*). The royal complex covered about 50 by 60 meters, comprising almost 50 rooms on the ground floor alone. A large central courtyard provided easy access to the palace's many rooms. —*J.B.*

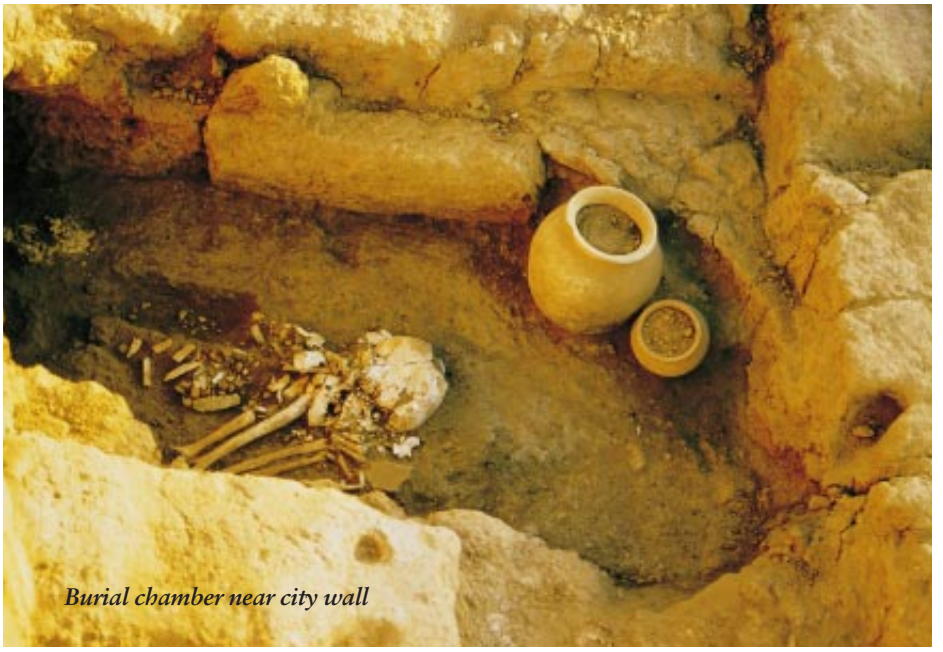


PATRICIA J. WYNNE



Tell Beydar in 1927, before excavations began

POI DEBARD; PATRICIA J. WYNNE (*inset*)



Burial chamber near city wall

JOACHIM BRETSCHNEIDER

of kilometers—as far as the city of Ebla.

Another text mentions the king of Nagar, who apparently ruled Nabada around 2350 B.C. This king visited the "province"—that is, Nabada—on occasions such as council meetings and ritual celebrations. A recently unearthed tablet lists delicacies provided to a woman named Paba, possibly the queen of Mari and spouse of King Iblul-il. Her visit illustrates the city's far-reaching political connections.

Daily activity in this ancient city is also revealed in the impressions made by seals. High officials in the palace possessed finely carved stone cylinders that they rolled over gobs of clay to seal doors, containers and documents [*see illustration on next page*]. We found many such seal impressions in Tell Beydar, on pots and doors of storage rooms and also at the entrances of the throne room and the temple. These rooms may have been opened only on special occasions. Many of the sealings are miniature masterpieces depicting celebrations, lively traffic on trade routes, war and diplomatic activity.

Historians now know that during the first half of the third millennium B.C., political and economic power in Mesopotamia were being concentrated in a few large centers. Thus, cities such as Kish, Lagash, Umma, Ur and Uruk in present-day Iraq and Chuera, Ebla, Mari, Nagar, Urkesh, Nabada and Tuttul in northern Syria came into being. (The earliest levels of Troy in northwestern Asia Minor and the early occupation of Byblos on the Lebanese coast also belong to this era.) Each of these cities contained fortification walls, palaces, storage areas and temple complexes.

The economic and political structure of southern Mesopotamia is relatively well understood. Burgeoning agricultural production allowed livestock to be raised communally. The surplus of food enabled specialized artisanal and administrative skills to flower, and a flourishing trade in raw materials such as copper spurred the evolution of centralized authority and power—as well as warfare. Perhaps the most far-reaching innovation was the craft of writing.

It is not clear that the northern cities followed a similar pattern. As in the south, the larger cities probably controlled extensive regions with vital trade routes. Nabada and its neighboring cities seem to have functioned as relay stations for caravans traveling the ancient routes between Anatolia (present-day Turkey) and Babylonia or between Egypt and Mesopotamia. Important families and a council of elders ruled, as in the south, and offered sacrifices to local and regional deities. A growing number of settlements also led to armed disputes over water rights, agricultural and pasture land, and control of trade routes.

Many questions remain. Historians would like to know who founded Nabada and other northern cities, where they came from, what language they spoke, and around what political and moral principles their society was organized. Moreover, Nabada appears to be quite distinct from the other circular tell that was examined, Tell Chuera. (Tell Chuera is, however, older than the levels so far excavated at Tell Beydar, and so a direct comparison cannot be made.) Chuera had monumental stone architecture (not just mud brick as in Nabada), and, more significantly, 15 years of digging have as yet yielded no evidence of writing. Nabada was probably more allied with southern Mesopotamia than with Tell Chuera, which had closer links with the civilization in Turkey.

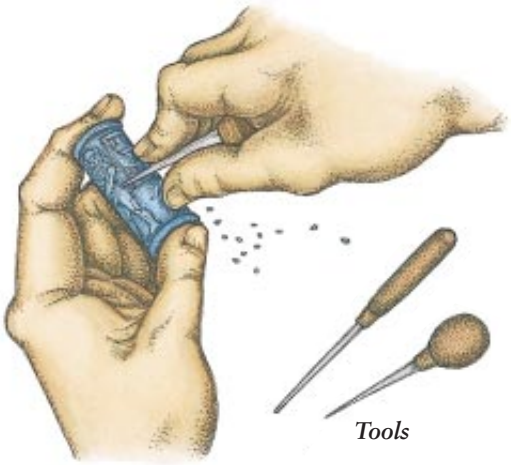
After 500 years of prosperity, Nabada

Writings from Nabada

A picture of what life was like in Nabada can be found in the impressions made by seals. To produce a seal, palace officials rolled a finely carved stone cylinder over wet clay; when the clay dried, it sealed shut pots, documents, even doors. Many of the seals depict celebrations, lively traffic on trade routes, war or diplomatic activity. The photograph at the right shows a seal illustrating the use of wagons in both religious ceremonies and in war. —*J.B.*

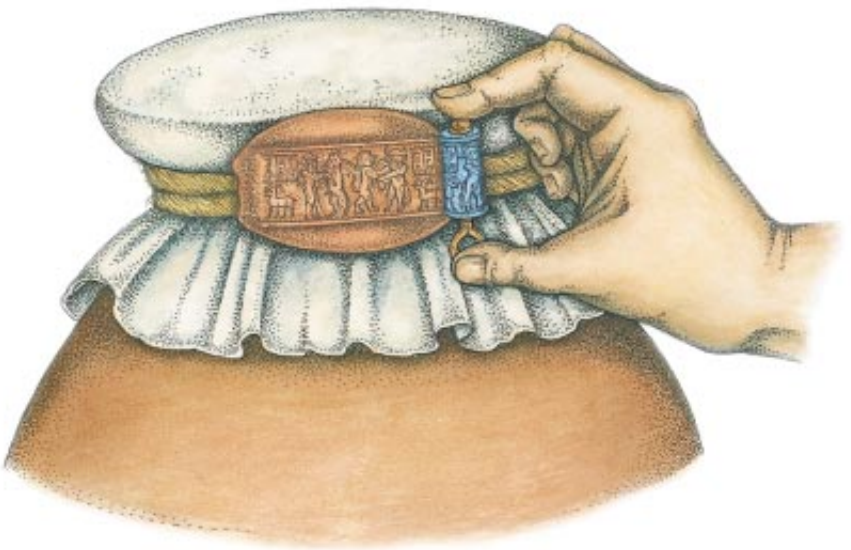
Seal found on a container at Nabada

Carving a stone cylinder



Tools

Rolling a seal onto a jar



was abandoned around 2350 B.C. Why, we do not know; the other known cities of northern Mesopotamia declined at the same time. Perhaps the kings of Akkad, who conquered many of the city-states of Mesopotamia, burned Nabada as they did Ebla. At Nabada, however, no indications of fire or other destruction have been found.

Some architectural changes may provide a clue to this mystery. Shortly before deserting the palace at Nabada, its inhabitants added a sanctuary and renovated several of the ritual rooms. Could this religiously motivated activity have been prompted by some natural event, such as a drought? Or do the deep fissures and cracks in the palace walls suggest a phase of intense earthquakes? Hartmut Kühne of the Free University of Berlin has suggested that a large increase in population caused the natural resources of the region to be overexploited, forcing a migration. It may simply be that a reduction in the number of caravans traveling the trade routes made the northern cities uneconomical.

Tell Beydar was briefly reoccupied a few hundred years later. Around 2250 B.C., however, it suffered from plundering, decline and erosion. In the next millennium the Hurrians, a people who originated in the mountains of East Anatolia, arrived in the region, founding new royal cities such as Urkesh (excavated by Giorgio and Marilyn Buccellati at Tell Mozan). Around 1600 B.C. a Hurrian empire called the kingdom of Mitanni developed in the Khābūr area. At its apex, around 1400 B.C., it extended all the way from the Mediterranean coast to the Zagros Mountains. The Hurrians settled amid the ruins of Nabada, 1,000 years after the fall of the early Bronze Age metropolis.

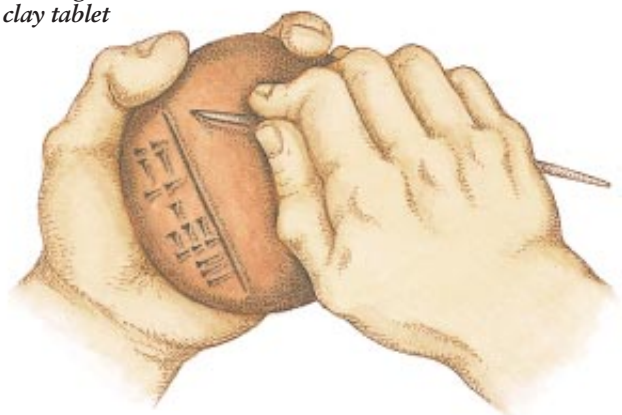
In the lower town surrounding the abandoned palace of Nabada, outside and to the west of the fortification wall, some preliminary excavations have located remains of the Mitanni period. We came across a ceremonial vase decorated with rams' heads and a collection of ivory reliefs depicting lions and bulls that were probably once used as inlays for furniture. The Hurrians did not, however, build a palace, and overall Tell Beydar was far less important at

A New Language

The writings found at Nabada are unusual: the script is Sumerian, but the language is Semitic. The tablets of Tell Beydar (an example is shown below) represent the largest collection of Old Semitic texts found in the Khābūr area. Language experts have translated the symbols; a sample dictionary is shown below. —*J.B.*

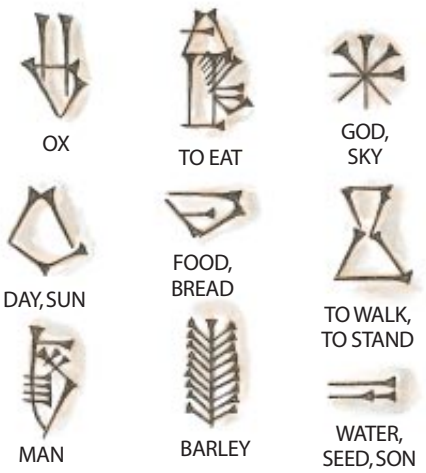
Tablet written in a Semitic language, with Sumerian script

Carving a clay tablet



Reed stylus

Cuneiform



this time than in its heyday as Nabada.

The empire of Mitanni perished in its turn, and in the eighth century B.C. Assyrians conquered the Khābūr region. From Assur, their capital city to the northwest of Babylon, these warlike people increased in influence after 1100 B.C. For a brief period, they even advanced as far as Egypt. The Assyrians were feared by their enemies: they are said to have deported entire populations and slaughtered opponents without respect to sex or age.

The invaders settled on top of the earlier Hurrian occupation in the lower regions of Tell Beydar. Their stay was brief; Babylonians and Medes (from what was then Persia) combined their forces and attacked them. "All who hear the news of you clap their hands at your downfall," wrote the prophet Nahum in the Old Testament when Nineveh, the last and greatest capital of the Assyrian Empire, was reduced to ashes in 612 B.C.

With the fall of the Assyrian empire, Tell Beydar once again returned to dust and silence. SA

The Author

JOACHIM BRETSCHNEIDER studied archaeology and Oriental studies at the Westfälische Wilhelms University in Münster, Germany. He is an assistant professor at the Catholic University in Leuven. Bretschneider is leading the German team—from the Altorientalische Seminar of the University of Münster—of the Euro-Syrian excavations at Tell Beydar.

