SCIENTIFIC AMERICAN

APRIL 1993
$3.95

Controlling the quantum jitters of atoms.
The implications of an aging human species.
High-tech materials for roads and bridges.

Night-hunting owl *can locate prey by sound alone. Studies reveal how the brain calculates direction from acoustic cues.*

SCIENTIFIC AMERICAN
April 1993 Volume 268 Number 4

Scientific American (ISSN 0036-8733), published monthly by Scientific American, Inc., 415 Madison Avenue, New York, N.Y. 10017-1111. Copyright © 1993 by Scientific American, Inc. All rights reserved. Printed in the U.S.A. No part of this issue may be reproduced by any mechanical, photographic or electronic process, or in the form of a phonographic recording, nor may it be stored in a retrieval system, transmitted or otherwise copied for public or private use without written permission of the publisher. Second-class postage paid at New York, N.Y., and at additional mailing offices. Authorized as second-class mail by the Post Office Department, Ottawa, Canada, and for payment of postage in cash. Canadian GST No. R 127387652. Subscription rates: one year $36 (outside U.S. and possessions add $11 per year for postage). Subscription inquiries: U.S. and Canada 800-333-1199; other 515-247-7631. Postmaster: Send address changes to Scientific American, Box 3187, Harlan, Iowa 51537. Reprints available: write Reprint Department, Scientific American, Inc., 415 Madison Avenue, New York, N.Y. 10017-1111, or fax: (212) 355-0408.

DEPARTMENTS

THE COVER photograph captures a Ural owl (*Strix uralensis*) flying back to its nest with dinner. Nocturnal owls such as the Ural rely on acoustic cues to help them catch their prey in the dark. Studies on another night hunter, the barn owl (*Tyto alba*), have revealed most of the steps by which the brain processes these cues (see "Listening with Two Ears," by Masakazu Konishi, page 66). The brains of mammals, including humans, probably use a similar system consisting of hierarchical steps and parallel pathways to process sound.

THE ILLUSTRATIONS

Cover photograph by Satoshi Kuribayashi/Nature Productions

Page	Source
47	Dan Wagner (photograph aging by Rick's Retouching)
48–52	Ian Worpole
55	Steve Murez/Black Star
56–60	Jared Schneidman/JSD
67	Masakazu Konishi
68	Jana Brenning; Michael S. Brainard (data for *b* and *d*) and Eric I. Knudsen, Stanford University (*photograph inset*)
69	Jana Brenning
70	Jason Küffer
71–72	Jana Brenning
73	Johnny Johnson
74–75	Visual Logic
76	Ullstein Bilderdienst, Berlin
77	Michael Goodman
78–79	Michael Goodman (*top*), Visual Logic (*bottom*)
81	Oxford Scientific Films Ltd./Carolina Biological Supply Company
82–83	Patricia J. Wynne
84	Johnny Johnson (*left*), Dwight R. Kuhn/Bruce Coleman, Inc. (*top right*), Kim Taylor/Bruce Coleman, Inc. (*center right*), Oxford Scientific Films Ltd./Carolina Biological Supply Company (*bottom right*)
86	George J. Pinwell, *Death's Dispensary,* circa 1866, Philadelphia Museum of Art, Art Media Collection
87	Joseph Polleross/JB Pictures
88	Ian Worpole
90	Ian Worpole (*top* and *middle*), William C. Brown, Science Source/Photo Researchers, Inc. (*bottom*)
91	Jim Pickerell, Tony Stone Worldwide (*left*), Ian Worpole (*right*)
92	Charles Dauget, Pasteur Institute; courtesy of Photo Researchers, Inc.
93	Ohmori Satoru/Gamma-Liaison
94–95	Courtesy of Ofer Bar-Yosef and Bernard Vandermeersch; Ian Worpole (*map inset*)
96	Courtesy of Ofer Bar-Yosef and Bernard Vandermeersch
97	Ian Worpole
98–100	Courtesy of Ofer Bar-Yosef and Bernard Vandermeersch
102–103	National Institute of Standards and Technology
104–105	Johnny Johnson (*left*), National Institute of Standards and Technology (*right*)
106–107	Neil McDonald
110	Maunsell Structural Plastics (*top* and *bottom*)
111	John James Danehy
120–121	Andrew Christie

SCIENTIFIC AMERICAN®

Established 1845

LETTERS TO THE EDITORS

Everyone's a Critic

Well, by and large, "Reproductive Strategies of Frogs," by William E. Duellman [SCIENTIFIC AMERICAN, July 1992], is the most disgusting damned thing I have ever seen.

J. A. NUNLEY
Milpitas, Calif.

For 30 years, I enjoyed, devoured and carefully stored *Scientific American* for reference. Now all is bleak 'round the battlements. Alas, "my" *Scientific American* is dead, replaced by a pale surrogate, an organ of leftist apocalyptic causes. This editorial swing leftward was expectable, considering the inexorable dilution of your once excellent staff by women.

LORING EMERY
Hamburg, Pa.

When will you publishers stop propagandizing for speculative ideas such as the big bang and black holes? When they are discovered not to exist, what rationale will you use, since you plastered your magazine full of this nonsense? You are the publicity agents for birdbrain professors of physics.

I give you till the end of the year to publish the fact that the observable universe is the last electron of plutonium.

LUDWIG PLUTONIUM
White River Junction, Vt.

Attention, West Virginia....

I hope you see some merit in my process for mass-manufacturing diamonds with subterranean nuclear explosions. One day in the not too distant future I may get to push a button and blow a coal mine in West Virginia all over creation. In the rubble will be diamonds you can pick up with a scoop loader.

Unless you are sure for some reason that the process cannot work, I do not understand why *Scientific American* will not report on the possibilities of this process. I have already met the expected red tape in Washington, but that is something persistence and being right have always overcome. I will continue to keep you informed of the progress of this project. I am very sure somebody would like to be there when the button is pushed.

JAMES W. LINCK
Kenner, La.

In regard to the failure of the *Hubble Space Telescope*: Yet again a very large amount of money has been lost probing the universe. The mirror makers, who are supposedly the best on earth, have been blamed for the poor pictures taken by the telescope. There may, however, be another explanation.

Beyond the solar system there is nothing real! There is only a set of illusory images created by the boundless void in which the solar system is encased and reflected, as in a virtual spherical mirror. We are totally alone!

There seems to be a need to refute this theory before squandering further terrestrial resources.

SHAFI AHMED
London

First Contact?

I was greatly intrigued by the picture appearing on pages 128–129 of "The Mind and Donald O. Hebb," by Peter M. Milner [SCIENTIFIC AMERICAN, January]. If I am not mistaken, the third man from the right, labeled as "Unidentified," seems to have two antennae protruding from his cranium. Was he the product of an unusual operation or an extraterrestrial attending Hebb's seminar? Any clarifications concerning this perplexing mystery would be appreciated.

JARED WHITE
Wayland, Mass.

Weight and See

In 1876 the entire membership of the American Society of Civil Engineers voted to use metric units only. It was internationally agreed in Paris in 1901 that mass is quantity of matter and that weight is force acting on mass. Yet there are universities, colleges, magazines and other entities that continue to use as units of measure the unsafe pound or the unsafe kilogram.

Net mass is required for fair trade; "net weight" is a government lie! The Olympic sport is masslifting, not weightlifting. A fat person is overmass, not overweight, and should lose mass if he wants to be thinner. How long can "educators" expect to fool the public with unsafe words and unsafe units? You are just as fat on the moon, but only one sixth the weight!

R. C. GERCKE
Los Angeles, Calif.

Toasting the Climate

Roses are red,
Violets are blue.
The fizz in beer and soda
Is CO_2.

Some "environ-mentalists"
Who have gone mad,
Tell people today that
CO_2 is bad.

Their stupid theory
I will strongly reject
Because I like the
"Greenhouse effect"!

We'll have warmer winters,
Which I like better,
And more ocean evaporation
Could made deserts wetter.

If glaciers melt
And the oceans rise,
A move to Alaska and Canada
Is easy and wise.

So bartender, serve me
More pop and beer.
The fizz will warm the winters
During each coming year.

JOSEPH GAYSOWSKI
Westchester, Ill.

50 AND 100 YEARS AGO

APRIL 1943

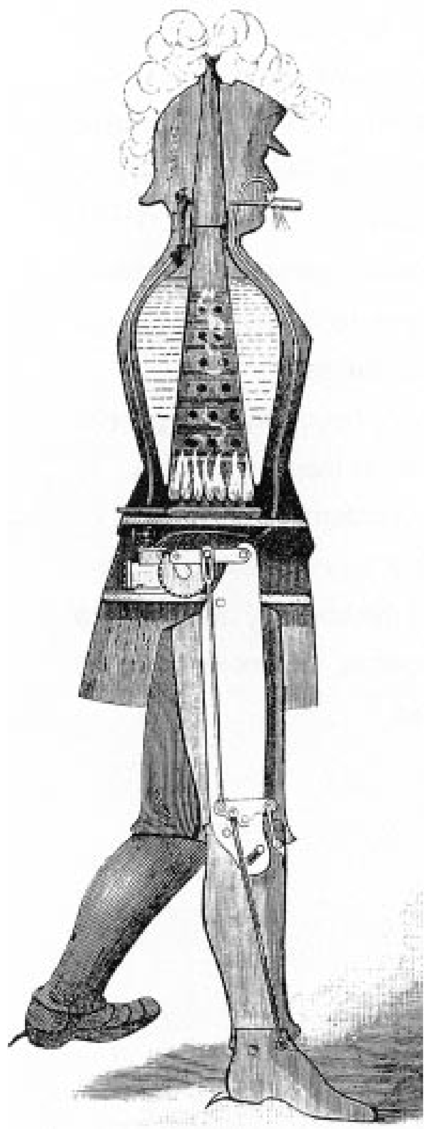
"A modern version of the discovery of the famous Damascus armorers of how to make sword steel that would bend and not break—without entailing the human suffering involved in the olden method—has been developed by 20th Century research. In the ancient method, human blood was the original 'quenching oil.' The technologists of the Gulf Research & Development Company concluded that the tissues of the body probably had more to do with the tempering than the blood itself. They were cognizant, however, of the fact that organic matter in the blood was made up generally of large molecules, and this knowledge was employed effectively in the experiments which led to the development of Super-Quench. It is said to have a cooling rate intermediate between other oils and water through the hardening temperature range and yet retaining the slow speed of other oils below the hardening range."

"The question of whether ill health can result from lead piping for household water supply has no categorical answer. The following is the reply given to a physician by *The Journal of the American Medical Association*. 'The amount of lead absorbed by most waters is negligible. Lead piping is effective in forming an insoluble coating of salts which inhibits its solution. It is *only when the water supply* is acid, particularly because of organic acids, that it is a potential danger. It may also dissolve when different metals are used in the plumbing, when galvanization may play a part. Water with highly solvent properties will dissolve some lead from a pipe on standing. The length of standing and the temperature of the water will influence the final concentration, but the actual quantities of lead will be small.' "

"Extra vitamin C is needed in the diet of soldiers under certain conditions and of workers exposed to industrial poisons, according to Prof. Harry N. Holmes of Oberlin College, president of the American Chemical Society. Vitamin C, which is destroyed by infection and by a number of industrial poisons of a military nature, is also lost in appreciable quantities in heavy perspiration, he points out. Prof. Holmes reports that one of the large rubber companies gave vitamin C daily to 100 workmen exposed to a so-called safe concentration of benzene and toluene vapors in the factory air. After a short time 37 of the workers felt 'less tired' at the end of the day, he says, 10 felt in better health generally, and only 31 reported no gain."

APRIL 1893

"Professor Hertz has shown that the rays proceeding from the cathode of a Geissler tube, which are capable of exciting phosphorescence, will pass through thin metal. If it were practicable to find a sheet of metal foil thick enough to be airtight and opaque, yet thin enough to be permeable by this discharge, it would be possible to allow these rays a passage into the open air by closing an opening in a discharge tube with such a piece of foil. This idea has been realized by Dr. Philip Lenard, assistant to Professor Hertz. A hammered aluminum plate 0.003 millimeter thick forms a shutter which Dr. Lenard calls the 'window,' because it allows the rays from a cathode at a distance of 12 centimeters to penetrate it freely. Substances capable of phosphorescence, if held near the window, shine with their peculiar light on the side nearest to it."