Further Information

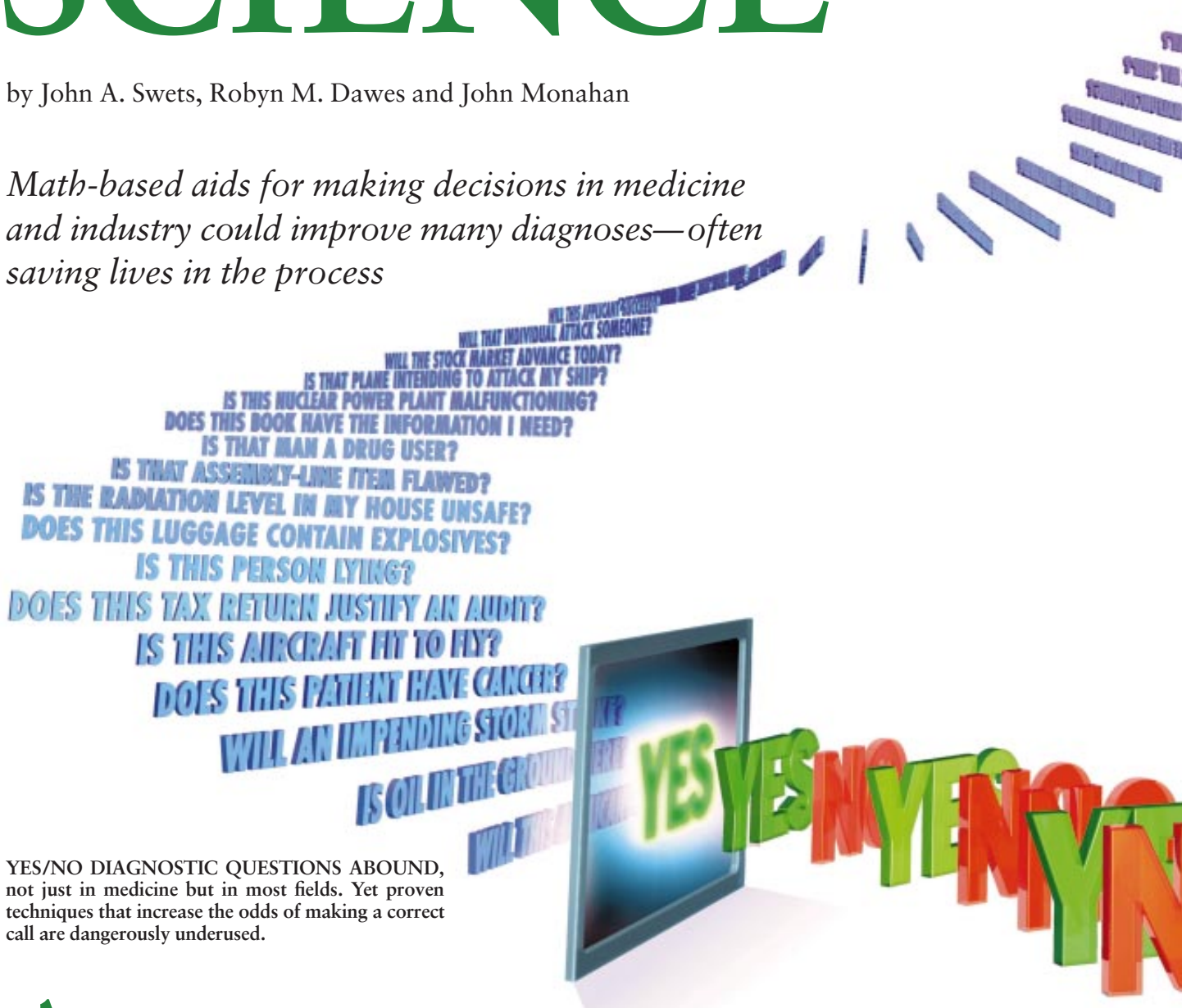
EARLY MESOPOTAMIA: SOCIETY AND ECONOMY AT THE DAWN OF HISTORY. J. N. Postgate. Routledge, 1994.
THE STORY OF WRITING: ALPHABETS, HIEROGLYPHS AND PICTOGRAMS. Andrew Robinson. Thames and Hudson, 1999.

PHOTOGRAPHS BY JOACHIM BRETSCHNEIDER; ILLUSTRATIONS BY PATRICIA J. WYNNE

Better DECISIONS through SCIENCE

by John A. Swets, Robyn M. Dawes and John Monahan

Math-based aids for making decisions in medicine and industry could improve many diagnoses—often saving lives in the process



YES/NO DIAGNOSTIC QUESTIONS ABOUND, not just in medicine but in most fields. Yet proven techniques that increase the odds of making a correct call are dangerously underused.

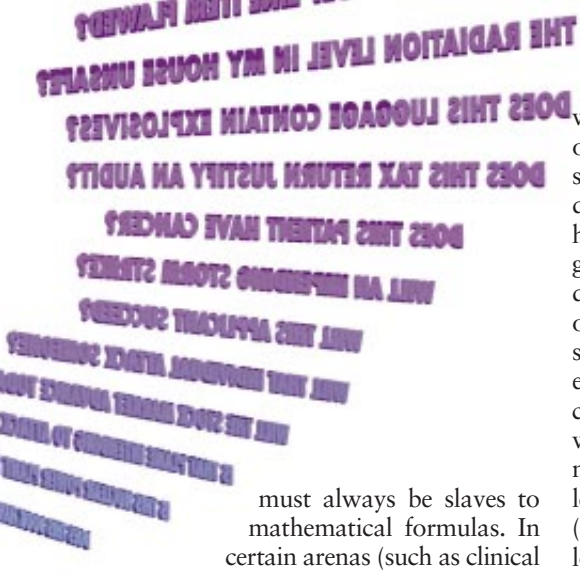
SLIM FILMS

A physician stares at a breast x-ray, agonizing over whether an ambiguous spot is a tumor. A parole board weighs the release of a potentially violent criminal. A technician at an airport worries over a set of ultrasound readings: do they suggest a deadly crack in an airplane's wing?

All these people are grappling with diagnostic decisions. In spite of incomplete or ambiguous evidence, they must determine whether or not a certain condition exists (or will occur). Such problems abound in health care, public safety, business, environment, justice, education, manufacturing, information processing, the military and government. And the stakes can be high. In many cases, a wrong verdict means that people will die.

Perhaps surprisingly, the diagnostic decision-making process turns out to be essentially the same across fields. Hence, methods that improve the process in one industry can usually serve in others. At least two such methods are already available. Sadly, though, they remain unknown or unused in many realms. One increases accuracy, enhancing the odds that any given decision will be the correct one. The other improves the "utility" of a decision-making approach, ensuring that the number of true cases found does not come at the cost of an unreasonable number of false positive diagnoses ("false alarms"). These methods are statistical, but math phobics have nothing to fear; the basic logic is easy to grasp.

No one is saying that diagnosticians



must always be slaves to mathematical formulas. In certain arenas (such as clinical medicine and weather forecasting), objective tools may function best as "second opinions" that inform a reviewer's decisions but do not have the final word. In other fields, however, statistical analyses have frequently been found to be more accurate than subjective judgments, even those made by highly experienced professionals.

We focus in this article on diagnoses that hinge on a choice between just two alternatives—yes or no (Is a tumor present? Is an airplane wing defective?). Certainly the world is full of problems involving a wider range of options, but serious yes/no decisions are prevalent.

Tools of the Trade

If diagnostic tests always produced straightforward answers, no one would need statistical decision-making tools. In reality, though, the raw results of diagnostic tests usually have to be interpreted. In a simple example, the fluid pressure in the eye is measured to detect whether a person has glaucoma, which robs vision by damaging the optic nerve and other parts of the eye. A very low score clearly means the eye is healthy, and a high score signifies glaucoma. But scores in between are ambiguous, unable to indicate which patients have the condition and which do not.

Statistics can clear some of that fog. For argument's sake, assume that pressure is the only diagnostic measure available for glaucoma. Assume, too, that pressures below 10 on the standard measuring scale always signify good health, pressures over 40 always signify disease, and readings between 10 and 40 can occur in affected as well as healthy eyes.

To cope with this ambiguity, analysts would first identify a large population of individuals whose scores on the pressure test were known. Then they would determine which people went on to have vision problems characteristic of glaucoma within a set period and which did not. And they would calculate the odds that people having each possible score will have glaucoma. Finally, guided by those probabilities (and by other considerations we will discuss), they would set a rational cut point, or diagnostic threshold: scores at or above that level would yield a positive diagnosis ("the patient has glaucoma"); scores below would yield a negative diagnosis ("the patient does not have glaucoma").

Of course, single diagnostic tests may not be as informative as a combination. To enhance the accuracy of a diagnosis, analysts can combine data from many tests that each provide unique information, giving greater weight to measurements that are most predictive of the condition under study. The mathematical algorithms that specify the best tests to include in a diagnostic workup and that calculate the likelihood, based on the combined results, that a condition is present are known as statistical prediction rules (SPRs).

Totally objective data, such as pressure readings, are not the only features that can be incorporated to enhance the accuracy of statistical prediction rules; subjective impressions can be quantified and included as well. They can be objectified, for instance, by making an explicit list of perceptual criteria (such as the size and irregularity of a possibly malignant mole) that can be rated according to a scale, perhaps from one to five.

If more than one statistical prediction rule is available, decision makers have to determine which ones are most accurate. This challenge, too, can be met objectively. The overall accuracy of prediction rules can be evaluated by reviewing what are called ROC (receiver operating characteristic) curves. Such curves were first applied to assess how well radar equipment in World War II distinguished random interference (noise) from signals truly indicative of enemy planes.

Programs that generate ROC curves consider what will happen if a particular raw score on a diagnostic test (or set of tests) is selected as the diagnostic threshold for a yes/no decision. What percent of individuals who truly have the condition in question will correctly be deemed to have it (true positive decisions, or "hits")? And what percent of individuals free of the condition will mistakenly be deemed to have it (false positive decisions, or false alarms)?

Then, for each threshold, the programs plot the percentage of true positives against the percentage of false positives. The result is a bowed curve, rising from the lower left corner, where both percentages are zero, to the upper right corner, where both are 100. The more sharply the curve bends, the greater the accuracy of the rule, because the number of hits relative to the number of false alarms is higher.

Obviously, true positives and false positives are not the only outcomes possible. A yes/no diagnosis based on any particular threshold will also generate true negatives (individuals are correctly deemed to be free of the condition being evaluated) and false negatives, or "misses" (individuals are incorrectly deemed to be free of the condition). But these results are the exact complements of the others and thus can be ignored when constructing ROC curves. A true positive rate of 80 percent, for instance, automatically means that the miss rate is 20 percent.

Given that few diagnostic methods are perfect at sorting individuals who have a condition from individuals who do not, institutions have to decide how important it is to find all or most true positives—because more true positives come at the cost of more false alarms. That is, they need to set a threshold that makes good sense for their particular situation.

Returning to our glaucoma example, clinicians who looked only at pressure could find virtually every case of glaucoma if they chose a very "lenient" diagnostic cutoff—say, a score of 10. After all, the test sample revealed that virtually everyone with glaucoma has a score above that level. Yet that cutoff would result in many healthy people being told they were ill; those people would then be subjected unnecessarily to both worry and treatment. To minimize such errors, clinicians could instead set a rather strict diagnostic threshold—an eye pressure of 35, perhaps; very few healthy people in the sample had pressures that high. But this strict criterion would miss more than half of all affected individuals, denying them treatment.

In setting a threshold, decision makers weigh such issues as the consequences of misses and false alarms and

the prevalence of the problem under consideration in the population being tested. Fortunately, some rules of thumb and mathematical aids for finding the optimal cutoff point have been developed. For instance, a high prevalence of a problem in a population or a large benefit associated with finding true cases generally argues for a lenient threshold; conversely, a low prevalence or a high cost for false alarms generally calls for a strict threshold.

Rules Come to Life

Although statistical prediction rules and ROC curves are often sorely underused by diagnosticians, real-life examples of their value abound. One of the most dramatic illustrations comes from psychiatry.

Increasingly, psychiatrists and clinical psychologists are asked to determine whether incarcerated or disturbed individuals are likely to become violent. People who seem most likely to endanger others need to be identified and treated for their own good and for others' safety. At the same time, interfering in the lives of people who do not need care is unacceptable.

Disconcertingly, in 1993 the most sophisticated study of clinicians' unaided assessments uncovered a startling lack of accuracy. Clinicians who diagnosed consecutive patients coming to the emergency department of a metropolitan psychiatric hospital proved no more accurate than chance at predicting which female patients would commit violence in the community within the next six months. Their success rate with male patients was only modestly better.

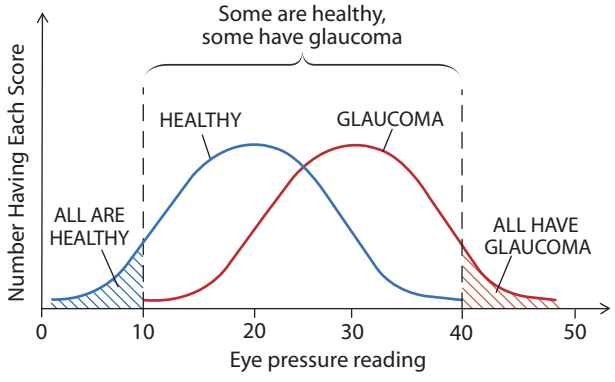
In response to such findings, a number of statistical prediction rules were developed for assessing the probability of violence. One of the most studied is the Violence Risk Appraisal Guide (VRAG), which measures 12 variables, among them scores on a checklist of features indicative of psychopathy and assessments of maladjustment in elementary school.

In a test of the rule's ability to predict whether criminals being discharged from a maximum-security hospital would commit violent acts over the next several years, the VRAG divided the subjects into two categories of risk: "high" and "low." Fifty-five percent of the high-risk group but only 19 percent of the low committed a new violent offense—an accuracy level well above that of chance. And a newer statistical prediction rule proved to be even better at forecasting violence in noncriminals about to be discharged from psychiatric facilities. Nevertheless, interested parties continue to disagree over whether

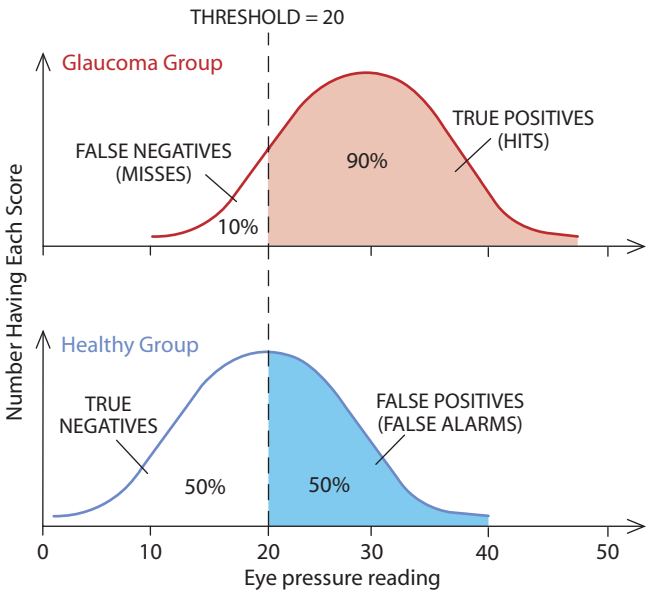
Better Decision Making, Step by Step

How can decision makers ensure that the diagnostic tests they use are as accurate as possible, doing the best job of distinguishing individuals who have a condition from those who do not? A major way involves constructing so-called ROC (receiver operating characteristic) curves. This approach is best described by example. Imagine the steps an analyst might take to evaluate how well glaucoma is diagnosed by measuring the fluid pressure in patients' eyes.

STEP 1 Find a large sample population of people whose eye pressure level and glaucoma status are known. Separate those who are healthy from those with glaucoma, and plot the number of individuals with each pressure level. The graph for this hypothetical population reveals that pressure readings in the 10 to 40 range cannot conclusively distinguish healthy people from those with glaucoma.



STEP 2 Calculate the probability that a "yes" diagnosis at or above any given score, or threshold, would be correct for a new patient. Find such probabilities by determining the fraction of patients in the sample population who would have been properly diagnosed if that threshold were applied. Below, the area under the curves represents 100 percent of each population. If the threshold were 20, 90 percent of people who truly had glaucoma would be diagnosed correctly (true positives), and 50 percent of healthy people would be incorrectly diagnosed as having the condition (false positives).



JENNIFER JOHANSEN

clinicians should treat such rules as advisory or make decisions based solely on the statistics.

Better Cancer Diagnoses

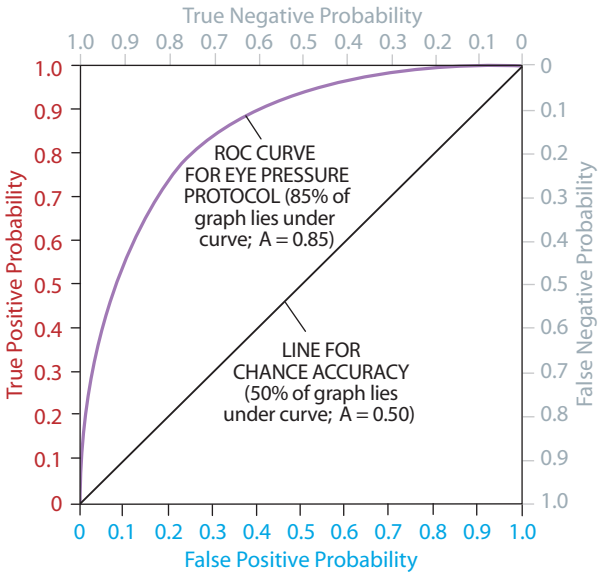
Statistical prediction rules have also had impressive success in studies aimed at helping radiologists diagnose breast cancer. In one such investigation, radiologists in community hospitals evaluated mammograms in their usual, subjective way. Months later they examined the same mammograms according to a checklist of perceptual features (such as how fuzzy the borders of a mass seem to be) developed by radiologists who specialize in reviewing mammograms. Then a statistical prediction rule converted the ratings into probability assessments indicating the likelihood for each patient that breast cancer was present. The radiologists reviewed these probabilities but ultimately made their own judgments. The extra data helped considerably. General radiologists who took the statistical data into account became more accurate, reaching the precision of specialists who had used the checklist.

Physicians who treat prostate cancer are already making extensive use of statistical prediction rules. One rule in particular is getting a serious workout. Once a man is "clinically" deemed to have cancer of the prostate gland (on the basis of a checkup, a simple needle biopsy and noninvasive tests), the question of the best treatment arises [see "Combating Prostate Cancer," by Marc B. Garnick and William R. Fair; Scientific American, December 1998]. Neither surgery to remove the affected gland nor radiation focused tightly on it (to limit side effects) will eliminate the tumor if it has grown beyond the gland or has spread to other parts of the body. Hence, physicians strive to determine the status of the tumor before any treatment is attempted. Unfortunately, a great many tumors that initially seem to be confined to the prostate later turn out to have been more advanced.

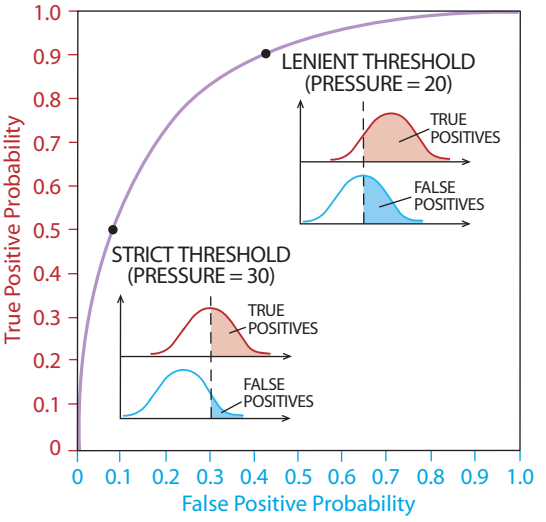
For years, doctors had few good ways of predicting which patients truly had confined disease and which did not. More recently, however, doctors and patients have been able to gain a clearer picture by consulting probability tables published in the May 14, 1997, issue of the *Journal of the American Medical Association*.

The researchers who created the tables knew that three assessments each had independent predictive value: the tumor's "clinical stage" (a determination, based on noninvasive tests, of tumor size and spread), the level in the blood of a specific protein (PSA, or prostate-specific anti-

STEP 3 Construct an ROC curve by plotting, for each potential threshold, the rate of true positives against the rate of false positives. A straight line would signify that the diagnostic test had 50/50 odds of making a correct diagnosis (no better than flipping a coin). As curves bow more to the left, they indicate greater accuracy (a higher ratio of true positives to false positives). Accuracy (A) is indexed more precisely by the amount of area under the curve, which increases as the curves bend. Our glaucoma protocol is moderately accurate.



STEP 4 If the accuracy is acceptable, select a threshold for yes/no diagnoses. Choose a threshold that yields a good rate of true positives without generating an unacceptable rate of false positives. Each point on the curve represents a specific threshold, moving from the most strict at the bottom left to the most lenient at the top right. Strict thresholds (*bottom inset*) limit false positives at the cost of missing many affected individuals; lenient thresholds (*top inset*) maximize discovery of affected individuals at a cost of many false positives. Which threshold is optimal for a given population depends on such factors as the seriousness of the condition being diagnosed, the prevalence of the condition in a population, the availability of corrective measures for those who are diagnosed, and the financial, emotional and other costs of false alarms.



gen) and the Gleason score (an indicator of tumor aggressiveness, based on microscopic analyses of a biopsy sample). The investigators therefore developed a statistical prediction rule that looked at virtually every combination of results for these three variables and calculated the odds that the initial diagnosis of "no spread" would be correct. Then they listed the probabilities in a user-friendly, tabular form.

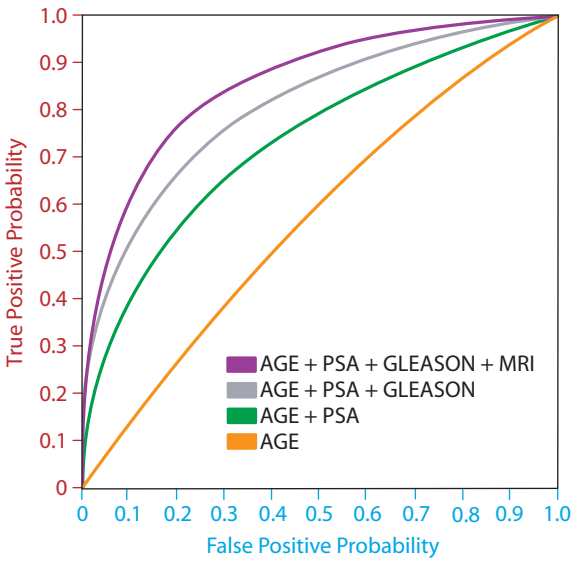
Good Chance of Rain

It would be a mistake to think that only medical practitioners use statistical prediction rules. In fact, meteorologists adopted the tools for weather forecasting more then 25 years ago.

The National Weather Service routinely feeds weather-related data into statistical programs designed to estimate the likelihood that tornadoes, hurricanes, heavy rains and other hazards will arise in different parts of the nation. The weather service then conveys these objective predictions to meteorologists in local areas, who modify the predictions in light of new information or of factors they think the computer programs did not address adequately.

Other groups have embraced the techniques as well—among them, graduate admissions committees at universities. In a typical example, a committee will project first-year grades from two variables—undergraduate grades and graduate school aptitude exams, on the assumption that students scoring above some preselected high level should generally be admitted and those scoring below a specified lower level should generally be rejected. Then the committee will more subjectively evaluate the credentials of applicants who have not been admitted or rejected by the school's statistical prediction rule.

One law school objectively rates two variables that were formerly assessed subjectively: the quality of the student's undergraduate institution and the extent of grade inflation at that institution. Along with the student's grade point average and scores on the aptitude exam required for law school, it considers the mean exam score of all students from the applicant's college who took the test and the mean grade point average of students from that college who applied to law school. The revised formula predicts first-year law-school grades significantly better than the two-variable scheme.



ROC CURVES compared the accuracy achievable by measuring one or more of the following variables to assess whether prostate cancer is advanced: patient age, blood level of PSA (prostate-specific antigen), tumor aggressiveness (represented by the Gleason score) and tumor appearance as judged by magnetic resonance imaging (MRI). The scheme that included all four variables (*top curve*) proved best.

Thorny Thresholds

So far we have highlighted success stories. But the merit of statistical analyses may be best illustrated by examples of failure to apply them for setting rational diagnostic thresholds—such as for tests that detect the human immunodeficiency virus (HIV), the cause of AIDS.

HIV screening relies initially on a relatively simple test that detects the presence of anti-HIV antibodies, molecules produced when the immune system begins to react against HIV. Sometimes these antibodies arise for reasons other than the presence of HIV, however. Hence, if the outcome (based on some antibody threshold) is positive, laboratories will run a different, more sophisticated test. This two-test requirement is meant to help limit false positives. The antibody tests are particularly problematic in that, illogically, the several approved tests differ in their accuracies and thresholds. Varied thresholds would make sense if each test were aimed at a distinct population, but that is not the case.

The thresholds are disturbing in another way as well. They were originally set to distinguish clean from tainted donated blood; then they were left unchanged when they were enlisted to identify people infected with the virus. Throwing out a pint of uncontaminated blood because of a false positive is a cheap mistake; sending an alarmed, uninfected person for further HIV testing is not. Worse still, the original thresholds have been applied mindlessly to low-risk blood donors, high-risk donors, military recruits and methadone-clinic visitors—groups whose infection rates vary over an enormous range. For the high-risk groups, the threshold should be set more leniently than for the low-risk populations (to maximize discovery), even if the price is a higher rate of false positives.

Recent years have seen the introduction of confirmatory tests that are more accurate and of HIV therapies that prolong life and health. Consequently, false positive diagnoses are rare these days, and people who are infected with HIV benefit much more from being diagnosed than was true in the past. These advances mean that the diagnostic problem has shifted from whom to call positive to whom to test.

The time has come for doctors to lower their thresholds for deciding when to test; they should not be waiting until patients show up with obvious symptoms of infection. We would even argue that almost every adult should be screened and that federal agencies should take the lead in encouraging such testing.

Objective methods for establishing thresholds are also being dangerously underused in parts of the aerospace industry. This industry must constantly diagnose conditions that are serious but arise relatively infrequently, among them cracked wings and life-threatening hazards during flights. The costs of missing a cracked wing are large and obvious: many passengers may die if the plane crashes. On the other hand, a false-positive decision takes a plane out of service unnecessarily, potentially causing inconvenience and lost income. At first blush, the benefits and costs point toward a lenient threshold, favoring lives over dollars. Yet such cracks occur rarely; therefore a lenient threshold yields an unworkable number of false posi-

JENNIFER JOHANSEN

tives. Unfortunately, no one has yet tackled this issue with the available statistical techniques.

Purchasers of cockpit alarms (such as airlines and the military) have similarly failed to come to grips with how best to set decision thresholds. Alarms go off in flight under many circumstances—when sensing devices determine that another plane is too close, that the plane is getting too near to the ground, that an engine is dying or that wind shear is threatening the landing area. But they cry wolf too often, largely because the sensors are only moderately accurate and because the thresholds set for them are rather lenient. Pilots are reluctant to act on the warnings unnecessarily, because doing so can be quite disruptive. This situation has raised fears that the high number of false alarms will cause pilots to ignore or respond slowly to a real emergency. To date, though, no one has forced manufacturers to consider the false positive rate when they establish alarm thresholds.