The steam man

"M. B. B. asks: If a ball be dropped into a hole that passes clear through the earth, would it stop when it reaches the center or pass by it? I hold that the ball would stop, and I wish to settle an argument. *A. The ball would have a hard rub in getting down to the center at all. Its circumferential velocity, derived from the earth's motion on its axis, would keep it against the east side of the hole, unless the hole was through the polar axis of the earth, when it might bob back and forth for a time until friction settled it at the center.*"

"A number of years ago what purported to be a steam man was widely advertised and exhibited in New York City. The remains of the individual in question were quite recently to be seen in one of the downtown junk stores. Within the last two years the project has been taken up by another inventor, and a practical steam man that actually walks and exerts considerable tractive power has been exhibited in actual operation in this city and elsewhere. It was invented and constructed by Prof. George Moore, a native of Canada. His steam man appears to be a native of America. In our illustration we show the section and general view of the steam man. In the body is the boiler, which is supplied with a gasoline fire. Below the boiler is situated the engine. While small in size, it is a high speed engine running up to 3,000 revolutions per minute or more, giving about $1/2$ horse power. The man, which is about 6 feet high, cannot, it is said, be held back by two men pulling against it."

Obstacle Course

Funding and policy stifle contraceptive research

The short list of birth control methods available in the U.S. is now longer by one, but the long list of obstacles facing contraceptive development is no shorter. For every advance, unsolved, unaddressed, sometimes unspoken, problems remain.

After 25 years of repeated review, an injectable synthetic hormone, Depo-Provera, was approved by the Food and Drug Administration last year. Approval of a female condom seems imminent, but not much else is waiting in the wings. The U.S. continues to have fewer birth control options than many other countries. And because use here reassures consumers at large that a product is safe, the country's contraceptive quandary can deter family planning elsewhere. "The U.S. is behind," states Rosemarie B. Thau, director of contraceptive research at the Population Council.

Nevertheless, Thau and many other researchers have found some hope in the early decisions of the new administration. President Bill Clinton issued an executive order stating that RU 486, the controversial French pill that induces menstruation, is no longer banned from personal use here. He also made explicit his intention to support family-planning programs by reversing what has been called the Mexico City Policy [*see box on page 22*]. "There is a new wind blowing, and it is attitudinal," comments Luigi Mastroianni, Jr., of the University of Pennsylvania, who directed a 1990 National Academy of Sciences study that detailed the reasons for the lag in U.S. contraceptive development.

The need for more options is vividly apparent. In the U.S. alone, there are about 3.5 million unintended pregnancies each year, 800,000 of them among teenagers, and 1.6 million abortions: these rates are among the highest for an industrialized country. Many forms of birth control have drawbacks—among them, an inability to protect against sexually transmitted diseases, of which there are 250 million new cases worldwide each year, according to the World Health Organization (WHO).

But if Clinton is going to counteract the policies of presidents Ronald Reagan and George Bush and provide the U.S. with a full range of contraceptive choice, he will have to back his intentions with funds. At present, most national support for birth control development comes from the National Institutes of Health's Contraceptive Development Branch. That program recently lost support for many of its grants and contracts when its budget plummeted from roughly $16 million in 1992 to about $9 million in the current fiscal year.

"What is in line for contraceptive development is less than it was a few months ago," says Nancy J. Alexander, chief of the branch. On hold, among other things, are studies on new condoms and diaphragms, transdermal patches that would deliver hormones and some aspects of birth-control vaccine development. "I just don't see any big influx of money into this research, much as it is needed, although I think there will be a shifting of priorities," she notes.

A significant share of the money NIH does have goes to three centers, established in 1991. Researchers at these sites as well as at other institutions are focusing on improving the methods already marketed here, winning approval for some that are available abroad and developing new approaches, such as contraceptive vaccines and a male pill. "Our main aim is to provide more methods so that various groups have access and so that men or women can switch methods," Thau notes.

Malcolm C. Pike and Darcey V. Spicer of the University of Southern California, for example, are improving on the pill concept. Using a compound that binds with receptors for gonadotropin-releasing hormone, the team has been able to prevent ovulation in a group of 14 women. The scientists simultaneously administer estrogen and progesterone to prevent postmenopausal symptoms, but they say the amounts of these hormones are significantly lower than those found in birth control pills.

CHRISTOPHER BURKE *Quesada/Burke*

DEVELOPING NEW CONTRACEPTIVES and making others more widely available are crucially important, says Rosemarie B. Thau of the Population Council.

The smaller dose may reduce the risk of breast cancer, which is associated with the pill. (At the same time, the pill seems to lower the risk of ovarian cancer.) The risk of breast and cervical cancer has led to opposition at various times, by some women's and consumer groups, to the approval of the pill.

Vaginal rings that release progestin, a progesteronelike compound, or a combination of estrogen and progestin are another form of hormonal manipulation. Because the hormones seep out steadily, "there are no peaks and valleys and, therefore, potentially fewer side effects," Thau says. Unlike the progestin-releasing NORPLANT, which is surgically implanted in the arm and which was developed by the Population Council, vaginal rings can be inserted and removed by the user. Although rings have been tested in many countries, they are not yet on the market anywhere.

Researchers are also no longer exempting men from hormonal vicissitudes. In a report in *Lancet* several years ago, researchers at WHO reported that injecting men once a week with a testosterone derivative could eliminate sperm in their ejaculate. Fertility was restored within a few months after stopping the injections. The group is now working to find longer-acting forms of testosterone so that the injections would be less frequent. And it is puzzling over one finding: the amount of sperm suppression varied geographically. Meanwhile Thau is working on a male implant that would also suppress sperm production.

A novel but longer-term approach seeks to harness immune responses. The reason that a woman's immune system does not perceive sperm as foreign remains a mystery—as does the reason that a man does not destroy his own sperm; since sperm do not appear until puberty, they could also be perceived as nonself. But studies of infertile couples who have somehow developed antibodies to each other's gametes are suggesting ways to develop birth control vaccines. The idea is to induce women and men to produce antibodies to proteins on sperm, explains Paul Primakoff, associate professor of physiology at the University of Connecticut, who has tested some vaccines in animals and observed reversible infertility.

Work on vaccines appears to be furthest along at the National Institute of Immunology in New Delhi. Researchers there, working in collaboration with the Population Council, have immunized men against luteinizing hormone–releasing hormone, a compound that contributes to the production of testosterone and sperm. Other collaborative trials there are looking at the effectiveness and safety of vaccinating women against human chorionic gonadotropin—a hormone produced by the embryo to maintain pregnancy.

Without increased funding, however, many efforts may never reach the public. "We can't really develop products with our limited budget. To make a product can cost between $300 and $350 million," laments Paul Van Look of WHO. "That is the sum total of money we have received in the past 20 years of our existence." Pharmaceutical companies have been reluctant to develop new contraceptives, despite $750 million in annual domestic sales of the pill. "A lot of them bowed out of this area because they felt liability was too high," Alexander says. In addition, companies viewed some of the FDA requirements for approval too intricate and too costly. In the past few years, however, the FDA has suspended several of its requirements.

Now, according to the Pharmaceutical Manufacturer's Association, seven companies are developing or considering developing contraceptives. "The industries are not interested in basic research, but they are interested in a major hit," says John C. Herr of the University of Virginia, who is also working on a contraceptive vaccine.

As a result, Van Look and others hope more companies will pick up their projects and take them to market. For example, a recent report in *Family Planning Perspectives,* a newsletter put out by the Alan Guttmacher Institute, a nonprofit organization, described a widely used but informal morning-after pill: two regular birth control pills taken within 72 hours of intercourse and two more, 12 hours later. Many family-planning experts hope companies will seek FDA approval for such a pill as well as for many methods available elsewhere. These include a variety of intrauterine devices, various permutations of the pill, RU 486 and related compounds, devices permitting reversible sterilization and different injectable contraceptives.

Changes on other fronts may be slower, though. Even if more methods were available, variety does not ensure use. Many family-planning organizations note that the lack of education and outreach as well as the cost of contraceptives can prevent people from using birth control. Although 95 percent of women of reproductive age in the U.S. use contraception, 37 percent of them rely on sterilization. Contraceptive fail-

Easing a Financial Gag

For nearly nine years, U.S. aid to family-planning programs was limited by a gag rule: no funds could be administered to any organization that performed abortions or provided counseling on abortions, even if U.S. dollars were not used for those purposes. In January, President Bill Clinton overturned this order, which was called the Mexico City Policy after the site where it was announced at a United Nations conference on population.

The policy "had a tremendous chilling effect, and the thaw is noticeable already," comments Mark Laskin, assistant secretary general of the International Planned Parenthood Federation (IPPF), which hopes to win back some of the $17 million a year that it lost as a result of the ban. "We will be able to help meet unmet need," Laskin adds, referring to the estimated 300 million couples worldwide who seek access to family planning.

But a lot more has to happen before the thaw is complete. Clinton must get approval from four congressional committees to reappropriate money. And while the IPPF may find some allocation forthcoming, the United Nations Fund for Population Activities (UNFPA) remains without U.S. backing for now. "There are hurdles still to be jumped," comments Alex Marshall of the UNFPA. The UNFPA was cut in 1985 as a result of the Kemp Amendment, which blocked subsidy of any organization thought to support programs forcing people to have abortions or to be sterilized. Repeated findings that the UNFPA was not involved in such activities did nothing to convince presidents Ronald Reagan and George Bush. To free money for the UNFPA now, Clinton must certify to Congress that the fund is not involved in such coercion.

It is also not clear whether the program in human reproduction at the World Health Organization (WHO) will receive funding. Because WHO works on compounds such as RU 486, which can induce menstruation after fertilization, U.S. aid is prohibited by the 1973 Helms Amendment and other congressional and administrative inhibitions. These policies stipulate that aid money cannot support abortion-related research. Changes on this front could take time since domestic issues will probably take priority, explains Sharon L. Camp, senior vice president of Population Action International. *—M.H.*

ure rates can be as high as 30 percent.

A better understanding of the sexual practices of Americans would help researchers pinpoint what is not working. "It is not just providing people with contraception, you also need individual education and community education: contraceptive failure rates are related to behavior," notes Lisa Kaeser of the Alan Guttmacher Institute. "All of us have been reliant on Kinsey data from the 1940s. We need a change."

But Senator Jesse Helms of North Carolina blocked funding for an NIH study of sexuality. In addition, support for the federal domestic family-planning program, which provides services for five million women, has fallen by two thirds since 1980, says Kathryn Kolbert of the Center for Reproductive Policy and Law.

And, of course, the abortion issue is unresolved. "Many of the problems with contraceptive development are attitudinal, and they have to do with the association of contraception with abortion," Mastroianni notes. "It is a paradox, because the best way to avoid abortion is to have more effective family planning."

The conflict over abortion is apparent in varying federal definitions of pregnancy and funding practices. The Agency for International Development defines pregnancy as fertilization, and thus, under the 1973 Helms Amendment, funding for research on compounds that act after fertilization is illegal. But because the NIH defines pregnancy as implantation, it can spend U.S. dollars researching methods that work after fertilization—methods that cannot be examined with U.S. foreign aid.

If this were not confusing enough, the NIH, in turn, is also prevented by law from studying methods to cause an abortion as well as contraception that interferes with implantation—unless the study is examined by the Ethics Advisory Board. The problem is, the board was disbanded in 1980. Thus, researchers must ignore aspects of a common medical procedure that causes some 125,000 deaths annually around the world.

The U.S. antiabortion lobby and longstanding abortion-related research policy have deterred the manufacturer of RU 486, Roussel-UCLAF, from seeking FDA approval. In February the company met with the FDA to explore the possibility of an agreement with another company or a research facility, which would apply for approval. Because of the threat of boycotts, Roussel-UCLAF reiterated its intention to avoid direct involvement.

But "the public has finally had enough of this," exclaims Mastroianni, with a warning that his age entitles him to climb on a soapbox anytime he has the opportunity. "Nothing is enduring. We just have to move the train again—get enough momentum up so that it will be hard to slow it down. We can't waste any time." *—Marguerite Holloway*

An Eternally Self-Reproducing Cosmos?

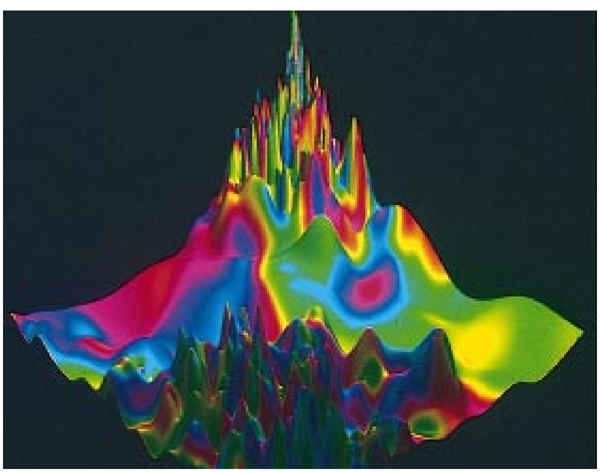
Until recently, people obsessed with the fate of the universe could ponder two rather bleak possibilities: either the cosmos keeps expanding forever, its matter dissipating into a cold, black void, or it collapses back onto itself in a cataclysmic "big crunch." For those who are willing to broaden their horizons, physicist Andrei D. Linde of Stanford University offers a less depressing scenario—the eternally self-reproducing universe.

Linde's theory builds on a concept he helped to devise called "inflation." It holds that just after the big bang, when the universe was fantastically small, hot and dense, it underwent a prodigious growth spurt before settling down to its current, relatively slow rate of expansion. The entire cosmos might have sprung from a minuscule fleck of space. "Most probably we are studying a universe that has been created by earlier universes," he adds.

Early versions of inflation, which relied heavily on particle physics, called for highly specialized, "fine-tuned" conditions. But Linde has shown that inflation might stem from more generic processes. According to quantum mechanics, space is never entirely empty; at very small scales, its energy content fluctuates violently. These chaotic quantum fluctuations, Linde says, could yield energy dense enough to trigger inflation.

Inflation is self-limiting: it rapidly attenuates the energy fueling it. But Linde contends that inflation is also self-perpetuating: quantum fluctuations will ensure that, somewhere, some mote of energy will keep sprouting into new universes. These universes may be radically unlike our own. Slight alterations in their initial conditions, Linde explains, could result in drastic changes in the way their physical laws are manifested after inflation ceases.

Working with his son, Dmitri, and others, Linde has simulated these ideas on a computer. "Whether you believe it or not, now we can show you," he says. The images depict a jagged, mountainlike terrain corresponding to a two-dimensional slice of space. Peaks indicate high-energy, inflationary regions; valleys represent regions of relatively low energy, such as our own, local universe, that have stopped inflating. Colors distinguish areas with different initial conditions—and laws of physics. Linde points out the mountainous pattern created by the differences in energy is fractal in nature: it recurs at scales ranging from trillions of times smaller than a proton to trillions of times bigger than the known universe.



FRACTAL FLUCTUATIONS in energy spawn infinite universes in a computer simulation by Andrei D. Linde.

Where's the evidence? Linde notes that the recent observations of "ripples" in faint microwaves thought to be the afterglow of our universe's fiery birth agree quite well with inflation's predictions. Estimates of the total mass of the universe also seem to be converging on the value predicted by inflation, enough to slow down but never quite stop the expansion of the universe—the local universe, that is. As for all those other universes blooming in the great beyond, they are separated from us by distances too vast to be breached by any currently conceivable method of observation.*—John Horgan*

Flat Chemistry

Enormous polymer sheets promise unusual properties

Polymer chemistry has entered a new dimension. Most polymers are nothing more than identical molecular units, or monomers, that are linked together to form one-dimensional chains. Now chemists have stitched together two-dimensional polymer sheets that have a variety of unusual properties. "There is a possibility of transforming all known monomers into two-dimensional objects," says Samuel I. Stupp, leader of the team at the University of Illinois that synthesized the polymer sheets. "If this possibility becomes reality, we would have a complete new set of materials with different properties."

Indeed, Stupp has already demonstrated that the polymer sheets have remarkable flexibility, strength and durability. The polymers might serve as lubricants, semiconductors, optical materials or selective membranes. "Until now, nobody has been able to make lots of two-dimensional objects that are self-contained and robust," comments Edwin L. Thomas, a materials scientist at the Massachusetts Institute of Technology. "The two-dimensional polymers may behave in ways that are not akin to things we already know."

Stupp's sheet polymers are among the largest molecules ever made by chemists, winning them the unattractive moniker "gigamolecules." The mass of a polymer is typically measured in daltons. A single carbon atom has a mass of 12 daltons. Amylopectin, one of the largest known polymers and the principal component of starches, is 90 million daltons. Stupp estimates that his molecules weigh much more than 10 million daltons. "The larger ones that we see by electron microscopy are beyond the molecular weight resolution of our instrumentation," he says.

To make the polymer sheets, Stupp reported in *Science,* he first prepares a precursor molecule by performing 21 different chemical reactions. The result is a rodlike molecule with two reactive sites: one in the center of the molecule and the other at one end.

It is perhaps easiest to understand how these precursors are assembled if one imagines that they are sharpened pencils. The eraser corresponds to the reactive end, and the brand name stamped on the pencil represents the central reactive site. The "brand name" encourages the pencils to align side by side in the same direction. The pencils therefore form a layer with the erasers on one side and the points on the other.

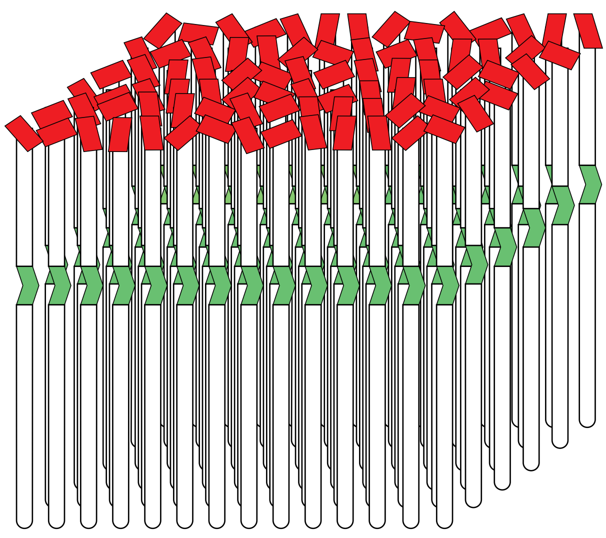
A second layer forms simultaneously on top of the first in such a way that the erasers in one layer touch those in the other. One of Stupp's key insights was to figure out how to sew these layers together. When heat is applied to the stacked layers, bonds are formed between the erasers and between the brand names, so connections are made within the two layers and between them. In this way, Stupp can construct a sheet whose area is typically one square micron and whose thickness is uniformly 0.005 micron. "The beauty of our method is we have some control over the size," Stupp remarks. "We can make either very small or very large sheets."

Chemists have been trying to synthesize polymer sheets for some time. During the past decade, workers at Harvard University and elsewhere have built two-dimensional molecular structures that were attached to sheets of gold or that rested on the surfaces of liquids. "The major problem inherent in these previous approaches is the poor stability of the structure," Thomas comments. So far Stupp is the only researcher who has succeeded in creating robust, free-floating polymer sheets.

The next major challenges for Stupp and his colleagues are, first, to attempt to make polymer sheets out of different building blocks and, second, to make bulk quantities of the polymers. "We have created four different kinds of polymer sheets by applying our original concept but using precursors that are easier to synthesize," Stupp explains.

Stupp is aware that he and other chemists have only limited bragging rights with respect to the two-dimensional polymers. Nature made them first. The membrane of red blood cells, for example, contains a protein gel, which is one kind of two-dimensional polymer. The gel is believed to serve as the flexible skeleton for the cells and plays a role in allowing them to change shape.

Although materials scientists have had little opportunity to characterize the gigamolecules, they are already thinking about some unusual applications. If the sheets are exposed to heat or placed in an acidic environment, they tend to roll up like a tobacco leaf around a cigar. Various substances could be wrapped up inside the polymer—a trick that might be useful for delivering pharmaceuticals into the body. Another possibility is building membranes that allow only certain molecules through. "I don't know what other applications might be possible," Thomas muses. "If I knew what they were, I'd be writing papers about them right now."—*Russell Ruthen*



***POLYMER SHEETS are made from rodlike precursors* (left) *with two reactive sites* (red *and* green). *A single gigamolecule can weigh more than 10 million daltons and be a few microns long, as shown in the electron micrograph* (right).**

Spot Marks the X

In females, one chromosome may lock itself inside an RNA

The mystery of calico cats is more than skin deep. The broad black and yellow patches in their fur are the outward manifestations of a more subtle genetic quirk. True calicoes are females, and like all female mammals, they carry two X chromosomes in their cells. Early in development, however, each embryonic cell randomly selects one X for future use and signals the other to condense permanently into an inert mass called a Barr body. (In this way, females achieve parity with males, which have only one X chromosome and a largely inactive Y.) In calico cats the resulting mosaicism is visible because each of their X chromosomes carries a different pigment gene.

After three decades of work, researchers are beginning to understand how mammalian cells manage to turn off an entire chromosome. The key appears to be a gene on the inactive X that produces an RNA molecule of unknown function. There are several explanations for how the gene accomplishes its feat. "My personal bias," remarks Carolyn J. Brown, one of the discoverers of the gene, "is that the RNA molecule is important in forming some kind of cage or structure that segregates the X and allows inactivation."

Brown and other members of the laboratory of Huntington F. Willard at Case Western Reserve University made their discovery while looking at gene expression on the Barr body. A few genes—about a dozen are now known in humans—evade the general "off" signal and therefore remain active on both X chromosomes. Yet in 1990 Willard's group found one gene that had a unique distinction: it was active only on the Barr body. Moreover, the gene was located in the small region of the X chromosome that previous research had determined was essential to X inactivation.

Those characteristics hinted that the gene, which Willard's group dubbed the X inactive-specific transcript gene (*Xist*), might play a pivotal part in turning off the X chromosome. Willard and Brown and their colleagues released word of *Xist* in January 1991. Several months later Sohaila Rastan and Neil Brockdorff and their colleagues at the Medical Research Council in Harrow, England, reported discovering a corresponding *Xist* gene in mice.

Last October in *Cell,* both the Willard and Rastan teams published their analyses of the human and mouse forms of *Xist.* The genes produce exceptionally large RNA molecules, and the human and mouse RNAs are generally similar to each other. Yet unlike most RNA, which leaves the cell nucleus and is translated into protein, the Xist RNA does not carry information for making proteins at all. Indeed, as Willard's experiments using fluorescent molecular probes showed, the Xist RNA never

Kitty, We Shrunk Your Brain

Chelsea Clinton and other cat lovers, don't take this the wrong way, but the brains of your pets aren't all that they used to be. The tabby curled on the sofa has lost almost a third of the neurons of its more robust Pleistocene ancestor. Such is the conclusion of Robert W. Williams of the University of Tennessee and Carmen Cavada and Fernando Reinoso-Suárez of the Independent University of Madrid. Their finding does not mean that cats have become more stupid—mercy, no. Rather it reveals a mechanism that may facilitate certain types of rapid evolutionary change.

The brains of domestic cats are not unusually tiny. If the brain sizes of lions, ocelots and all other feline species are plotted against their body weights, the domestic cat's brain falls neatly on the curve. "Its brain is exactly the size you'd expect based on its body size," Williams says. But, he observes, "even though people had studied those curves ad nauseam, nobody ever really knew what they meant in terms of cell number and cell size. What does it mean to say that the brain got smaller? Did it lose parts, or did the parts get smaller?"