A Plea

Clearly, statistical prediction rules can often raise the accuracy of repetitive diagnostic decisions, and formulas for setting decision thresholds can improve the utility of those decisions. But these tools provide other advantages as well. By standardizing the features that are assessed to make a diagnosis, the prediction rules can hasten the speed with which professionals recognize key diagnostic features. They also give decision makers a way to communicate more easily and precisely about impressionistic features. And they can help teach newcomers to a field.

Yet they are often met with resistance, especially if they are seen as replacing or degrading clinicians. Further, diagnosticians want to feel that they understand their own diagnoses and recommendations and that they can give a narrative of their thought processes. The results of a statistical prediction rule may be hard to include in such an account, particularly if the logic behind the analysis is not self-evident.

We understand all these concerns. Nevertheless, the benefits that statistical tools provide surely justify consideration by decision makers who hold others' lives and futures in their hands. SA

A QUESTION OF TASTE

Architects and connoisseurs of wine have invented two of the more offbeat applications of statistical prediction rules. The architectural rule applies to opera houses and was developed by having conductors rate the overall sound quality of 23 facilities. The conductors favored the houses in Buenos Aires, Dresden, Milan and Tokyo. Next, acoustical engineers physically measured several individual acoustical properties in each of the 23 buildings—such as the time delays between directly received and reflected sound, and the diffusion of sound waves caused by irregularities in walls and ceilings. Statistical analyses then revealed which properties combined to give the favored opera houses their exceptional sound and which of the acoustic characteristics were most important. The resulting rule can now guide the construction of future facilities.

COURTESY OF TAK ARCHITECTS, YANAGISAWA, TOKYO

OPERA HOUSE in Tokyo's New National Theater has stellar acoustics that can be emulated, thanks to a statistical prediction rule.

The wine rule predicts the eventual quality of red Bordeaux wines (as measured by auction price) when they are still young and undrinkable. Classically, experts have attempted to predict later quality "clinically," by smelling and tasting the new product. But about 10 years ago, researchers noted that years marked by a dry August and September and a warm growing season yield excellent wines if those years also follow a wet winter.

They then formulated the "Bordeaux equation," a statistical prediction rule that combines weather conditions and years of aging to predict the probability that wine quality will be great years ahead. That equation works quite well, accounting for 83 percent of the variance in price of mature Bordeaux red wines at auction. But it has not met with universal acclaim. "Somewhere between violent and hysterical" is how the reaction of the wine-tasting industry was described in a newspaper report soon after the equation was unveiled. *—J.A.S., R.M.D. and J.M.*

The Authors

JOHN A. SWETS, ROBYN M. DAWES and JOHN MONAHAN recently collaborated on a more technical paper on this topic published in the inaugural issue of a journal, from the American Psychological Society, that reviews psychological research on pressing issues of broad importance (see "Further Information"). Swets is chief scientist emeritus at BBN Technologies in Cambridge, Mass., senior research associate in radiology at the Brigham and Women's Hospital in Boston and lecturer on health care policy at Harvard Medical School. Dawes is the Charles J. Queenan, Jr., University Professor in the department of social and decision sciences at Carnegie Mellon University and author of *Rational Choice in an Uncertain World*. Monahan, a psychologist, holds the Doherty Chair in law at the University of Virginia and is director of the Research Network on Mental Health and the Law of the John D. and Catherine T. MacArthur Foundation.

Further Information

COMBINATION OF PROSTATE-SPECIFIC ANTIGEN, CLINICAL STAGE, AND GLEASON SCORE TO PREDICT PATHOLOGICAL STAGE OF LOCALIZED PROSTATE CANCER. A. W. Partin et al. in *Journal of the American Medical Association,* Vol. 277, No. 18, pages 1445–1451; May 14, 1997.

THINK HIV: WHY PHYSICIANS SHOULD LOWER THEIR THRESHOLD FOR HIV TESTING. Kenneth A. Freedberg and Jeffrey H. Samet in *Archives of Internal Medicine,* Vol. 159, No. 17, pages 1994–2003; September 27, 1999.

PSYCHOLOGICAL SCIENCE CAN IMPROVE DIAGNOSTIC DECISIONS. John A. Swets, Robyn M. Dawes and John Monahan in *Psychological Science in the Public Interest* (supplement to *Psychological Science*), Vol. 1, No. 1, pages 1–26; May 2000.

CONTACT_LENSES

The Hard and the Soft

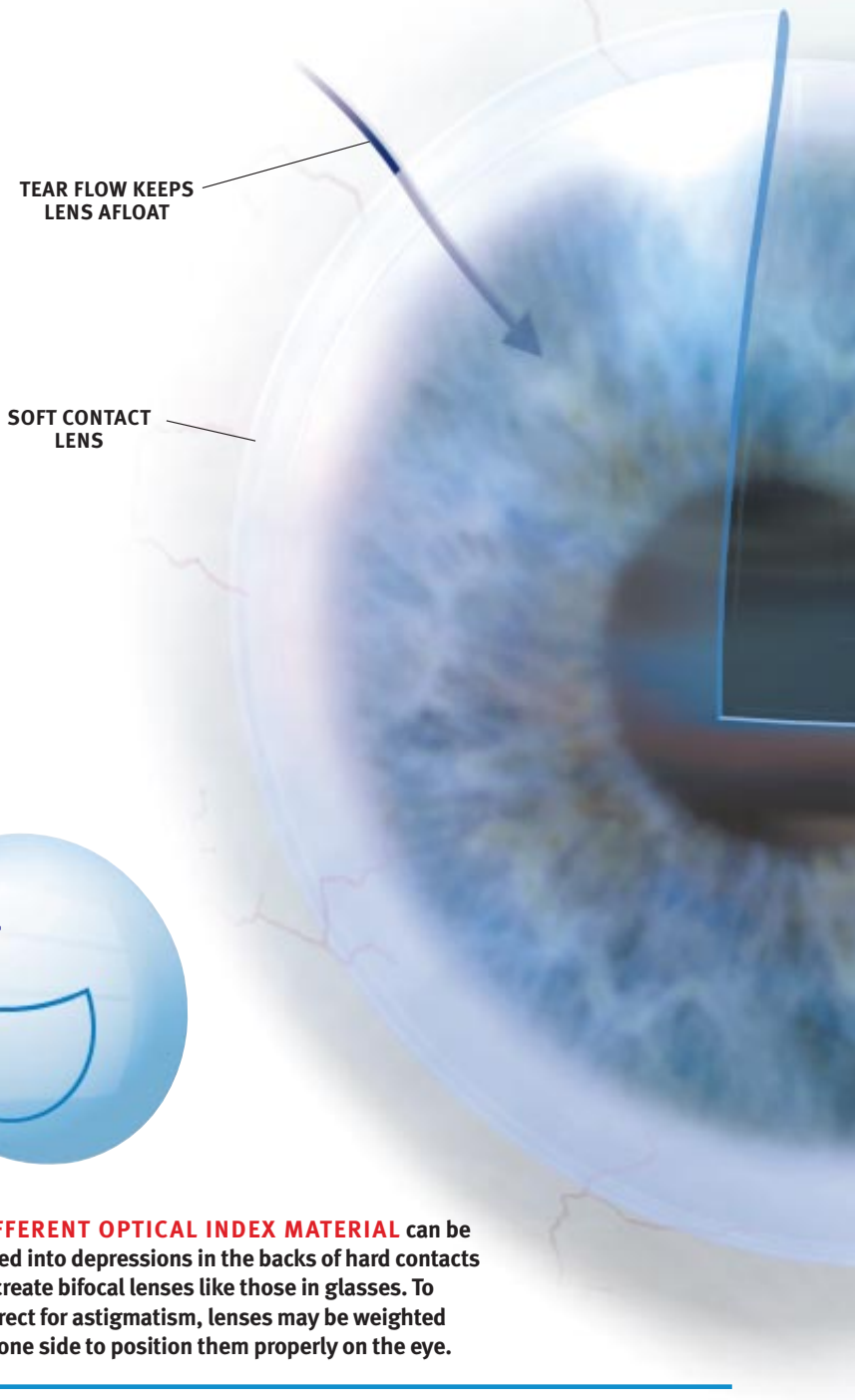
In the eyes of millions of Americans float little plastic disks, bending the light that enters their eyes and making their vision clearer. Last year approximately 34 million people in the U.S. wore contact lenses, according to the Contact Lens Council. The majority wear soft (or "hydrogel") lenses, intending to throw them away after one day, two weeks or several months. These lenses need cleaning daily. "Extended wear" soft lenses can be worn continuously—day and night—for up to a week (the Food and Drug Administration is considering 30-day lenses). Fewer than 15 percent of contact lens wearers use what are called rigid gas-permeable lenses, which can last for years if taken care of properly.

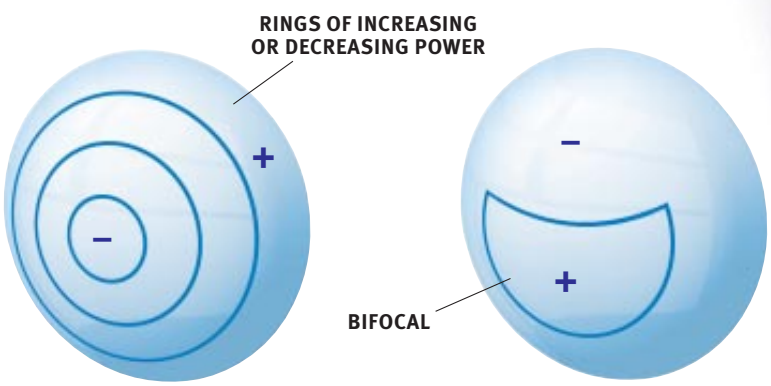
These rigid lenses are made of plastic polymers—such as cellulose acetate butyrate, polyacrylate-silicone or silicone elastomers—that do not absorb or attract water. David T. Grubb, a materials scientist at Cornell University, describes their molecular structure as open, with large gaps through which oxygen can pass. Hydrogel polymers, in contrast, are hydrophilic, absorbing or attracting water in amounts that vary from commercial brand to brand. Water moving through the molecular structure of soft lenses carries oxygen with it.

When the eyelids are closed, contacts block the flow of oxygen to the clear, delicate cornea, hindering routine tissue repair. Some hydrophilic polymer blends carry more oxygen than others, but according to George Grobe of Bausch & Lomb, such blends also encourage proteins in tears to stick to the lens. The protein deposits can shelter pathogens, increasing the chances of infection. The challenge in building an extended-wear lens, Grobe says, is finding the right polymer mix.

—Naomi Lubick, staff writer

SOFT CONTACT LENSES mold to the shape of the eyes, which are not perfectly spherical. (Hard lenses must be carved to fit and will slip if not correctly formed.) The bulge of the cornea makes a contact look a bit like a mountain, with flat foothills, steeper slopes and a conical apex. Soft lenses cover the adjacent limbal and scleral areas of the eye (whereas rigid lenses sit only on the cornea) and, because of their size, may be held in place partly by the eyelids. The liquid film of tears contributes surface (capillary) attraction to hold both soft and rigid lenses in place.

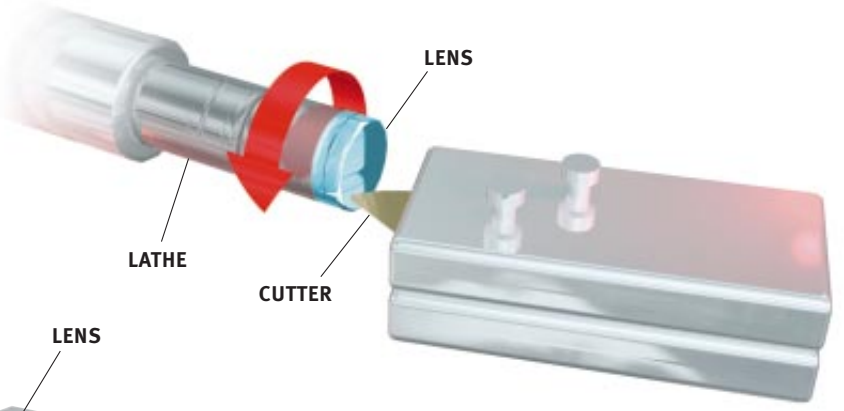


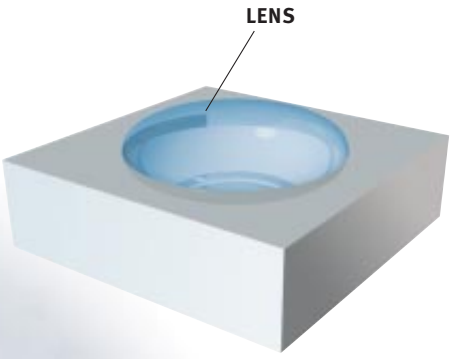


BULL'S-EYE BIFOCAL LENSES, which can be either soft or hard, have concentric rings of material that "correct" vision to varying degrees. To look at an object, the brain automatically chooses whichever view offers the sharpest focus.

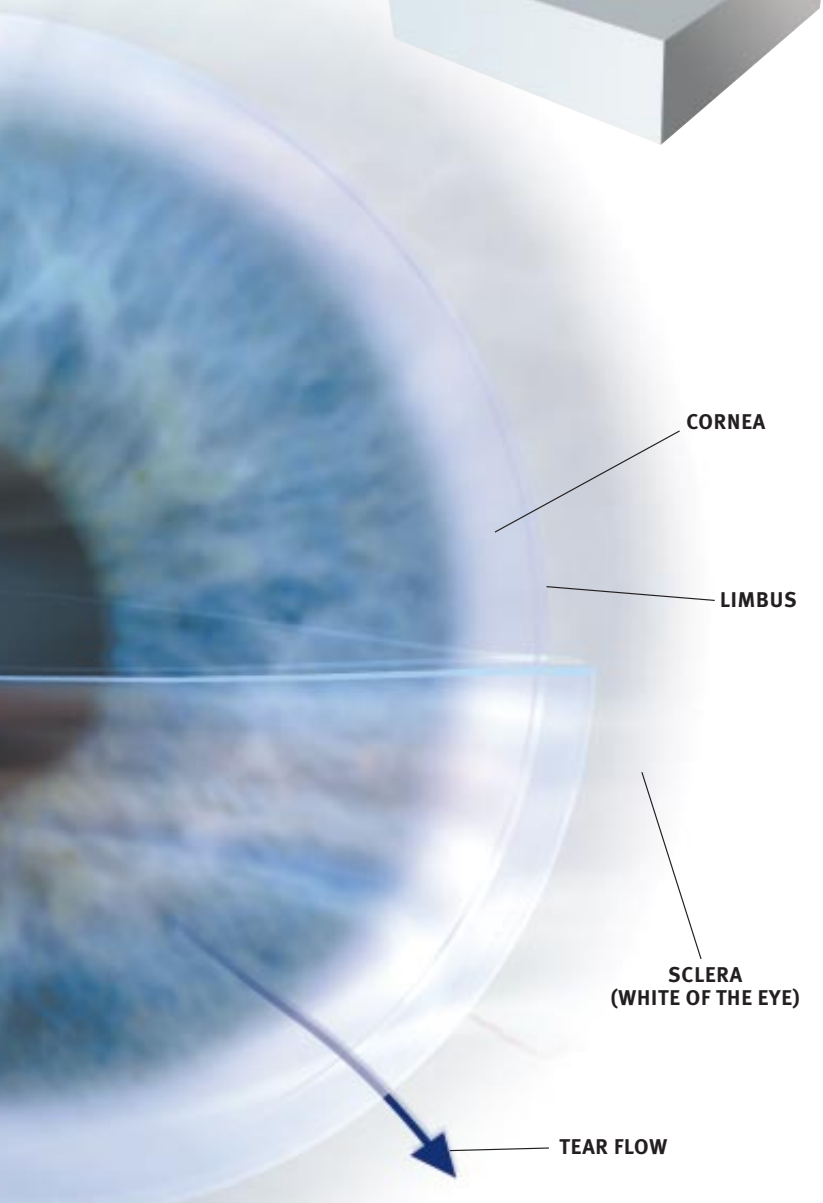
DIFFERENT OPTICAL INDEX MATERIAL can be fused into depressions in the backs of hard contacts to create bifocal lenses like those in glasses. To correct for astigmatism, lenses may be weighted on one side to position them properly on the eye.

COMPUTER-DIRECTED LATHES carve the plastic "buttons" from which rigid gas-permeable lenses and some soft lenses are made. Technicians direct this carving using a topographic map of the patient's eye.





CAST MOLDS can be used to mass-produce soft contact lenses quickly. Ciba Vision, one manufacturer, makes plastic molds from metal master molds for every possible vision type and eye shape. The polymer is injected into the mold and "cured" by ultraviolet light. The lenses are then hydrated and finished.



DID YOU KNOW ...

- Tears are made mainly in the lacrimal gland. Mostly water and salt, they also carry proteins, glucose, potassium, sodium, calcium, chlorine, bicarbonate, ammonium, nitrogen, citric and ascorbic acids, lysozymes, and other substances. The mix varies from person to person. No wonder tears leave behind unwanted deposits on contact lenses—but they also cleanse and protect the cornea.

- Keratitis is inflammation of the cornea. Dozens of variations stem from either hypersensitivity or microbial infection, both of which may be promoted by contact lens wear and are usually the result of poor hygiene. *Pseudomonas* and *Acanthamoeba,* organisms present in tap water, are the most likely pathogens, causing redness, lesions, ulcers and infections. Keratitis is rare: the disease occurs in only 41 of 100,000 daily-wear contact users (but the rate is five times higher in extended-wear users), according to Dr. Robert Davis of the American Optometric Association.

- Contact lenses may also affect the conjunctiva, a thin mucous membrane that lines part of the eye and eyelids. In extended-wear users, conjunctivitis is four to 10 times more likely to occur than in daily-wear users, although it, too, is rare.

ILLUSTRATIONS BY BRYAN CHRISTIE

ELECTRO_BALANCE

Down among the Micrograms

Shawn Carlson describes how to construct a high-precision laboratory scale

I live for Fridays. That's because I usually spend that day hiking through the San Diego badlands with an eclectic assembly of iconoclasts, including several brilliant technologists and some of my dearest friends. We connect through our love of instrumentation and our shared passion for developing inexpensive solutions to various experimental challenges. This common interest leads to friendly rivalries, the results of which often feed this column.

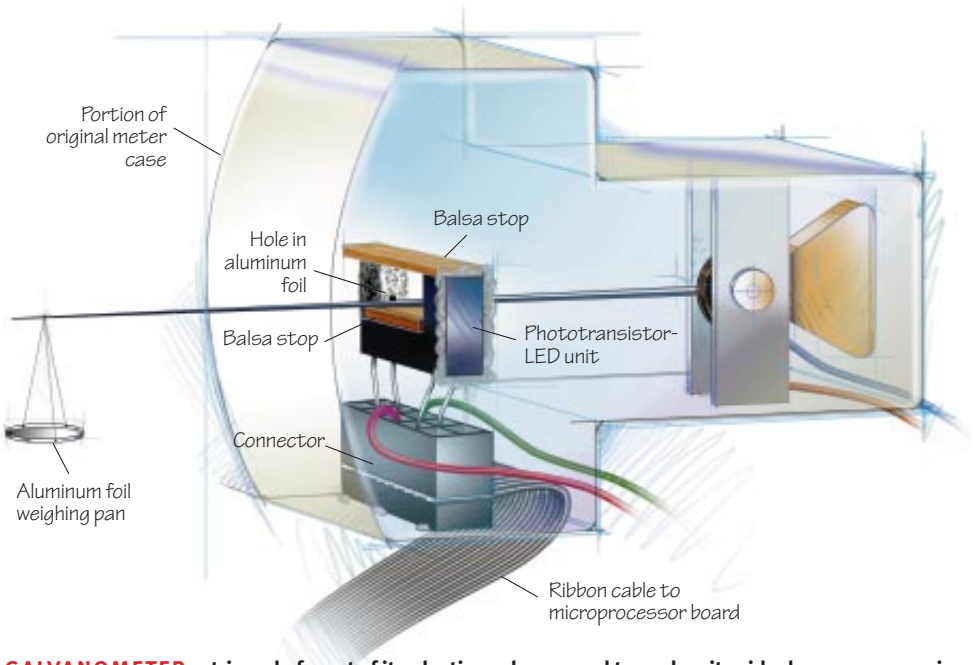
Take for instance the problem of measuring extremely tiny masses. George Schmermund developed a fantastic approach, which I described in these pages in June 1996. George extracted the coil and armature from a discarded galvanometer and mounted them upright, so that the needle of the meter moved in a vertical plane. He then connected the coil to a variable voltage and adjusted it until the needle was exactly horizontal. A tiny mass of known weight placed at the end of the needle pulled it downward. George then increased the voltage until the arm returned to its starting position. Because a heavier mass required a proportionally larger increase in voltage to balance it, the change in voltage indicated the weight of a sample. George's electrobalance was able to weigh masses as small as 10 micrograms (that is, 10 millionths of a gram).

That achievement was stunning enough for me, but recently the organizer of our weekly outings, Greg Schmidt, realized that even this amazing performance could be improved on. Greg's design eliminates the need to adjust the needle manually: the balance automatically zeros (or "tares") and levels itself, and it can continuously track how an object changes in mass—the rate at which a single ant loses water through respiration, for instance. The result is an extremely versatile electrobalance with microgram sensitivity that can be built for less than $100.

Here's how it works. Greg took George's basic design and added an inexpensive microcontroller (a small computer with its central processing unit and memory all on a single chip), instructing it to send 2,000 weak current pulses through the coil each second. The inertia of the armature and needle prevents them from responding to each short pulse, so the deflection reflects the average current in the coil. The individual pulses do, however, seem to be large enough to vibrate the bearings of Greg's galvanometer. He believes that this slight jitter reduces "stiction," the tendency of a bearing to lock in place when it is not moving. This effect seems to account for why an inexpensive meter like his can respond to the tug of such tiny masses.

Greg didn't design his circuit to reduce stiction, though. This feature turned out to be an unforeseen benefit of using "pulse width modulation" to control the average current sent through the coil. With this scheme, the time between successive pulses is kept the same, but the microcontroller varies the duty cycle—the fraction of the cycle during which the current remains on. Pulse trains with short duty cycles energize the coil for only a smidgen of the total time and so can lift only the smallest weights, whereas pulse trains with longer duty cycles can hoist heavier loads. Greg's microprocessor can generate 1,024 different values for the duty cycle. That number sets the dynamic range of the balance. If the maximum current is set so that the apparatus can lift up to one milligram, for example, the smallest detectable mass will be about one microgram.

Such sensitivity is pretty impressive. Yet the microcomputer that runs the show need not be anything special. Indeed, one has a dizzying array of choices to pick from. But if you haven't a clue how to go about selecting and programming a microprocessor, don't worry: Greg developed



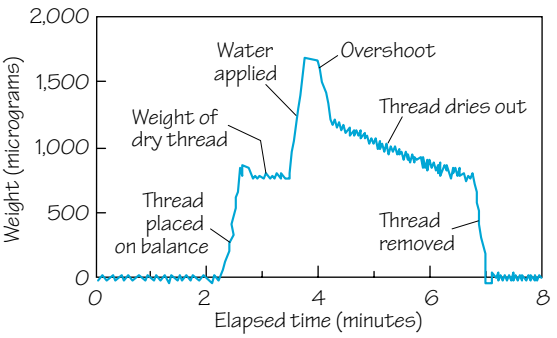
GALVANOMETER, stripped of most of its plastic enclosure and turned on its side, becomes a sensitive scale. A U-shaped optical device glued into place allows an inexpensive microcomputer to apply just enough current to hold the needle level when weight is added.

ALL ILLUSTRATIONS BY DANIELS & DANIELS

his instrument with the novice in mind. He used the Atmel AT 89/90 Series Flash Microcontroller evaluation kit, which includes a fully functional and extremely versatile microcomputer, one that links directly to a personal computer. This kit (model STK-200) includes everything you need to get going and costs less than $50 (see www.atmel.com for a list of suppliers).

Unfortunately for Macintosh users, this system supports only IBM compatibles. In any case, you don't have to program everything from scratch, because Greg developed all the software needed to run the device, including instructions that show the weight in real time on a small liquid-crystal display (catalogue number 73-1058-ND from Digi-Key; 800-344-4539). You can download his code for free from the Web site of the Society for Amateur Scientists (www.sas.org).

As with George's original design, almost any galvanometer plucked from a surplus bin will work. Just make sure that it measures small currents and that its needle tends to stay in place when the unit is rocked rapidly from side to side. Whereas George's prototype required the operator to squint at the needle, Greg's electrobalance senses the position of the needle electronically using a phototransistor and a light-emitting diode, which you can also purchase from Digi-Key (catalogue number QVA11334QT-ND comprises a single unit). Pierce a small piece of aluminum foil with a pin and center the hole on the phototransistor, as shown on page 90.



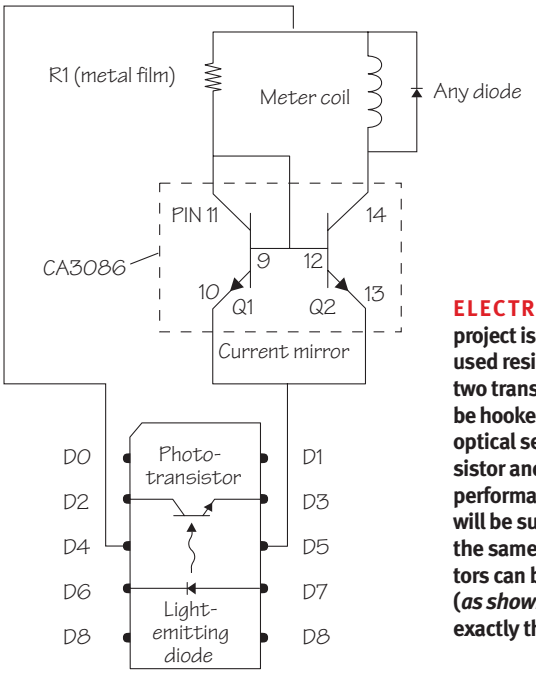
CONTINUOUS RECORDING **of the changing weight of a one-centimeter length of moistened thread demonstrates the versatility of this inexpensive instrument.**

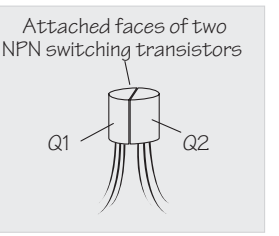
With the foil covering most of the phototransistor, the signal will go from full on to full off very rapidly when the needle interrupts the light from the diode. Attach a sliver of balsa wood as shown to stop the needle exactly at that point.

If too little current is in the coil, the needle will rest on the bottom piece of balsa and block the light. Too much current lifts the needle completely out of the light path. Greg's software uses a sophisticated algorithm to keep the needle balanced between these two states. After the device has been properly calibrated and tared, this pulse width reflects the mass of the sample.

The control circuit that helps accomplish all this magic is shown above. You will need to adjust the value of R1 to set the maximum current to something your meter can handle. The full-scale current might be indicated on the meter. Otherwise, use a variable resistor, a nine-volt battery and a current meter to measure it. Because Greg's galvanometer topped out at five milliamperes, he programmed the microcontroller to create a five-milliampere current by delivering a five-volt pulse across a one-kilohm resistor.