In search of an answer, Williams, Cavada and Reinoso-Suárez compared the visual systems of modern house cats with those of Spanish wildcats (*Felis sylvestris tartessia*). Fossil evidence indicates that the Spanish animals are virtually indistinguishable from the wildcats that roamed northern Africa and Europe 20,000 years ago. The Spanish wildcats are taller and usually about twice the weight of the more familiar *F. catus.* Unlike feline homebodies, which are primarily nocturnal hunters, the wildcats hunt by day.

The clear-cut results of the comparison showed that "the reduction in brain weight involved the loss of brain cells," Williams says. Domestic cats had only about half as many neurons in the ganglia (nerve clusters) that connect their brain to their retinas. The wildcats had about 50 percent more neurons in their lateral geniculate nuclei, the brain structures that first receive signals from the optic nerves. In the retinas of the wildcats, the density of the cone photoreceptors—which make color vision and vision in bright light possible—was also more than twice as great. The researchers are confident that similar losses have occurred throughout the cat brain.

Twenty thousand years is relatively little time for so much change to have evolved. Williams thinks he and the others have found "a scintilla of evidence" about the mechanism. When they examined a wildcat embryo, they found that its brain contained approximately the same number of neurons as that of a domestic cat embryo. "So it looks plausible to us that the way the domestic cat got a smaller brain was by losing more cells rather than by producing fewer cells," Williams concludes.

Programmed cell death is a common feature of embryonic development for most animal species. In domestic cats, about 80 percent of the cells in the visual ganglia die before or shortly after birth—far more than in other vertebrates. Conceivably, then, the smaller modern cat species might have arisen fairly rapidly through a change in the developmental program that generally raised the amount of cell death. Williams cautions, however, that the idea "still really needs to be nailed down."

To Williams's knowledge, the study is the first attempt to compare species within an evolutionary lineage. The shrinkage in cats is not entirely human doing: most of it occurred long before people began domesticating cats less than 5,000 years ago. Indeed, because many mammals have become smaller since the last ice age, further work on other animals may find similar massacres of gray matter. Williams believes dogs are likely to be another example of "absurdly rapid evolution," much of it at the hands of human breeders. Cat fanciers may find some consolation in that thought: Who knows how much was deleted en route from Great Danes to Chihuahuas? *—John Rennie* seems to leave the nucleus. Instead it clusters tightly around the inactivated X chromosome that makes it.

Those results suggest several models for how inactivation might occur. One is that as the Xist RNA is produced, it binds to the chromosome, perhaps in association with other molecules. The resulting cage of RNA may directly incapacitate most genes. Alternatively, the presence of the RNA might enable the chromosome to interact with other factors on the nuclear membrane or elsewhere that deactivate it. Yet another possibility is that the RNA itself does not serve a function but that the act of transcription in that region induces conformational changes in the chromosome that lead to its inactivation.

In recent months the association between *Xist* and X inactivation has been further strengthened by Larry J. Shapiro of the University of California School of Medicine at San Francisco, Jacob Wahrman of the Hebrew University of Jerusalem, John R. McCarrey and Donald D. Dilworth of the Southwest Foundation for Biomedical Research in San Antonio and others. In independent studies, those investigators have found that the transcription of *Xist* precisely mirrors the inactivation of X chromosomes in various tissues.

In January, Graham F. Kay, another member of Rastan's group, announced that the transcription of *Xist* in early embryonic cells seems to precede X inactivation by a day or so. "That implies to us that *Xist* expression is not simply a consequence of X inactivation and supports the case that it could be causal," Brockdorff comments. Brown agrees that *Xist* is "a smoking pistol" but emphasizes that its importance during inactivation remains to be proved.

New experiments should settle that issue. "The idea we're working on is to knock out the *Xist* genes in an embryonic stem cell," Brockdorff explains. "If *Xist* is required, we should abolish the ability of those cells to undergo X inactivation." Investigators can also insert active copies of *Xist* into cells to see whether neighboring genes are shut off.

Other questions also remain. "If *Xist* is involved in X inactivation, then there is something that is turning it on or turning it off," Brown says. Researchers are keenly interested in determining how the Xist RNA interacts with the chromosome. At this point, they can only speculate about how the information concerning which X chromosome should be inactivated is passed from one cell to its progeny. Until those answers are found, researchers' understanding of X inactivation is likely to stay as patchy as the calico cat herself. *—John Rennie*

Illuminating Zero

Descending toward the coldest state of matter

If atoms of hydrogen could be cooled to absolute zero, they would not freeze into a solid or even condense into a liquid. Instead they would form an unusual type of gas known as a Bose condensation. In such a state the hydrogen atoms would have no velocity, and, by the laws of quantum mechanics, there would be no way to determine the precise positions of individual atoms. The entire gas would behave like one gigantic atom.

This ultimate state of matter may not be as faraway as absolute zero, however. Physicists are betting that a Bose condensation of hydrogen can be achieved at a balmy 30 microkelvins, that is, 30 millionths of a degree above absolute zero. And Jook T. M. Walraven and his colleagues at the University of Amsterdam have developed a new cooling trick that should help researchers reach the final frontier. He has succeeded in combining two techniques: laser and evaporative cooling.

In laser cooling, light is used to form an electromagnetic field that opposes the motion of atoms in a gas; this "optical molasses" slows atoms and thereby cools the gas. In evaporative cooling, the fastest atoms are allowed to escape from the gas, leaving the slow, cold atoms behind.

During the past decade, physicists have cooled atomic hydrogen using the evaporative technique, but the powerful laser method has been unavailable to them. The problem is that researchers have had difficulty generating light at a wavelength that an atom of hydrogen can readily absorb when it is in its lowest energy state. The key to Walraven's work was producing light of the appropriate wavelength. He and his co-workers employed a variety of conventional amplifiers and filters to transform a beam of visible laser light into weak pulses of ultraviolet photons (specifically, a wavelength of 121.6 nanometers).

To achieve ultralow temperatures, Walraven traps hydrogen in a magnetic field. The atoms are then exposed to ultraviolet pulses, which slows them in one direction. As the atoms interact with one another and the trap, they cool in all directions. In this way, he can reach temperatures around 8,000 microkelvins.

In the process the coldest atoms migrate to the center of the trap, whereas the hotter atoms oscillate from one side to the other. The hot atoms at the sides of the trap can be pushed out, once

SCIENTIFIC AMERICAN

CORRESPONDENCE

Reprints are available; to order, write Reprint Department, SCIENTIFIC AMERICAN, 415 Madison Avenue, New York, N.Y. 10017-1111 or fax inquiries to (212) 355-0408.

Back issues: $6 each ($7 outside U.S.) prepaid. Most numbers available. Credit card (Mastercard/Visa) orders for two or more issues accepted. To order, fax (212) 355-0408.

Index of articles since 1948 available in electronic format. Write SciDex™, SCIENTIFIC AMERICAN, 415 Madison Avenue, New York, N.Y. 10017-1111, or fax (212) 355-0408.

Photocopying rights are hereby granted by Scientific American, Inc., to libraries and others registered with the Copyright Clearance Center (CCC) to photocopy articles in this issue of SCIENTIFIC AMERICAN for the fee of $3.00 per copy of each article plus $0.50 per page. Such clearance does not extend to the photocopying of articles for promotion or other commercial purposes. Correspondence and payment should be addressed to Copyright Clearance Center, Inc., 27 Congress Street, Salem, Mass. 01970. Specify CCC Reference Number ISSN 0036-8733/93. $3.00 + 0.50.

Editorial correspondence should be addressed to The Editors, SCIENTIFIC AMERICAN, 415 Madison Avenue, New York, N.Y. 10017-1111. Manuscripts are submitted at the authors' risk and will not be returned unless accompanied by postage.

Advertising correspondence should be addressed to Advertising Manager, SCIENTIFIC AMERICAN, 415 Madison Avenue, New York, N.Y. 10017-1111, or fax (212) 754-1138.

Subscription correspondence should be addressed to Subscription Manager, SCIENTIFIC AMERICAN, P.O. Box 3187, Harlan, IA. 51537. The date of the last issue of your subscription appears on each month's mailing label. For change of address notify us at least four weeks in advance. Please send your old address (mailing label, if possible) and your new address.

again using the laser pulses. This process, a type of evaporative cooling, yields a gas colder than 3,000 microkelvins. "There is no fundamental reason why we couldn't go much lower," Walraven says.

"To put the whole thing together and make it work is a tour de force," comments Daniel Kleppner of the Massachusetts Institute of Technology, who describes himself as a friendly competitor. But, he adds, "it isn't so clear that what Walraven's done is really going to be that useful in getting to the Bose condensation."

Kleppner and his collaborators use an evaporative-cooling technique that is more conventional than Walraven's method. After trapping atomic hydrogen in a magnetic field, they allow the hottest atoms to escape by decreasing the strength of the field somewhat at one end of the trap. The procedure has been very successful, chilling atomic hydrogen to a record 100 microkelvins.

Kleppner believes he can produce a Bose condensation without resorting to laser cooling. Walraven begs to differ. His team can employ both standard evaporative cooling and the laser technique. "If you have light around, why not use it?" he asks.

Walraven and Kleppner must also contend with Carl Wieman and his co-workers at the University of Colorado at Boulder. Wieman has used a laser-cooling method on cesium atoms, instead of hydrogen, to attain the lowest temperature ever—one microkelvin. Wieman may not, however, be any closer than his rivals to achieving the ultimate goal. Because a cesium atom is 100 times heavier than hydrogen, cesium atoms will form a Bose condensation at a temperature much lower than hydrogen, according to theory.

Meanwhile Kleppner and his co-workers are struggling to develop a laser system that could detect and measure the Bose condensation. "It's anyone's guess about who is going to get there first," he remarks. *—Russell Ruthen*

Ancient Sleepers

Some bacteria cheat adverse conditions by folding themselves into tight, little balls and entering a state of suspended animation. As desiccated motes with all systems switched off, these endobacterial spores travel through time in search of food and water, the advent of which wakens them from their slumber.

No one knows how many centuries such microbial Methuselahs can traverse, in part because no one has bothered to scrutinize the scattered reports about them. Biologists are also deterred from making such systematic inquiry by the false positive results that have plagued efforts to recover ancient DNA. If the fragments of fossil genes can be so elusive, they reason, what are the chances of finding an entire organism whole—and viable?

Yet Max J. Kennedy and Sarah L. Reader of the New Zealand Institute for Industrial Research and Development support such an undertaking. Last year they announced in *Nature* the establishment of a data base for antediluvian microbes. "We saw all these anecdotal references in the literature," Kennedy says, "and thought a data base would be a great resource for evolutionary thinking." The researchers are also interested in the potential of long-lost organisms for industrial production of chemicals. "If these organisms were truly ancient, they might produce different chemicals from those that bacteria make now," he adds.

One case in point predates the data base by several years: a brand of the beer known as porter, brewed with yeast cultures salvaged from an 1825 shipwreck in the English Channel. Keith Thomas, a microbiologist at the University of Sunderland in England, was most interested in the chemical analysis of the first bottle dredged up. But then he found cells. "We opened the second bottle under sterile conditions and found cells again," Thomas says. He cultured the residue and—isolating the yeast from the bacteria and molds—applied it to an 1850 recipe for porter. The result was Flag Porter—some 50,000 bottles of it a year.

But two centuries are as nothing when compared with 117 of them. Gerald Goldstein, a microbiologist at Ohio Wesleyan University, believes he has succeeded in culturing bacteria that lived some 10,000 years ago in the gut of a mastodon entombed in a bog, now a water hazard of an Ohio golf course. The remains yielded convoluted, pink, smelly material from the region near the mastodon's bones. "I inoculated the material into a medium and cultured *Enterobacter cloacae,* which is normally found in the intestines of mammals," Goldstein says. "Of the 38 or 40 metabolic reactions we have carried out, there was only one difference with the species that exists today: it can digest a sugar called maltose."