That current is not, however, directed through the coil. Rather it flows through a circuit called a current mirror, which forces an identical current to pass into the coil. This trick dramatically improves the long-term stability of the balance. Why? The resistance of the coil depends on its temperature, which rises whenever electrical energy is dissipated inside it. But the mirror circuit keeps the current constant no matter what the temperature of the coil is.





ELECTRONIC WIRING **required for the project is minimal because the microcomputer used resides on a self-contained board. Only two transistors, a resistor and a diode need be hooked up, in addition to the integrated optical sensor (which contains a phototransistor and light-emitting diode). Although performance of the "current mirror" circuit will be superior if its two transistors reside on the same silicon chip, separate NPN transistors can be used if their casings are attached (*as shown above*) so that they both stay at exactly the same temperature.**

Of course, the resistance of R1 will itself vary somewhat with temperature, which could cause the calibration to drift. So you'll want to use a component with a low temperature coefficient. A 1 percent tolerance metal-film resistor, for instance, typically shifts a mere 50 parts per million for each degree Celsius. You will also need to keep the two transistors in the current mirror at the same temperature to prevent that circuit from drifting. It's best to use a set of matched transistors on a single silicon chip, like the CA3086 (48 cents from Circuit Specialists; 800-528-1417). Otherwise, wire two identical NPN switching transistors together with their casings touching as shown above.

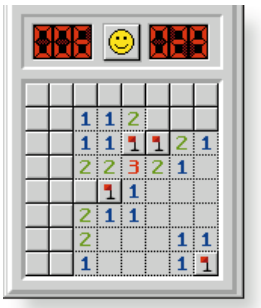
A delightful demonstration of the sensitivity his apparatus achieves is shown in the graph at the left. Greg soaked a centimeter of fine thread in water. He then monitored its weight as the water slowly evaporated. Remarkable. SA

As a service to the amateur community, the Society for Amateur Scientists will supply all the electronics components for this project, including the matched transistors, as well as a set of calibration weights for $60. The package includes the LCD display, but it does not contain a circuit board, the galvanometer coil or the Atmel microcontroller starter kit. To order, call SAS at 619-239-8807. You may write the society at 4735 Clairemont Square PMB 179, San Diego, CA 92117. Or for more information about this and other projects described in "The Amateur Scientist," surf over to www.sas.org and click on the "Forum" button.

COMPUTATION_ALGORITHMS

Million-Dollar Minesweeper

Ian Stewart explains how a computer game can make you rich



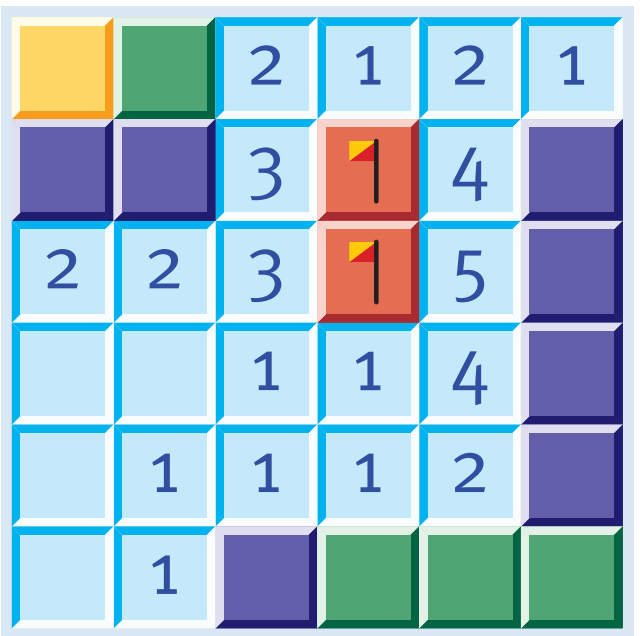
Who wants to be a millionaire? The Clay Mathematics Institute, a nonprofit educational foundation in Cambridge, Mass., is offering million-dollar prizes for the solutions to seven infamous unsolved problems. One is the notorious P vs. NP question. Although it is a fiendishly difficult puzzle, a clever amateur may be able to solve it with the help of Minesweeper, the popular game that runs on most personal computers.

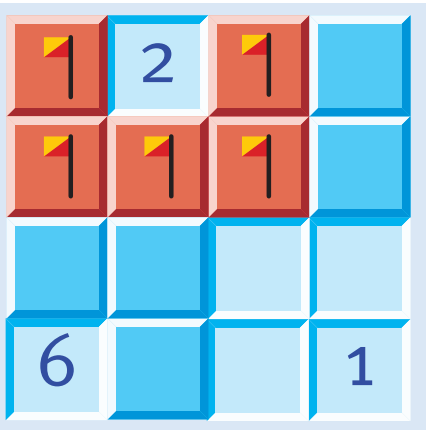
Richard Kaye of the University of Birmingham in England pointed out the connection between the game and the P vs. NP question in a recent article entitled "Minesweeper Is NP-Complete" (*Mathematical Intelligencer,* Vol. 22, No. 2, Spring 2000). First, let's review how Minesweeper is played. The computer starts the game by showing you a grid of blank squares, some of which conceal explosive mines. Your task is to figure out where the mines are without detonating any of them. In your first move, you choose to uncover any square in the grid. If there's a mine underneath it, you're out of luck: the mine is detonated and you lose the game. If there's no mine, however, the computer writes a number in that square, telling you how many mines there are in the eight adjacent squares.

You then use this information to determine the next square you should uncover, and again you either detonate a mine or reveal the number of mines in the adjacent squares. When you have deduced the presence of a mine in a square, you mark it with a little flag. A typical game position after a few moves is shown in the left illustration below. You win the game by locating all the mines in the grid.

Now we turn to the P vs. NP puzzle. Recall that an algorithm is a step-by-step procedure that can be run on a computer to solve a problem. A central question in the mathematics of computation is, How efficiently can an algorithm solve a given problem? In other words, how does the running time—the number of computations needed to produce the answer—depend on the initial data? The main distinction is between problems that are of type P, which stands for polynomial time, and those that are not. A problem is of type P if it can be solved using an algorithm whose running time grows no faster than some fixed power of the number of symbols needed to specify the initial data. If the solution can't be found this way, the problem is non-P. Type P problems can be handled efficiently by computer, whereas non-P problems cannot be solved in a practical manner, because any algorithm used to tackle them will take a ridiculously long time to yield an answer.

You can prove that a problem is of type P by finding an algorithm that solves it in polynomial time. For example, sorting a list of numbers into ascending or descending order is a type P problem, which is why commercial database programs can efficiently sort even very large sequences of numbers. In contrast, the Traveling Salesman Problem—finding the shortest route whereby a salesman can visit every city on some itinerary—is widely believed to be non-P, although this has not been proved. Finding the prime factors of





TYPICAL MINESWEEPER POSITION after the game's first few moves shows two squares flagged as mines (*left*). By examining the numbers that have been uncovered so far, one can deduce that the purple squares must contain mines but that the green squares do not. The status of the yellow square is uncertain but can be determined after one more move. In contrast, the position shown above is not logically consistent: no allocation of mines in this grid could match the numbers shown in the squares.

ALL ILLUSTRATIONS BY BRYAN CHRISTIE

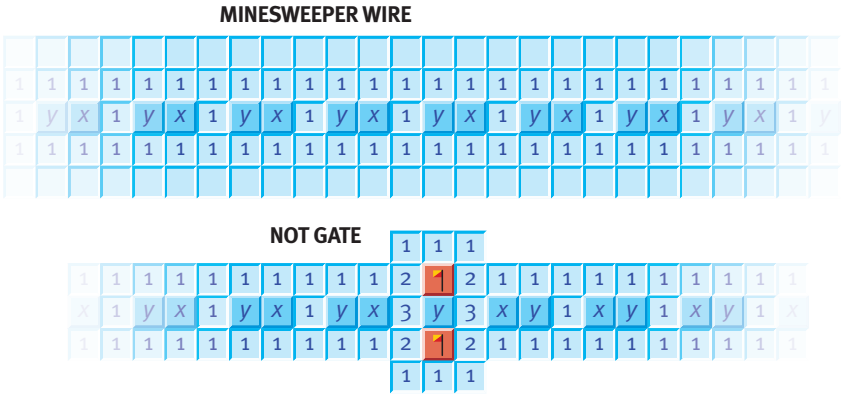
a given integer is also thought to be non-P, but this hasn't been proved either.

Why is it so hard to prove that a problem is non-P? Because you can't do it by analyzing any particular algorithm. You have to contemplate all possible algorithms and show that none of them can solve the problem in polynomial time. This is a mind-boggling task. The best that has been done is to prove that a broad class of candidate non-P problems are all on the same footing: if any one of them can be solved in polynomial time, then they all can. These problems are said to have a nondeterministic polynomial running time and are therefore called type NP.

NP is not the same as non-P. A problem is NP if you can check whether a proposed solution is correct in polynomial time. This is—or at least seems to be—a much less stringent condition than being able to actually find the solution in polynomial time. My favorite example is a jigsaw puzzle. Solving the puzzle can be very hard, but if someone claims to have solved it, it usually takes no more than a few seconds to check whether the solution is right. Just look at each puzzle piece and make sure that it fits with its neighbors. The time required to do this is roughly proportional to the number of pieces, so the check runs in polynomial time. But you can't construct a jigsaw puzzle in polynomial time. If you attempt to solve the problem by trying all potential solutions and checking each one in turn, the running time will be enormous, because the number of potential solutions grows much faster than any fixed power of the number of pieces.

It turns out that a lot of NP problems have equivalent running times. Specifically, an NP problem is said to be NP-complete if the existence of a polynomial-time solution for that problem implies that all NP problems have a polynomial-time solution. Thus, if you can solve one NP-complete problem in polynomial time, you've solved all NP problems in polynomial time. The P vs. NP question asks whether types P and NP are the same, despite all appearances to the contrary. The expected answer is no. But if any NP-complete problem turns out to have a polynomial-time solution, then NP must equal P.

Many problems are known to be NP-complete. One of the simplest is called the SAT Problem, which involves Boolean circuits. Such circuits are built from logic gates with names like AND, OR and NOT. The inputs to these circuits are either T (true) or F (false). Each gate combines the inputs in a specified way and yields the result of that combination as its output. For instance, a NOT gate turns a T input into an F output, and vice versa. The SAT Problem asks whether there exist choices of inputs that produce the output T for a given Boolean circuit. The problem is child's play for a simple circuit but becomes quite difficult when the circuit contains large numbers of gates and inputs.



MINESWEEPER ELECTRONICS converts Boolean circuits into game positions. If the squares marked *x* contain mines, the squares marked *y* do not, and vice versa. A Minesweeper wire (*top*) propagates a signal (*y* followed by *x*) along its length. A NOT gate (*bottom*) reverses the order of *x* and *y*.

The link to the computer game comes when we introduce the Minesweeper Consistency Problem. The challenge here is not to find the mines but to determine whether a given Minesweeper position is logically consistent. If, for example, during the course of play you encountered the position shown in the illustration at the bottom right of the opposite page, you would immediately know that the programmer had made a mistake. No allocation of mines could be consistent with the numbers in the squares.

In his article, Kaye proves that the SAT Problem for any given Boolean circuit can be transformed into a Minesweeper Consistency Problem for some game position. What is more, he shows that the procedure for converting Boolean circuits into Minesweeper positions runs in polynomial time. In this conversion, the inputs and outputs of the circuit become the pattern of mines in the grid: a square with a mine stands for T, and a square without a mine stands for F. The wire that connects the circuit's gates is represented by the Minesweeper position shown in the top illustration above. In this Minesweeper wire, the squares marked *y* have the opposite values of the squares marked *x*—if the *x* squares contains mines (T), the *y* squares don't (F), and vice versa. Notice that the numbers shown in the other squares are correct whether *x* is T or F. The effect of the wire is to "propagate" either signal (T or F) along its length.

The same procedure can also depict the logic gates themselves. The Minesweeper position in the bottom illustration shows a NOT gate. The block of numbers in the middle of the figure forces an interchange of the positions of *x* and *y*: *x* follows *y* in the input wire but precedes *y* in the output wire. The values of the squares are reversed whether *x* is T or F, which is exactly what a NOT gate does in a Boolean circuit.

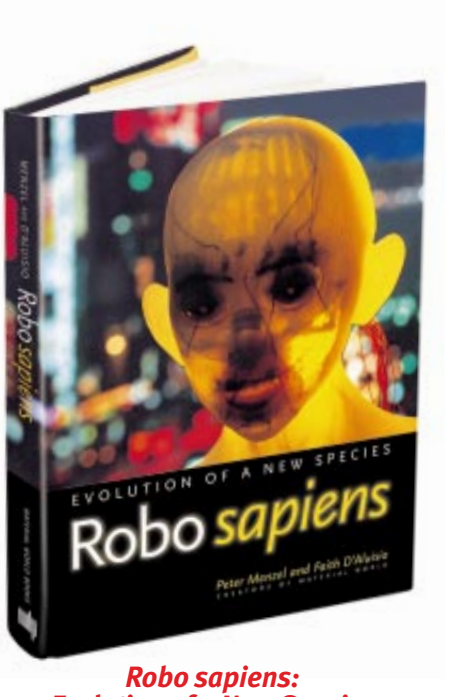
Minesweeper electronics gets much more complicated, of course. You need to be able to bend and split the wires as well. Kaye manages to solve all these problems in his article. The upshot is that if you could solve the Minesweeper Consistency Problem for a given game position in polynomial time, you would have solved the SAT Problem for the equivalent circuit in polynomial time. In other words, Minesweeper is NP-complete. So if someone finds a polynomial-time solution to the Minesweeper Consistency Problem, then all NP problems have polynomial-time solutions, and hence P equals NP. Alternatively, if someone can prove that no such solution exists for the Minesweeper problem, then P does not equal NP. Either way, the question would finally be answered.

Before you get your hopes up, though, remember that the Minesweeper Consistency Problem is a tough nut to crack. Determining whether a game position is logically consistent becomes very difficult when you're considering a gigantic grid, and most mathematicians and computer scientists believe that there's no general solution that runs in polynomial time. Moreover, the Clay Institute has imposed strict rules for the contest: before it will accept a solution as valid, it must be published by a major journal and be "generally accepted" by the mathematical community within two years of publication.

REVIEWER_VERNOR VINGE

Apocalyptic Optimism

Menzel and D'Aluisio foresee a superintelligent hybrid species



Robo sapiens: Evolution of a New Species
by Peter Menzel and Faith D'Aluisio
The MIT Press, Cambridge, Mass., 2000
($29.95)

Computers are the most important thing to come along since ...." It would be interesting to ask people from over the past 30 years to complete the preceding sentence. In the 1970s most people might finish the sentence with "television" or "the automobile" or even "atomic energy." In the 1990s the ante was raised, and I imagine that the average person would complete the sentence with "the industrial revolution" or even "the invention of writing." Now, as the 21st century dawns, some would say: "Computers are the most important thing to come along since the rise of humankind on Earth."

Even this last comparison may be conservative. Computer hardware power is doubling every two years or so. If that trend continues, then in another 20 years, our machines will be more powerful than many estimates of human brainpower.

These supermachines might come about in several ways. In their book *Robo sapiens,* Peter Menzel and Faith D'Aluisio examine one of the most significant and plausible possibilities. The prospect of their title is explained on the very first page of the introduction: *"Robo sapiens:* ... A hybrid species of human and robot with intelligence vastly superior to that of purely biological mankind; began to emerge in the twenty-first century."

After their excellent introductory comments, the authors indulge in little speculation. Instead they show us the current state of the art. Photographer Menzel and journalist D'Aluisio traveled to dozens of research sites around the world and conducted what must be hundreds of interviews. They have organized their work into 53 short chapters. Each consists of one or more interviews, photos of the robots, and sidebars giving project specifications and background information.

Project by project, they build the overview for us. We see the opinions that researchers have of the field and of one another's work. We see the conflicts that exist among different research approaches. No consensus has been reached as to the path to success or what may be possible. We are in the midst of an intellectual turmoil, driven by human genius and the steadily increasing power of our computer hardware.

The breadth of research projects is enormous. We are introduced to autonomous devices and teloperated ones, uncrewed rovers and robotic aircraft and pipeline crawlers. Pets and companions, prosthetics and hands-at-a-distance, housekeepers and welders, weapons and guardians—someone somewhere has found money (military, commercial, medical) to work on these possibilities. Many researchers are trying to build machines by imitating life—and some bioscientists are using robot models to improve their understanding of life.

It is very difficult to write clearly and simply about such things and still remain faithful to the underlying technology. Menzel and D'Aluisio accomplish this feat. The interviews and sidebars are consistently intriguing but also show great effort in avoiding the misinterpretations that could come from comparing so many different research enterprises. (Considering the range of research studied, an appendix categorizing the different projects would have been helpful.)

Many of the robots are mechanically complex. Often they are unlike anything we have seen before. Imagine that you had never seen a horse: you would need clear and objective pictures to get a good idea of the creature's appearance. Similarly, Menzel's color photos are essential to understanding the projects described. The photos are beautiful and, for the most part, extremely effective. In a few cases, however, the higher-level goals of the photography make the details of the machines unclear—for example, the motion-blurring effect used in picturing the M.I.T. Spring Flamingo robot on page nine or the Honda P3 robot on page 34 [*see opposite page*].

The chapters related to robotic pets and artificial affect are especially fascinating. Setting aside issues of intelligence and


1


2

1 Researcher Cynthia Breazeal of the M.I.T. Artificial Intelligence Laboratory holds a mirror to her creation, Kismet; **2** Kismet's disassembled head reveals the mechanisms that move its features; **3** Honda P3 strides along at the company's secret research facility near Tokyo; **4** M.I.T. Leg Lab researcher Hugh Herr, himself a double amputee, is developing a robotic knee; **5** My Real Baby, a collaboration between toy company Hasbro and iRobot, can mimic the facial expressions of a human infant.

THE EDITORS_RECOMMEND

FROM *THE GILDED DINOSAUR*

MARK JAFFE'S *The Gilded Dinosaur: The Fossil War between E. D. Cope and O. C. Marsh.* Crown Publishers, New York, 2000 ($25).

The early years and troubled maturation of American paleontology are Jaffe's subject, which he treats mainly by recounting the 19th-century rivalry between Edward Drinker Cope and Othniel Charles Marsh. Cope was the scion of a wealthy Philadelphia family and a self-made scientist. Marsh, born on a hardscrabble farm near Buffalo, probably would have become a carpenter or country schoolteacher if he had not had a rich uncle, George Peabody, who financed the education that enabled him to become a professor of paleontology at Yale in 1866 (an appointment doubtless facilitated by a substantial donation to the university from Peabody).

For many years, Cope and Marsh competed in finding or acquiring fossils, particularly dinosaur bones, and trying to classify them. Marsh got the better of it. "When it came to dinosaurs," says paleontologist Peter Dodson of the University of Pennsylvania, "Cope was always a day late and a dollar short." Science writer Jaffe makes a rich tale of all this, describing not only the tense relations between Cope and Marsh but also what it was like to hunt for fossils in the frontier atmosphere of the U.S. West and to fit them into a coherent classification that helped to pave the road to modern paleontology.

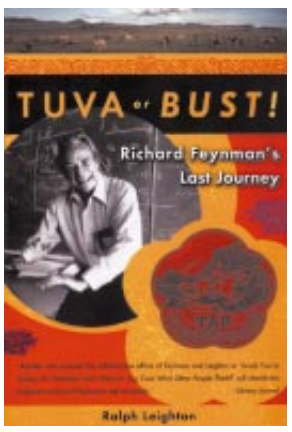
LEWIS WOLPERT'S *Malignant Sadness: The Anatomy of Depression.* Free Press, New York, 2000 ($24).

Wolpert, a professor of biology as applied to medicine at University College London, is a recovered victim of clinical depression. "It was the worst experience of my life," he writes. "Severe depression borders on being beyond description: it is not just feeling much lower than usual." Wolpert tells of his experience and treatment in this book, which he wrote with a fourfold purpose: "to help those who are living or working with a sufferer to understand the nature of depression, since depressives, whether parents, children or companions, are not easy to be with; to help depressives to understand themselves; to remove the stigma associated with depression; and, foremost, to try and understand the nature of this dreadful affliction in scientific terms." He examines, among other things, the factors that make people vulnerable to depression, why about twice as many women as men suffer from severe depression, and the psychological and biological theories that have been advanced to explain the disease.

Acknowledging that psychotherapy succeeds about as well as medication in treating depression, he foresees that it "is nevertheless in relation to chemical processes occurring in the brain that future research is very promising, together with the development of new drugs to restore normal function of the neurons involved."

RALPH LEIGHTON'S *Tuva or Bust!: Richard Feynman's Last Journey.* W. W. Norton, New York, 2000 ($13.95).

It was characteristic of Richard Feynman, the eminent Caltech physicist, to take an intense interest in something for a quirky reason. In the case of Tuva, once an independent nation between Siberia and Mongolia and later an autonomous republic of the Soviet Union, it was his discovery that the capital is Kyzyl. "A place that's spelled K-Y-Z-Y-L has just *got* to be interesting," he said. And so he and his friend, literary collaborator and fellow recreational drummer Leighton, decided to go there. The effort to do so plunged them into a harrowing but often hilarious 10-year struggle with bureaucrats and language difficulties. Feynman never made it; he died in 1988, a few months before Leighton at last got to visit Kyzyl. It is a delightful tale that Leighton tells in this updated paperback version of the hardcover published in 1991. "Like Magellan," Leighton says, "Richard Feynman completed his last journey in our minds and hearts. Through his inspiration to others his dream took on a life of its own."



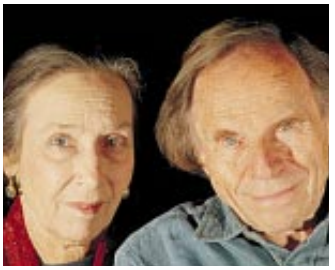
mobility, when you pat your dog on the head and look into its eyes, you can see that *someone is home.* Very few of us would say any such thing about a machine. What's missing from the machine? Perhaps it's simply that our instincts say "warm, soft, responsive creature with large dark eyes" = "living being with a soul." Commercial products such as Sony's AIBO and (soon) Hasbro/iRobot's My Real Baby are early attempts to address this issue. Will people be attracted, repelled or simply bored by such creations?

The arrival of *Robo sapiens* is not assured. Maybe we have not yet stumbled on some basic impossibility. Or maybe our researchers are ignoring some critical technique: to one degree or another, almost everyone interviewed is working from the bottom up, imitating the lower levels of animal capability with a view toward later building the higher functions. The alternative, top-down approach would seek to model the highest functions of thought first, perhaps in a supercomputer or computer network—and then, with luck, the hands and legs and wheels would be easy add-ons. For the past 10 or 15 years, the top-down approach has been in eclipse.

A theme builds through the interviews: living creatures are superior in almost every way to existing machines. Even some of the superiorities that we typically grant machines over animals—durability, speed, strength—do not hold in a natural environment. Similarly, real-time computations that involve reasoning with the chaos of the material world are also beyond the computers we have now.

But remember, computer power is doubling about every two years. What we see now in robotics and AI research is still the historic beginning. Menzel and D'Aluisio have given us a richly detailed picture of this tumult as it is at the end of the 20th century. Their book provides a baseline for comparison with projects in the years ahead. What develops in the next 10 years should yield strong evidence about where things are going. If we get effective household robots, if we get vehicles that can navigate real terrain autonomously, if we get lovable robotic pets … then we are on track to create something very like *Robo sapiens.*

VERNOR VINGE taught mathematics and computer science at San Diego State University from 1972 to 2000. He is also a science fiction writer; his most recent novel is A Deepness in the Sky *(Tor Books, 1999).*



Hybrid Vigor!

Philip and Phylis Morrison go for a ride in a practical car that guzzles less gasoline

The miller is at home near his millstones and beside his dam. His place is fixed. Our big modern power plants evidently share this medieval trait. But in our restless times, power for travel is needed as urgently as that from stationary sources. In travel, though, the circumstances vary. Even long ago mobile power sources were often hybrids, two or more distinct power supplies right at hand to fit the various demands of a journey. The Viking long ship is one: a square sail well set for long voyages and a party of tough oarsmen ready for sudden action. A present-day example is the diesel-electric locomotive. Economical diesel engines dislike slowing, stopping and starting, but rail travel demands all three. In the throbbing hybrid locomotive, queen of the rails since World War II, the diesel engine on board runs at its most efficient rpm but turns not a train wheel. Rather it spins an electrical generator for fully controlled onboard power, feeding electric motors that turn the wheels with flexibility and efficiency costly gears could not supply.

Internal-combustion engines of all sizes, mainly using gasoline, are remarkably well developed in cost and in reliability. They convert fuel to mechanical power with an efficiency high for heat engines, 25 to 30 percent. Alternatives are not lacking, though they improve slowly. Purely all-electric autos, powered by photovoltaic cells under the sun, are here, but fit so poorly into the present context that they remain a sport. A typical car engine delivers about 75 kilowatts of mechanical power, whereas sunlight provides no more than half a kilowatt per square meter of surface area. If the capturing solar cell were 30 percent efficient in energy capture (a high value), your car would spread out over the area of a spacious house. Such paragons could move by day, but their night use demands stored energy, a hybrid in itself, whose cost, size and weight are high. Practical solar cars cannot yet compete on present market terms; the designs we see lack the range and speed motorists now expect.

Hybrid vigor, a favorite phrase among flower breeders, is pretty apt for power engineering as well. The automobile species seems to be at the cusp of hybridization on a really large scale. The new gasoline-electric hybrid car has been brilliantly engineered over the past decade or so. It is here now, in small production numbers at low prices. In Japan tens of thousands of them have been on the road in private hands for several years, and thousands are entering the U.S. this year. Their practical hybrid design is based on engineering insight into the true nature of the present passenger automobile, for almost every car on the road is already a hybrid. Gasoline engines deliver the overall energy used in most cars, nearly all of it directly to the wheels. But they are assisted by an indispensable system that stores and supplies the electrical fraction, a few percent at most, from a storage battery kept charged by the main gas engine. That has been true since the end of World War I, when electric starters entered in number. (Before that, cars were started by the hardworking driver turning the crank.)