As with most such finds, Goldstein's claims are being challenged. Carl R. Woese of the University of Illinois doubts the methodology. "There are other strains that don't metabolize maltose, and he happened to pull one of them out of the mastodon's gut," Woese says. "I don't know how to rule out a contaminant. Bacteria do seem to occur throughout the surface of the earth and to work their way into rocks." Indeed, contributors have debunked the most exciting citation in the new data base—a 1963 report of spores revived from salt deposited some 650 million years ago. The bacteria turned out to be of recent origin.

Yet there is a way to confirm that a bacterial culture is ancient, maintains Raul J. Cano of California Polytechnic State University: compare its genome with that of the original sample. The two DNA sequences should be identical. Then look at the corresponding sequences in kindred bacterial strains, together with the reconstructed sequence of the family's common ancestor. If the microbe in question is truly ancient, it should be more closely related to its ancestor than to any modern relative.

Cano says he sees no obvious limit to the life span of an endospore, although he allows that a billion years "might be a little too much." He has high hopes that he has revivified spores from the gut of a stingless bee entombed in amber between 25 and 40 million years ago. Muscle tissue from such bees yielded DNA—the oldest on record—as Cano and George O. Poinar, Jr., of the University of California at Berkeley reported in September in *Medical Science Research.* "These bees carry a bacillus that digests some of the more complex polysaccharides," Cano observes. "For bacteria, 40 million years should be enough to note important changes. You would expect them to have had different enzymes back then, because foods have changed."

Cano says federal regulations governing recombinant DNA and other exotic genetic material oblige him to keep his culture under tight security, but he deprecates the fear that the bugs might harm people. The modern species live on their own or inside insects and behave in culture in much the same way as denizens of his test tube. If released into the wild, he adds, "they'd probably just pick up where they left off." *—Philip E. Ross*

Asteroid Hunters

There's a rock out there with our name on it. Ho hum.

Astronomers who stalk the stray rocks that hurtle through the earth's part of the solar system are literally a rare breed. "Fewer people are involved in searching for near-earth asteroids than work in a McDonald's," reports David Morrison of the National Aeronautics and Space Administration's Ames Research Center. One of the most noteworthy is Steven J. Ostro of the Jet Propulsion Laboratory in Pasadena, Calif.—the world's sole expert in studying asteroids by radar.

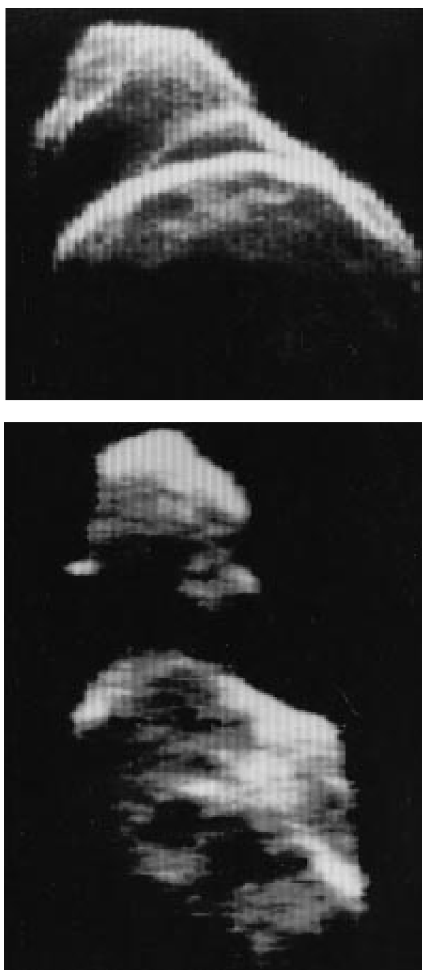
Last December, Ostro and his collaborators bounced a 400,000-watt radio signal from the Goldstone Deep Space Communications Complex in California off the asteroid Toutatis as it passed within a celestial hairbreadth of the earth. When Ostro's team analyzed the echoes, it recovered a "breakthrough" portrait of a remarkable object that consists of two battered rocks stuck together like Siamese twins. Three years ago Ostro and his colleagues produced a much fuzzier image of the asteroid Castalia, which indicated that it, too, is binary. "It's an amazing thing," says Morrison, who likens the discovery of twin asteroids to Galileo's observation that many stars are double.

The stunning images of Toutatis were possible mostly because the asteroid passed just 3.6 million kilometers from the earth, less than one tenth the distance to any other planet. Toutatis is but one of a whole class of interlopers whose orbits carry them well away from the main asteroid belt between Mars and Jupiter along paths that pass close to the orbit of the earth.

Their passage by the earth does not always result in a near-miss. A rocky body now estimated to have been approximately 60 meters wide flattened hundreds of square kilometers of forest in Siberia in 1908. The consensus estimate is that such impacts occur every 300 years or so. Objects the size of Toutatis, which measures roughly four kilometers across, strike far less often but are many orders of magnitude more destructive. An asteroid about 10 kilometers in diameter may have so disrupted the terrestrial environment that it caused the demise of the dinosaurs. "We're realizing that the earth exists in an asteroid swarm that time and again has dramatically altered the evolution of life on this planet," Ostro explains.

In 1992 two workshops sponsored by NASA addressed the question of how to detect and, in principle, deflect potential killer asteroids. Then, last November, Brian G. Marsden of the Harvard-Smithsonian Center for Astrophysics announced that Comet Swift-Tuttle might smack into the earth on August 14, 2126. That prediction, though since retracted, helped to publicize the impact threat that a handful of astronomers have worried about for years. "The probability of being hit by a large asteroid is exceedingly small," notes Tom Gehrels of the University of Arizona, a pioneer of near-earth asteroid hunts and the organizer of a recent symposium on asteroid hazards. "But if it happened, it would eliminate society."

Despite the high level of popular fascination, "we know terribly little about near-earth asteroids," Ostro laments. Even the most basic statistic—the number of the rocky missiles lurking out there—is unknown. Asteroid watchers have so far identified about 200 bodies whose orbits could bring them close to the earth. Richard P. Binzel of the Massachusetts Institute of Technology estimates that the total number of near-earth asteroids more than a kilometer across is about 10 times higher. But if smaller bodies are included, the tally balloons even more dramatically. By some calculations, there are perhaps a million objects upward of 50 meters in diameter. Over millions of years, many of these asteroids will inevitably slam into the earth, Binzel says.



RADAR SNAPSHOTS of the asteroid Toutatis reveal an irregular, heavily cratered binary object. These views were captured two days apart last December.

Counting how many near-earth asteroids really are out there is neither fast, easy nor particularly lucrative. Gehrels's project, known as Spacewatch, nearly perished in 1984 for lack of funds. In order to keep it alive, he resorted to supplemental fund-raising. He currently counts 230 individual contributors, including one person whose donation is contingent on the condition that Gehrels not tell the donor's wife where the money is going. Gehrels proudly reports that "public funding is quite strong" but adds that, even so, he could use more money to help finance a new 1.8-meter telescope on Kitt Peak.

Marsden notes that most near-earth asteroid survey programs are supported "on a shoestring" using retirees and volunteers. Participants in the surveys sometimes exhibit a kind of gallows humor about the marginal status of their work. Morrison, commenting on the paucity of researchers able to make radar studies of asteroids, quips that "we're all hoping Ostro isn't run over by a truck."

Current programs are turning up near-earth asteroids at the rate of a few dozen a year. One of the NASA-sponsored workshops outlined a more ambitious search called Spaceguard. The effort would use electronic detectors and a set of dedicated telescopes to uncover 90 percent of the threatening objects larger than about a kilometer across within about 25 years—at a cost of about $50 million up front and $10 million a year thereafter.

David J. Tholen of the Institute for Astronomy in Hawaii points to a major obstacle standing in the way of such a project: the sense of urgency, or rather the lack of one. "We could find 90 percent of the near-earth asteroids in a couple hundred years" using existing equipment, he points out. "If nobody's worried about getting hit in the next couple hundred years, there is no need for Spaceguard." Morrison calculates that kilometer-size asteroids (which are a serious hazard and are large enough to be reliably detected using present technology) hit once every 300,000 years or so.

Therein lies the dilemma of rationally evaluating a risk that is rare but potentially catastrophic. "Mass extinctions don't happen very often, but in reality you need only one," Marsden comments.

At the same time, he recognizes the problem of what he calls "the giggle factor" that afflicts asteroid-hazard research—in particular, the skeptical publicity engendered by proposals from some researchers, especially those at the Department of Defense, to deflect or destroy asteroids using tools ranging from nuclear weapons to giant solar sails. Binzel recounts being pleasantly surprised that the workshop at the University of Arizona set a sensible first goal: "to know what's out there."

As Ostro pursues that aim, his thoughts are far from mass extinctions. "We're seeing thousands of new 'worlds,'" he exclaims. "It's comparable to Columbus's exploration." He is hard at work producing refined images of Toutatis that will show details less than 100 meters wide, offering a window into the tumultuous history of near-earth asteroids. In 1995 the radio antennas at Goldstone and at Arecibo in Puerto Rico will be upgraded, at which time Ostro expects it will be possible to make comparably high resolution observations of asteroids approximately once a year. That information will help astronomers study the near-earth asteroids as an overall population and understand their place in the evolution of the solar system.

The recent attention to military solutions to the asteroid hazard is about to produce a significant scientific spin-off. In January 1994 the Strategic Defense Initiative Organization (SDIO), eager to find a compelling new project, will launch the Clementine mission. The $50-million *Clementine* spacecraft will fly within 100 kilometers of the asteroid Geographos in August of that year. Results from the flight will be shared with NASA and passed on to civilian scientists. Gehrels applauds the military efficiency with which the Clementine mission came together. "Fortunately, they did not do it through a committee report," he says acerbically.

Learning more about near-earth asteroids will undoubtedly be easier than devising a reasonable way to weigh the risk they pose. If Spaceguard goes ahead, it "will find things all the time that have a one in 10,000 chance of hitting the earth. That's just a fact of life we'll have to learn to live with," Binzel says. Marsden relates that asteroid orbits are sufficiently chaotic that even the most accurate data can predict no further than a century or two. Astronomers must come to terms with their double roles as solar system explorers and potential messengers of doom. As Binzel puts it, "The near-earth asteroids are our friends, but like all friends, they require respect." —*Corey S. Powell*

PROFILE: JOHN H. GIBBONS

The Nicest Guy in Washington

John H. Gibbons should look harried—at the very least—on this afternoon in February. Gibbons, who for 14 years advised Congress on technology-related matters as head of the Office of Technology Assessment (OTA), is the new science adviser to President Bill Clinton. He moved into the Old Executive Office Building, an excessively columned edifice a stone's throw from the White House, just after he was confirmed by Congress two weeks ago. He has been too busy to finish unpacking since then; boxes of files lie heaped around his large corner office.

Yesterday the president announced that he was fulfilling a campaign promise of trimming the White House staff by 25 percent. The order hit the Office of Science and Technology Policy (OSTP) and the Space Council, both of which Gibbons oversees, disproportionately hard; he must hack the combined staffs down from 95 to 46. Meanwhile there is policy to plot. This morning, Gibbons sat in on an hour-long Cabinet meeting on technology-investment strategies. That was followed by a two-hour conference with Vice President Al Gore.

When I finally meet Gibbons—at four o'clock, just after a Canadian minister and before another functionary—he greets me with a big grin and handshake. I see no sign of stress. He is a fit-looking 64-year-old, bald and ruddy-faced, with a demeanor both easygoing and earnest. Asked about the staff cuts, he responds as though he wishes they were his idea. With improved office technologies and administrative methods, he says, he and other White House officials should be able to make the reductions without any loss of productivity. "We're employed by the American people, and we ought to be at least as efficient as the private sector in these areas."

Trying to sum up what he sees as the essence of his new job, he says he hopes to "give the president and the vice president and other members of government—and in fact the American people—more effective access to the specialized knowledge of science and technology." He makes bureaucratic boilerplate ring like a silver bell.

ROBERT PROCHNOW

NEW PRESIDENTIAL ADVISER John H. ("Jack") Gibbons spent the past 14 years counseling Congress on technological issues.

When I ask how he has managed to work for so long in Washington without making any enemies, he laughs. "There's a story Tennessee Ernie Ford told," he replies, cranking his faint Southern twang up a notch, "about sitting on the side of a mountain drinking a big orange drink and watching these fellows down in a cow pasture playing this game, and he finally figured out the rule of the game was to take that little ball and run from one end of the cow pasture to the other without getting knocked down or stepping into something."

Then his compulsion to present all sides of the issue kicks in, and he tells me where I might find his critics. He notes that proponents of space-based defense, magnetically levitated trains and other megatechnologies treated skeptically by the Office of Technology Assessment sometimes called it the "Office of Technology Harassment." Some opponents of biotechnology also deplored a 1991 OTA report discounting the alleged dangers of milk from cows treated with genetically engineered hormones. "Call Jeremy Rifkin," he says.

Biotechnology gadflies aside, critics of Gibbons are vanishingly scarce in Washington. Certainly they are hard to find in Congress. During his confirmation hearing, members of the Senate Committee on Commerce, Science and Transportation spent two hours telling him how pleased they were and lobbing him softball questions on industrial policy and the proper mix of big and little science (totally ignoring the critical issue of alien nannies). The Senate confirmed him unanimously two days later.

Even the neoliberal magazine *The New Republic,* which eviscerated most of the president's other choices, gushed: "It's nice to note at least one Clinton appointment that wasn't motivated by diversity, cronyism or any criterion other than the nominee's demonstrated abilities." Frederick Seitz of the Marshall Institute, a cantankerous, conservative think tank, credited "Jack" Gibbons with having maintained the integrity of the OTA in spite of political pressures from both the left and the right. Seitz added, gratuitously, that Gibbons is "such a nice person you really can't say anything bad about him."

Before Gibbons was selected, some observers had suggested a biologist should be appointed science adviser, to reflect the fact that biology is supplanting physics as the most technologically and economically potent of the sciences. Others, particularly research scientists, had lobbied for a Nobel laureate or other luminary who could seek more funding for science. Gibbons is neither a biologist nor a Nobel winner.

But after Clinton made his choice, it was immediately apparent that no one was better suited to the job of science adviser to elected officials than someone who had held that job for 14 years. "It would be hard to find much daylight between his résumé and his job description," says John E. Pike, an analyst for the Federation of American Scientists who normally skewers inside-the-beltway technocrats.

Gibbons admits he thinks he's a pretty good choice, too. "In times past, I've frequently wondered what I wanted to do when I grow up," he said during his confirmation hearing. "Now I believe this new job is just that, since it will draw so completely on my past experience." He can, and does, claim to have seen science and technology from a variety of perspectives: bench scientist, academician and entrepreneur as well as administrator and policy adviser. Like almost all other science advisers, he was trained in physics; he obtained his undergraduate degree from Randolph-Macon College in Virginia (his home state) in 1949 and his doctorate from Duke University in North Carolina in 1954.

He spent the next 15 years at Oak Ridge National Laboratory studying nuclear physics, forging heavy elements in reactors in order to understand their origin in the solar system. "I call it solar system pediatrics," Gibbons says. In 1962 Gibbons and some co-workers used this expertise to start a company that sold radiation detectors and other instruments. Called Ortek, it was eventually sold to the electronics firm EG&G Corporation. Gibbons has also served on the boards of several other companies. This business experience, he says, should help him fulfill the administration's goals of building "new, productive bridges of cooperation and co-venturing between the private sector and the people of this nation."

In the late 1960s Gibbons's boss at Oak Ridge, the eminent nuclear physicist Alvin M. Weinberg, pointed out that after more than a decade of enjoying publicly funded research Gibbons should consider "shouldering some of the broader burden." Even before Weinberg approached him, Gibbons recalls, "I'd gotten interested in broader energy issues and the environment."

In 1969 Gibbons initiated a program at Oak Ridge that addressed how to conserve energy and minimize the impact of energy production and consumption on the environment. In 1973 he went to Washington to head up the first federal program on energy conservation. Two years later he returned to Tennessee to direct the University of Tennessee's Energy, Environment and Resources Center, and in 1989 he arrived at the OTA.

If the past is any guide, Gibbons will need to draw on all his experience and political skills in his new job. The OSTP is a descendant of the President's Science Advisory Committee, which originally consisted of prominent scientists who made recommendations on scientific issues regardless of the political consequences. The group's independence—especially over the issue of arms control—led President Lyndon B. Johnson to ignore it and President Richard M. Nixon to abolish it. Although scientists lobbied successfully for the creation of the OSTP in 1975, the office has had little influence since then.

Gibbons suggests the U.S. can no longer pursue big science projects without international help.

For example, President Ronald Reagan did not even consult his adviser, George A. Keyworth, Jr., before announcing the Strategic Defense Initiative. Keyworth was then reduced to serving as a cheerleader for the so-called Star Wars program. President George Bush's science adviser, D. Allan Bromley, a physicist at Yale University, managed to maintain somewhat more dignity during his tenure, but he reportedly had little influence on environmental issues, defense research and other areas.

"The office, inherently and for cause, is going to reflect the personalities and outlooks of the president and vice president," Gibbons notes. "One reason I was attracted to this job was my conviction, from statements the president and vice president have made, that they feel science is a source of new options. I think they called it the 'engine of growth.' "

So far it seems that Gibbons might enjoy greater clout than his predecessors—in spite of the cuts in his staff. First, Clinton nominated him in December; Bush did not select Bromley until three months after the inauguration, and Bromley did not assume the post for five months after that. Gibbons's early appointment allowed him to chime in on the many lower-level jobs still to be filled. The president also made Gibbons a member of the new Economic Policy Council, which is expected to play a major role in implementing the administration's economic plans.

Finally, there is Gibbons's relationship to Gore, who shares his passion for issues involving science and technology; they particularly agree on the need for maintaining a balance between economic growth and environmental conservation. "We resonated on this, because I think that's where the facts lead you," Gibbons remarks. "It's that conviction that a wise use of technology can provide human amenities with far less environmental impact, far less use of material resources, that is compelling to both of us." Yet Gibbons suggests that his links to Gore may have been exaggerated. He has "no idea" whether Gore recommended him for the job of science adviser, as some reports have surmised. "I never asked," he says.

Gibbons predicts that he will ruffle more feathers in his new job than he did in his old one. At the OTA, "we gave options rather than trying to come down on one side or the other of a particular decision," he remarks. "If you only give options, you don't tend to make a lot of enemies." In his new job, he adds, "I'm going to have to go further than that, in trying to focus on particular outcomes, so I probably won't enjoy such an easy and wide company of friends."

Indeed, scientists who have been calling for greater support for basic research may not like what Gibbons has to say on this topic. He notes that some scientific fields, including particle physics, have grown much faster than the overall economy during the past few decades. "That's known as a divergent series," Gibbons says. "It seems to me to be indefensible to say that science should forever have a rate of growth of support that is multiples of the growth of our resources."

Gibbons hints that the big science projects that have served as symbols of American ambition and prowess may survive only by attracting international support. "There are many things that we really not only can't but logically shouldn't do on a national basis," he says. Examples? He cites the space station, the Superconducting Super Collider, the Human Genome Project and the effort to build a fusion reactor. But Gibbons then sweetens his tough talk. The internationalization of science, he notes, "could be one of the most important things in the human experience." Look out: there's a nice new science adviser in town. *John Horgan*

SCIENTIFIC AMERICAN

The Aging of the Human Species

Our species has modified the evolutionary forces that have always limited life expectancy. Policymakers must consequently prepare to meet the needs of a population that will soon be much older

by S. Jay Olshansky, Bruce A. Carnes and Christine K. Cassel

For the first time, humanity as a whole is growing older. The demographic aging of the population began early in this century with improvements in the survival of infants, children and women of childbearing age. It will end near the middle of the next century when the age composition of the population stabilizes and the practical limits to human longevity are approached. No other species has ever exerted such control over the evolutionary selection pressures acting on it—or has had to face the resulting consequences.

Already the impact of the demographic transformation is making itself felt. In 1900 there were 10 million to 17 million people aged 65 or older, constituting less than 1 percent of the total population. By 1992 there were 342 million people in that age group, making up 6.2 percent of the population. By 2050 the number of people 65 years or older will expand to at least 2.5 billion people—about one fifth of the world's projected population. Barring catastrophes that raise death rates or huge inflations in birth rates, the human population will achieve a unique age composition in less than 100 years.

Demographers, medical scientists and other workers have anticipated the general aging of the human species for several decades, yet their attention has been focused almost exclusively on the concurrent problem of explosive population growth. We believe, however, that population aging will soon replace growth as the most important phenomenon from a policy standpoint. In a more aged population, the patterns of disease and disability are radically different. Many economic and social institutions that were conceived to meet the needs of a young population will not survive without major rethinking. Attitudes toward aging and the aged will have to be modified to address the demands of a much larger and more diverse older population.

Age structure is a characteristic of populations that reflects the historical trends in birth and death rates. Until recently, the shape of the human age structure was fairly constant.

Before the mid-19th century the annual death rates for humans fluctuated but remained high, between 30 and more than 50 deaths per 1,000 individuals. Those elevated, unstable rates were primarily caused by infectious and parasitic diseases. The toll from disease among the young was especially high. Often almost one third of the children born in any year died before their first birthday; in some subgroups, half died. Because childbirth was very hazardous, mortality among pregnant women was also high. Only a small segment of the population ever lived long enough to face the physiological decrements and diseases that accompany old age.

The only reason *Homo sapiens* survived such terrible early attrition was that the number of births more than compensated for the deaths. It was common for women to give birth to seven or more children in a lifetime. The higher birth rates were part of a successful survival pattern that reflected an array of favorable evolutionary adaptations made by humans.

Together the evolutionary constraints and adaptations produced a long-term average growth rate for the human species that, at least before the mid-19th century, hovered just above zero. The age structure of the population had the shape of a pyramid in which a large number of young children made up the broad base. At the apex were the few people who lived past their reproductive adulthood. The mean age of the population was low.

Clearly, much has changed since then.

S. JAY OLSHANSKY, BRUCE A. CARNES and CHRISTINE K. CASSEL have worked extensively on estimating the upper limits to human longevity. Olshansky is a research associate at the department of medicine, the Center on Aging, Health and Society and the Population Research Center of the University of Chicago. In 1984 he received his Ph.D. in sociology from that institution. Carnes, a scientist in the division of biological and medical research at Argonne National Laboratory, received his Ph.D. in statistical ecology from the University of Kansas in 1980. Cassel is chief of general internal medicine, director of the Center on Aging, Health and Society and professor of medicine and public policy at Chicago. She received her M.D. from the University of Massachusetts Medical Center in Worcester in 1976.

ELDERLY PEOPLE OF TOMORROW are only children today. For the first time, much of the population is living into advanced old ages. That demographic change carries potential risks. Reforms in social policy and further biological research may determine whether the additional years of life available to the population will be healthy and prosperous ones.

During the 20th century, the disparity between high birth rates and low death rates led to population growth rates that approached 2 to 3 percent and a population doubling time of only about 25 years. In the U.S. today, people aged 65 and older make up 12.5 percent of the population; by 2050 they will constitute 20 to 25 percent. This change is the result of declining mortality during the early and middle years. It was initially brought forth by improvements in sanitation and was later assisted by other public health measures and medical interventions. Collectively, they asserted control over the death rates from infectious and parasitic diseases and from maternal mortality.