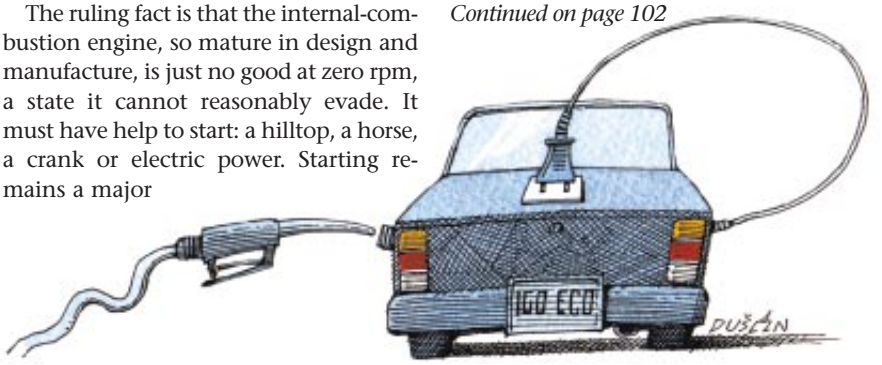
*The automobile species seems to be at **the cusp of hybridization** on a really large scale.*

The ruling fact is that the internal-combustion engine, so mature in design and manufacture, is just no good at zero rpm, a state it cannot reasonably evade. It must have help to start: a hilltop, a horse, a crank or electric power. Starting remains a major task for the new hybrid's batteries, just as it is in the cars that fill the roads, but in the hybrids electric power is applied to motion as well. Toyota and Honda, the two most widely imported makes in the U.S., offer such new hybrids in the U.S. this year, nor will Detroit and Stuttgart be far behind.

We have ridden watchfully and excitedly in the Honda Insight, the first hybrid in regular U.S. sale. The car is a handsome two-seater coupe, low and sleek, though not a model meant for the market mainstream. (Toyota's hybrid, the Prius, is a neat sedan, half a foot shorter than its best-sellers.) The Honda designers did not ignore the many ways to improve on today's cars. It is light (under a ton), with an aluminum monocoque body (no steel frame, although it has steel door beams for safety), tires with low rolling losses and the best aerodynamic performance on the mass market. Raindrops do not strike the windshield in the rain! The gas engine is an in-line, three-cylinder aluminum casting offering about 50 kilowatts of power, with lean-burning stirred inflow, well-controlled valving and timing, and ultralow emissions at near-steady optimum rpm. No stripped version, Insight has full power assists, antilock braking, air-conditioning, a full panel of digital instruments, even lots of cup holders, at about a $20,000 sticker price.

Lift the hood to see the front end, two slender radiators to cool engine and people, manual transmission below, amid accessories that crowd all engine rooms.

Continued on page 102

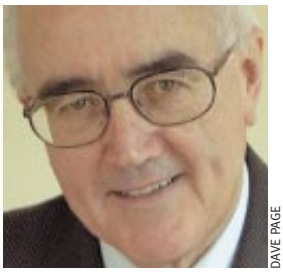


DUSAN PETRICIC

Getting High

Booze, dope, metaphysics and other rhapsodic matters—

James Burke airs a few Renaissance thoughts



DAVE PAGE

As I came out of the library the other day, an ambulance raced past. The AMBULANCE on the front was quite a coincidence, as I'd just been looking at some of Leonardo da Vinci's mirror writing. Talk about a role model for the Renaissance man. Leonardo's talents and interests were so enormous I'll leave it at that, except to say I particularly liked his high-tech anatomical drawings done sometime around 1510. Especially the one of the kidney, a major source of problems in an age when the rich and famous went for 10-course meals that were washed down with copious amounts of booze. Result: gout. Effect: kidney stones. By the 18th century this condition was the biggest earner for physicians to the well-heeled.

In 1753 one treatment attempted in Edinburgh was to dissolve the stones with caustic limewater. Another possibility, offered by medical student Joseph Black, was to use a preparation of magnesium carbonate. Wrong, but during the experiments Black came across carbon dioxide, extinguishing candles and small birds with it, and thus inspired subsequent, more famous work by such illustrious suffocaters as Joseph Priestley and Antoine Lavoisier. (Black is probably best known as the guy who told James Watt about latent heat, thereby triggering Watt's idea of a separate condenser.)

In 1766 Black took the Edinburgh chair of chemistry when his old friend and former teacher William Cullen moved over to medicine. Cullen's boffo success in the lecture hall had come from his introduction of the hip new trend of lecturing in English rather than Latin. He also got caught up in the contemporary craze for categorizing diseases by their symptoms, a.k.a. "nosology." (A sample from his work on the subject: "Melancholy. Symptoms: aversion to action and the duties of life.")

At one point, Cullen employed a penniless young teacher as tutor to his children. This wannabe doctor, John Brown, repaid the kindness years later, when he, too, had become a medic, with a scathing attack on Cullen's practices. By this time Brown had made so many enemies with his big mouth that he was never going to get a professorship. So instead he wrote *Elements of Medicine* and started a whole school of therapy based on the idea that all diseases were either a shortage or an excess of excitability, to be treated with either a stimulant (alcohol) or a sedative (opium).

Brunonian theories created such a furor that the man had his defenders and detractors fighting it out in the streets of Göttingen. Alas, in the end, Brown was to expire from taking (too much of) his own medicine. This in 1788, seven years before a late edition of *Elements* was put together by one Thomas Beddoes, who had his own medical ax to grind: treatment of all diseases by inhalation at his Pneumatic Institute (equipment by Watt, funding by Josiah Wedgwood). Beddoes's only real claim to fame (the institute failed) was that he hired a 19-year-old unknown named Humphry Davy.

Who is now so famous I'll just mention a couple of reasons why: the miner's lamp, nitrous oxide anesthetic and (from 1801) Royal Institution lectures with so much pizzazz that they became an intellectual fashion statement among the elite. One of whom was an old pal of Davy's (Davy had pretensions, fortunately discouraged, to writing poetry): Samuel Taylor Coleridge, by this time a well-known scribbler and another opium addict. Who introduced Davy to a young hanger-on, an as yet unknown scribbler and *another* opium addict named Thomas De Quincey. He would become famous with his *Confessions of an English Opium-Eater* in 1821.

De Quincey lived in such chaos that on many occasions he was unable to write a piece because he couldn't locate his notes. After years of writing biogs for the *Britannica* and lightweight magazine essays, and assiduously cultivating Wordsworth, De Quincey moved to Scotland, where he assiduously cultivated Sir William Hamilton. Now I'll give this to you straight, because reading about it turned my brain to porridge, so it's all I can do: Hamilton, Edinburgh prof of logic and metaphysics, quantified the predicate. Okay? (If you need to know more, try J. S. Mill's *An Examination of Sir William Hamilton's Philosophy,* London, 1889, and serve you right.)

One of Hamilton's pupils was a real unfortunate, name of Archibald S. Couper. A two-handkerchief tale. In 1858 Archy is working in a Paris lab and asks the lab boss to get somebody to present to the French Academy a paper he's done. Archy says hurry up when boss delays. Archy is ejected from the lab and returns to obscurity in Scotland. Meanwhile somebody else comes up with the same ideas and gets all the glory. Archy is never heard from again and eventually goes gaga.

"Melancholy. Symptoms: aversion to action and the duties of life."

The somebody else? Only the mighty Friedrich August Kekule von Stradonitz, whose ideas (atoms can link up thanks to valence: hydrogen

PATRICIA J. WYNNE

with one other atom, oxygen with two, and so on; and the concept of a closed ring of linked carbon atoms, as in "benzene ring") laid the foundations for modern structural chemistry. Or rather, did not, because poor old unimportant Archy Couper really got there first.

Kekule did, however, organize the first International Congress of Chemists in 1860 at Karlsruhe, Germany, where they sorted out the confusion in chemical terminology. Among those who showed up was a Russian, Dmitri Ivanovich Mendeleyev, who would later prove to be too democratic for his Czarist masters and lose his university job. Ended up director of the Bureau of Weights and Measures.

Ironic, that, because in 1869 Mendeleyev had discovered the periodic table of elements, which arranged them in order of increasing atomic weight (wonder if anybody saw the joke). One of his European traveling companions (who also turned up at the Karlsruhe congress) was a fellow chemist and the illegitimate son of a prince, Aleksandr Porfiryevich Borodin. Who in 1862 went back to St. Petersburg to a professorship, a life of chemistry and efforts to get women into the medical profession.

For relaxation, Borodin hung out with a group of twentysomethings he'd met through an old childhood pal, who all encouraged him to turn his hand to the activity that has since made him more famous than his chemistry: music, especially his "Polovtsian Dances," now a staple at concerts worldwide. At the time, Russian music was split between the nationalists (Borodin and friends) and the Westernizers, headed by Tchaikovsky. One of whose protégés was Sergei Rachmaninoff (whose "Rhapsody on a Theme of Paganini" is another staple). When the October Revolution hit, Rachmaninoff hightailed it for the U.S., after which his music was banned in Russia, as "representing the decadent attitude of the lower middle classes."

In 1923 Rachmaninoff got a letter from a fellow immigrant that impressed him so much he sent the guy $5,000, as a result of which Igor Sikorsky was able to start up his Aero Engineering Corporation in a barn on Long Island. In 1937 Pan Am flew the Atlantic in Sikorsky's four-engined Clipper. And in 1939 Sikorsky realized a lifelong dream: to build and fly the device that had inspired him when he first saw it as a child, in one of Leonardo's other drawings—a helicopter.

Must fly.

Wonders, *continued from page 99*

Our focus is the drivetrain. Begin with the engine itself, a box about the size of an elderly laptop computer; this 1.0-liter three-cylinder does the work of a usual 1.5-liter four. It drives directly the slender electric motor next in line, where wires dive below to the inside "rear seat," packed with circuitry under microchip rule. Forty pounds of rechargeable nickel metal-hydride 1.2-volt cells, 120 cells in series, are there, of the size we know from familiar cylindrical flashlights. The 10-gallon gas tank supports long range, because Insight drivers expect roughly 65 to 70 miles per gallon on the road. The instruments report instantaneous and cumulative mileage, and much more, in glowing digits. The warranty calls for a tune-up at 105,000 miles and maybe a battery replacement? The enthusiastic owner who took us out one at a time—our son—had mounted an apt new license plate: IGO ECO.

This car *works*. Drive enjoyably along, and stop for a traffic signal. The engine stops, too, wholly silent until you engage and accelerate, whereupon it very quietly restarts, to carry you to highway speeds in sports-car time. The 10-kilowatt permanent magnet motor-generator smoothly assists engine startup until gasoline power takes over to feed energy back into the batteries; frugal regenerative braking at any slowdown is equally automatic.

The grand average passenger car mileage in the U.S. is 21 miles per gallon. But the average life of a vehicle on U.S. roads is about eight years, a quarter of them beyond 12 years old—deep change cannot be fast. Not prophets, we expect hybrids to grow in number until they populate most roads, reducing fuel use to a third and emissions to an eighth. Fuel cells promise two or three times the fuel efficiency of any heat engine and may well become the next standard; we do not know what fuels will be used, perhaps solar-generated hydrogen itself? Benign water vapor could replace the carbon dioxide wastes of half a billion light, economical cars in the world of six or eight decades ahead. The fuel industry will be lean, personal autos less dominant.


DAN WAGNER

Founding Father of Invention

Now is the time for all good men and women to reflect for a moment about one of the key figures in American inventiveness, notes **Steve Mirsky**

Three out of every four years autumn brings crisper air, breathtaking foliage and, barring a strike, a World Series. Well, technically, it's four out of four years, but every fourth year those pleasant things are drowned out by the trumpeting elephants and braying asses of the presidential election circus. (Although I write these words in early August, I have no doubt that by the time you read them you're feeling completely bushed and totally gored.)

In the heat—well, warmth—of this historic moment, perhaps it would be refreshing to contemplate a previous occupant of the White House and reflect upon his immortal words: "It occurred then that this globe might be made to perform the functions of a dial. I ascertained on it two poles, delineated its equator and tropics, described meridians at every 15 degrees from tropic to tropic, and shorter portions of meridian intermediately for the half hours, quarter hours, and every five minutes." Thus did Thomas Jefferson describe part of his design for a spherical sundial, which, for the benefit of the current crop of presidential aspirants, is like a big watch that works only during the daytime.

When he wasn't busy thinking up the United States, Jefferson was coming up with all kinds of useful devices. He mathematically deduced an improved plow that cut through the soil with increased efficiency. He thought up a contraption, called a wheel cipher, that could encode secret written messages for safe passage to American agents overseas. And he originated channel surfing, or at least the 18th-century version, when he designed a revolving stand that enabled the busy reader to keep five books in play simultaneously. (Note to George W.: Five! Books! Hey, it's still fun to spin.)

Jefferson never bothered to patent his various inventions, as he had mixed feelings about ownership of intellectual property that might limit its dissemination. And he could afford to have such mixed feelings, being more upper crusty than the homemade apple pie he was as American as. But he also recognized that the economic motivation of at least temporary ownership of a concept for a device would spur invention. (Note to Al Gore: Nice try with the I-invented-the-Internet thing.)

So it came to pass that one of the most important things Jefferson helped to invent was the U.S. Patent Office, in 1790. As secretary of state under the original George W., Jefferson led the three-man board that reviewed patent applications. Although a big fan of inventions who went so far as to personally test some of the new devices for which patents were sought, Jefferson kept grumbling something to the effect of, "Hey, I'm the secretary of state here." In 1793 Congress finally heard his grouses, which sometimes roamed the woods around Monticello, and absolved him of his patent responsibilities.

Of course, not all patents have gone to inventions as useful as those Jefferson himself dreamed up. "The abuse of frivolous patents," he once noted, "is likely to cause more inconvenience than is countervailed by those really useful." He most likely then would have disapproved of patent number 5971829, which was given to the inventor of the motorized ice cream cone. This gem of an idea rotates the scoop of vanilla so that the eater can simply stick his or her tongue out, as if entering the voting booth on November 7. Jefferson probably would also have been unhappy about patent number 5457821, granted to the guy who put a white, irregular brim together with a yellow dome to create a baseball cap that looks like a fried egg. Putting the egg on your face approximates the effect of walking out of that same voting booth. SA

"It's that plan to reintroduce wildlife into their original territory."

SIDNEY HARRIS