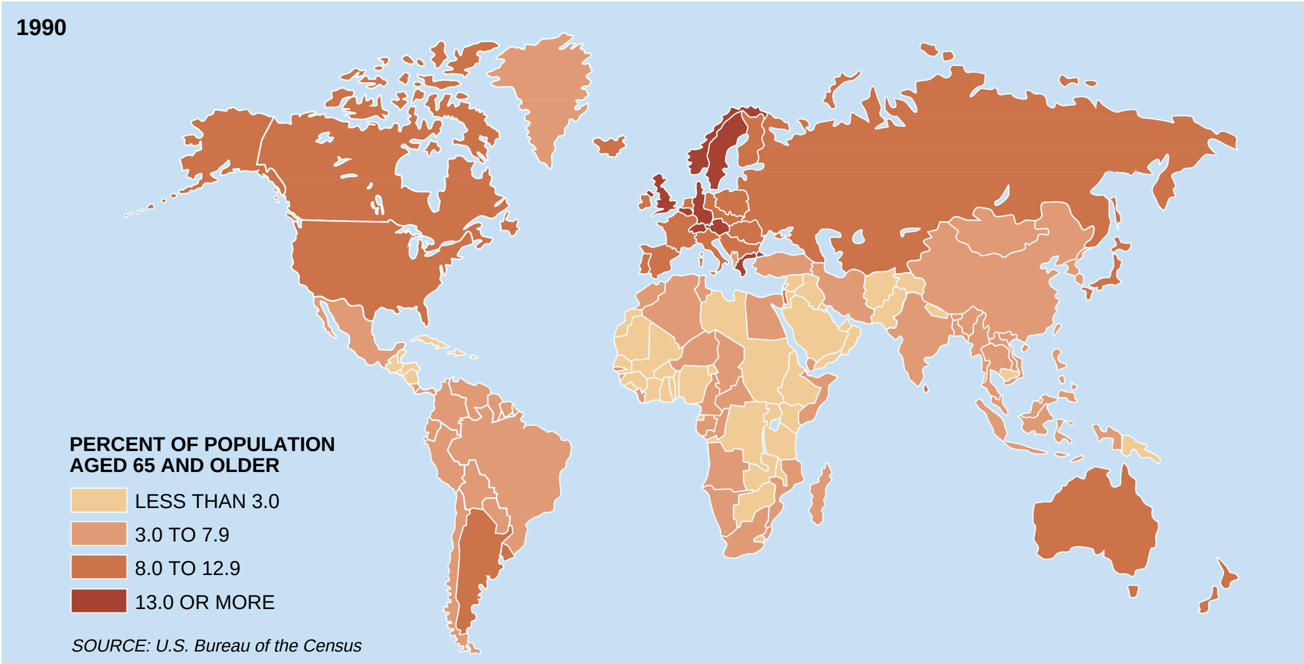
The series of steps by which a population ages has been the subject of considerable research. Indeed, the patterns of this demographic transformation and the speed with which they occur are central to understanding the social problems now on the horizon.

Initially, declines in infant, child and maternal death rates make the population younger by expanding the base of the age pyramid. Yet that improvement in survival, along with social and economic development, leads to a drop in birth rates and the beginning of population aging. Fewer births produce a narrowing of the pyramid's base and a relative increase in the number of people who are older.

As risk of death from infectious and parasitic diseases diminishes, the degenerative diseases associated with aging, such as heart disease, stroke and cancer, become much more important. Whereas infectious and parasitic diseases usually occur in cyclic epidemics, the age-related diseases are stable and chronic throughout an extended life. Consequently, the annual death rates fall from high, unstable levels to low, steady ones of eight to 10 persons per 1,000. Abdel R. Omran, when at the University of North Carolina at Chapel Hill, was the first to describe this change as an "epidemiologic transition." The rate of change and underlying causes of the transition differ among subgroups of the population.

In the final stage of the epidemiologic transition, mortality at advanced ages decreases as medical and public health measures postpone the age at which degenerative diseases tend to kill. For example, heart disease, stroke and cancer remain the primary causes of death, but healthier ways of life and therapeutic interventions permit people with those diseases to live longer. Disease onset and progression can also be delayed.

Once the birth and death rates in a population have been in equilibrium at



AGING OF THE WORLD POPULATION will become much more apparent during the 21st century. The trend is already pronounced in the industrialized countries. Within just a few decades, much of the population in the developing world will

low levels for one average life span—approximately 85 to 100 years—the age structure becomes almost permanently rectilinear: differences in the number of persons at various ages almost disappear. Thereafter, more than 90 percent of the people born in any year will live past the age of 65. About two thirds of the population could survive past 85, after which death rates would remain high and the surviving population will die rapidly. Such age structures have been observed in laboratory mice and other animals raised in controlled environments.

A crucial feature of the rectilinear age structure is its stability. If birth rates increase and temporarily widen its base, its rectilinear shape will gradually reassert itself because nearly all the members of the large birth generation will survive to older ages. Conversely, if the birth rate falls, the aging of the population will temporarily accelerate because the young become proportionally less numerous. The rectilinear age structure persists as long as early and middle-age mortality remain low.

The trend toward stable, low death rates has already been observed for a substantial segment of the world's population. Nevertheless, no nation has yet achieved a truly rectilinear age structure. Countries such as Sweden and Switzerland are much further along in the demographic transformation to population equilibrium than are other developed nations.

In the developed nations, two major phenomena have had a particularly noteworthy influence on the transformation of the age structure. The first is the post–World War II baby boom, the rise in birth rates that occurred during the middle of the century. Although 100 years is usually enough time for an age structure to become stable, the high birth rates of the baby boom postponed the aging of the population by widening the base of the age structure again. As the baby boomers grow older, however, the average age of the population will increase much faster. The stabilization process will probably take about 150 years for the developed nations, in which rectilinear age structures should become common by 2050.

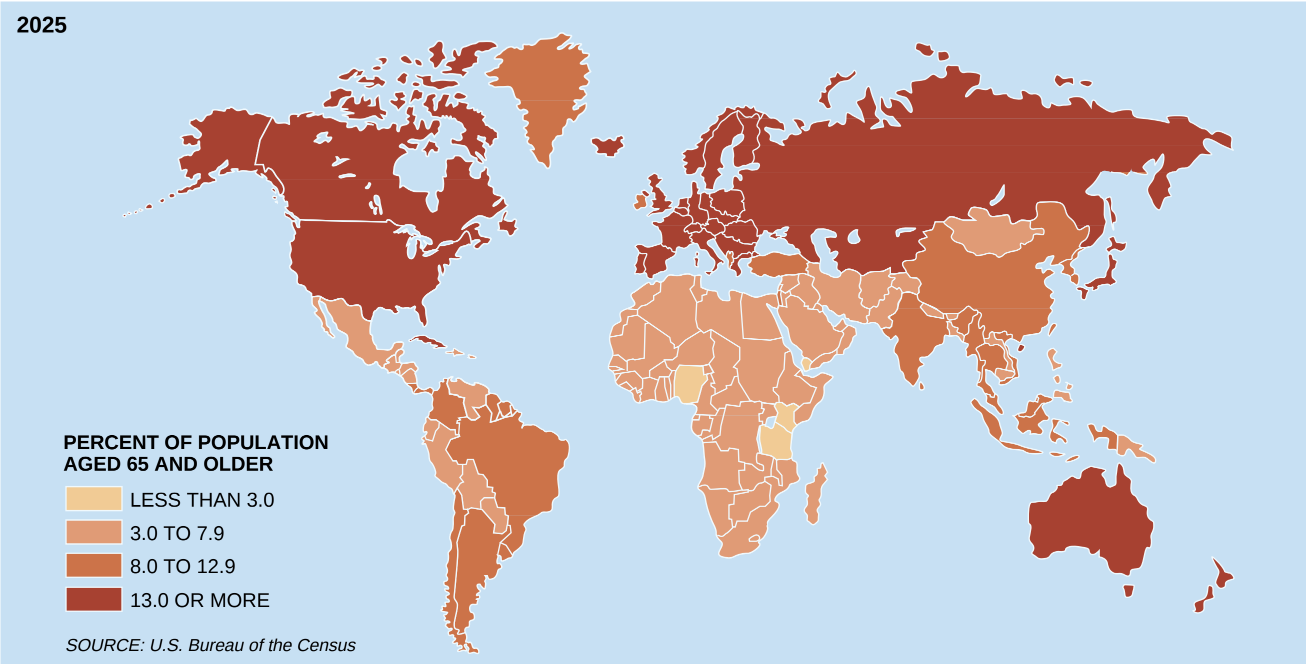
The second factor that influenced population aging in developed nations was the unexpected decline in old-age mortality that began in the late 1960s. Few scientists had anticipated that death rates from vascular disease could substantially be reduced at older ages. A fall in old-age mortality accelerates population aging by raising the age at which death becomes more frequent and the age structure begins to narrow. Death has become an event that occurs almost exclusively at older ages for some populations.

In many developing countries and in some groups within developed nations, human populations still face intense selection pressures. Consequently, some developing nations are not likely to reach equilibrium even by the middle of the 21st century. Nevertheless, the pace at which the population ages will accelerate throughout the developing world for the next 60 years.

For example, in China, which has both the largest population and the largest number of elderly people, the population aged 65 and older will increase from 6.4 percent (71 million people) to about 20 percent (270 million people) by 2050. China will then contain more people over 65 than the U.S. now has at all ages. India, which has the second largest elderly population, should experience even greater proportional increases.

We must emphasize that the demographic momentum for both population growth and population aging is already built into the age structures of all nations: the people who will become old in the next half century have, of course, already been born. These demographic forces will present a formidable set of social, economic and health problems in the coming decades—many of which are as yet unforeseen by policymakers and are beyond the capacity of developing countries to handle.

By the middle of the 21st century the transformation to an aged population should be complete for much of humanity. No one yet knows whether medical science will thereafter succeed in postponing the age at which rapid increases in the death rate begin. Will the apex of the age distribution retain its shape but shift to older ages, or will mortality be compressed into a shorter time span?



also be dramatically older. This demographic transformation is occurring because mortality at young ages has diminished. The social, medical and economic changes that accompany the aging of the population will pose significant problems.

The answer, which could profoundly affect economic and health issues, depends on whether there is an upper limit to longevity and a lower limit to the death rate.

For decades, the question of how low death rates can go has puzzled researchers. In 1978 demographer Jean Bourgeois-Pichat of Paris calculated that the average human life expectancy would not exceed 77 years. He arrived at that figure by theoretically eliminating all deaths from accidents, homicides, suicides and other causes unrelated to senescence. He then estimated the lowest death rates possible for cardiovascular disease, cancer and other diseases associated with aging. In effect, he eliminated all causes of death except those that seemed intrinsic to human biology. Yet shortly after its publication, Bourgeois-Pichat's life expectancy limit had already been exceeded in several nations. Other demographers have speculated that life expectancy will soon approach 100 years, but their theoretical estimates require unrealistic changes in human behavior and mortality.

In 1990 we took a more practical approach to the question of longevity. Rather than predicting the lower limits to mortality, we asked what mortality schedules, or age-specific death rates, would be required to raise life expectancy from its current levels to various target ages between 80 and 120 years. To determine the plausibility of reaching the targets, we compared those mortality schedules with hypothetical ones reflecting the elimination of cancer, vascular problems and other major fatal diseases. We demonstrated that as the actuarial estimate of life expectancy approaches 80 years, ever greater reductions in death rates are needed to produce even marginal increases in life expectancy.

Our conclusion was that life expectancy at birth is no longer a useful demographic tool for detecting declines in death rates in countries where mortality rates are already low. Furthermore, we suggested that the average life expectancy is unlikely to exceed 85 years in the absence of scientific breakthroughs that modify the basic rate of aging. Like others before us, we demonstrated that even if declines in death rates at older ages accelerate, the gains in life expectancy will be small.

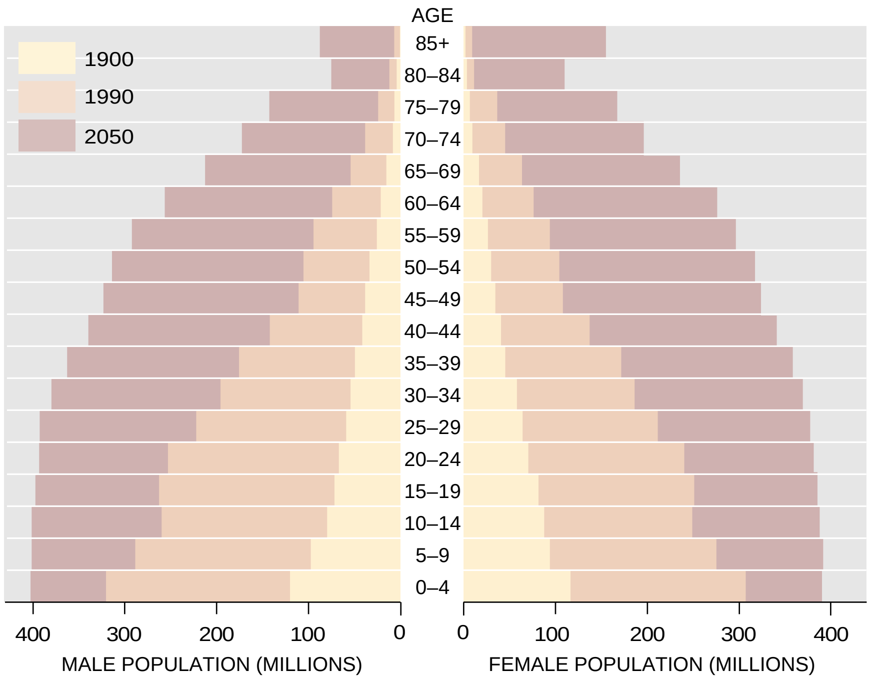
Why is the metric of life expectancy so insensitive to declining old-age mortality in low-mortality countries? First, for as long as reliable mortality statistics have been collected, the risk of death has always doubled about every eight years past the age of 30. That characteristic of human mortality has not changed despite the rapid declines in death rates at all ages during this century. A 38-year-old man today has a longer life expectancy than one from a century ago, but he is still twice as likely to die as a 30-year-old man.

Moreover, there is no indication that humans are capable of living much past the age of 110 regardless of declines in death rates from major fatal diseases. Thus, as death becomes ever more confined to older ages, the decline in death rates will inevitably stop. The point of deceleration occurs as life expectancy approaches 80 years.

Finally, in low-mortality countries, cardiovascular disease and cancer account for three of every four deaths after age 65. Those diseases are, in effect, competing for the lives of individuals, particularly at advanced ages. If the risk of dying from any single disease were reduced to zero, the saved population would simply be subject to high mortality risks from other causes—yielding a surprisingly small net gain in life expectancy. As deaths become concentrated into older ages, the competition among causes of mortality grows more pronounced.

Conceivably, however, medical researchers may learn how to slow the rate of senescence itself, thereby postponing the onset of degenerative diseases and the causes of old-age mortality. Toward that goal, many scientists working in the fields of evolutionary and molecular biology are now trying to learn why organisms become senescent.

In an influential paper written in 1957, evolutionary biologist George C. Williams, who was then at Michigan State University, proposed a mechanism for the evolution of senescence. His theory and subsequent predictions rested on two arguments. First, indi-



AGE STRUCTURE of the population is changing dramatically. For the past 100,000 years, the human age structure had the shape of a narrow pyramid. Since 1900, it has become wider and more rectilinear because relatively larger numbers of people in the growing population are surviving to older ages. By the middle of the 21st century it will be very nearly rectangular.

vidual genes are involved in multiple biological processes—a widely accepted concept known as pleiotropy. Second, he proposed that certain genes conferred survival advantages early in life but had deleterious physiological effects later. He then linked those assumptions to the prevailing concept that an individual's evolutionary fitness is measured by the genetic contribution that he or she makes to subsequent generations.

Williams then argued that an individual's odds of reproducing successfully would inevitably diminish over time because he or she would eventually die from an accident or some other uncontrollable cause. As individuals fulfill their reproductive potential, selection pressures should diminish, and any genes that had damaging effects later in life could not be eliminated by natural selection. Williams argued that this process, called antagonistic pleiotropy, provided a genetic basis for aging.

Another theory, proposed in 1977 by biologist T.B.L. Kirkwood of the National Institute for Medical Research in London, is a special case of antagonistic pleiotropy. He assumed that organisms must always divide their physiological energy between sexual reproduction and maintenance of the soma, or body. The optimum fitness strategy for a species, he argued, involves an allocation of energy for somatic maintenance that is less than that required for perfect repair and immortality. Senescence is therefore the inevitable consequence of the accumulation of unrepaired defects in the cells and tissues. Under Kirkwood's disposable soma theory, senescence is the price paid for sexual reproduction.

The disregulation of genes may provide a mechanism that links the antagonistic pleiotropy and disposable soma theories into a unified concept of disease and senescence. Two concepts central to the modern paradigm of molecular biology are required: gene regulation and pleiotropy. It is assumed in molecular biology that genes are carefully regulated and that the proteins produced by gene activity are typically involved in multiple, often interacting processes. Over time, a gradual accumulation of random molecular damage could disrupt the normal regulation of gene activity, potentially triggering a cascade of injurious consequences. Richard G. Cutler, a gerontologist at the National Institute on Aging, has referred to this process as the dysdifferentiative hypothesis of aging.

The severity of the consequences will depend on how critical the affected processes are at the time of their disregulation and the ability of the organism either to compensate for or to repair the damage. If the damage disrupts the regulation of cell growth or differentiation, cancer could result. Antagonistic pleiotropy describes cases where the temporal expression of a gene becomes disregulated. For example, a gene that is essential early in life may be harmful if expressed later. Gene disregulation and pleiotropy also provide a biological mechanism for the disposable soma theory. Aging may occur when the normal repair and maintenance functions of cells become disregulated and gradually degrade physiological function.

The accumulating evidence suggests that sites of molecular damage may not be entirely random. Some regions of the genome appear to be inherently unstable and may therefore be more susceptible to the disruption of gene regulation. When the damage occurs in somatic cells, disease or senescence, or both, may occur. The consequences of damage to the germ cells (eggs and sperm) run the gamut from immediate cell death to genetic changes that can be passed to the next generation. Propensities for disease and competency of somatic maintenance and repair are probably inheritable traits.

If there is a biological clock that begins ticking when a sperm fertilizes an egg, it probably does not go off at some predetermined date of death encoded in the genes. Rather the breakdown in gene regulation is a product of purely random events acting over a lifetime on a genome that contains inherited instabilities. As our understanding of biomolecular mechanisms grows, it may eventually become possible to manipulate disease processes and to slow the rate of senescence, thereby extending the average life span.

Although its link to molecular mechanisms is uncertain, one method of lengthening life span is known: dietary restriction. Early in the 20th century, researchers found that laboratory rats fed a low-calorie diet lived longer than those allowed to consume food at will. Those findings have been repeated for several species, including mice, flies and fish. Work by Richard Weindruch and his colleagues at the National Institute on Aging and by Roy L. Walford and his colleagues at the University of California at Los Angeles has suggested that dietary restriction may slow some parameters of aging in nonhuman primates.

These studies suggest life span can be extended by postponing—without eliminating—the onset of fatal diseases. Caloric restriction does not alter the rate of physiological decline in the experimental animals, nor does it change the doubling time for their death rate. Instead the animals appear to live longer because the age at which their death rates begin to increase exponentially is delayed. Dietary restriction seems to help preserve somatic maintenance for a longer time. Although it is not practical to expect enough people to adopt a calorically restricted diet to increase the average human life span, research may be able to identify the mechanisms at work and thereby extend longevity by other means.

Few observers had imagined that the demographic evolution of the human age structure would reveal a new set of diseases and causes of death. Will future reductions in old-age mortality reveal even more, new senescent diseases? Or will the prevalence of existing senescent diseases simply increase? Given the health care industry's focus on further reducing the impact of fatal diseases and postponing death, these issues will become critical to policymakers attempting to evaluate the consequences—both medical and economic—of an aging population.

One of the most important issues is whether the trend toward declining old-age mortality will generally benefit or harm the health of the overall population. In a controversial paper published 12 years ago, physician James F. Fries of Stanford University hypothesized that the biological limit to human life is fixed at about 85 years. Better life-styles and advances in medical technology, he said, will merely compress mortality, morbidity and disability into a shorter period near that limit. His underlying premise was that changes in diet, exercise and daily routines will postpone the onset age both of the major fatal diseases (heart disease, cancer and stroke) and of the debilitating diseases of old age (including Alzheimer's disease, osteoporosis and sensory impairments).

Fries's compression-of-morbidity hypothesis has since been challenged by many scientists who posit an expansion of morbidity. They argue that the behavioral factors known to reduce the risks from fatal diseases do not change the onset or progression of most debilitating diseases associated with aging. Further reductions in old-age mortality could therefore extend the time during which the debilitating diseases of aging can be expressed. In effect, an inadvertent consequence of the decline in old-age mortality may be a proportional rise in the untreatable disabilities now common among the very old. This view has been referred to as trading off longer life for worsening health.

The expansion-of-morbidity hypothesis serves as a consequence and a corollary to the evolutionary theories of aging. As a larger and more heterogeneous population survives into more advanced ages, the opportunities increase for the known senescent diseases to become more prevalent. New diseases associated with age (possibly resulting from the pleiotropic effects of gene disregulation) may also have a greater opportunity to manifest themselves.

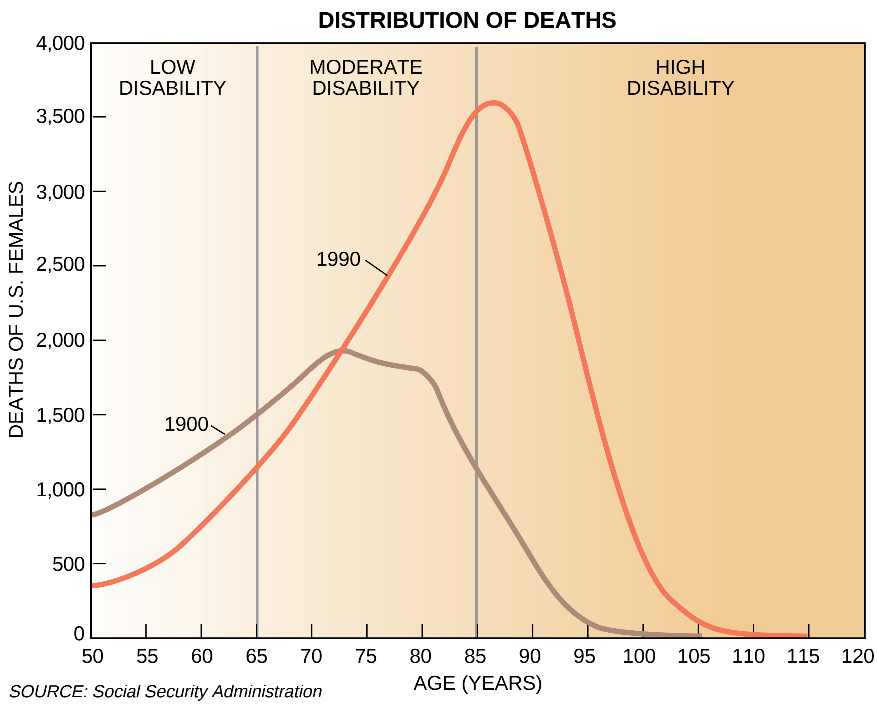
The ramifications of the expansion-of-morbidity hypothesis are so alarming that an international organization of scientists has been formed under the direction of demographer Jean-Marie Robine of INSERM in France to test its validity. The group's focus is the complex relation between declining old-age mortality and the relative duration of life spent healthy or disabled. Robine and his colleagues have demonstrated that women in Western societies can expect to spend up to one quarter of their lives disabled and men up to one fifth. Wealthier people are more likely to live longer and be healthier than those who are less well-off.

The data also suggested that recently the average number of years that people spend disabled has grown faster than those that they spend healthy. In other words, although people are enjoying more healthy years while they are young and middle-aged, they may be paying the price for those improvements by spending more time disabled when they are older. Because of the known problems of data reliability and comparability and of the short periods observed, current trends in morbidity and disability must be interpreted with caution.

The dilemma we face as a society is that medical ethics oblige physicians and researchers to pursue new technologies and therapeutic interventions in efforts to postpone death. Yet that campaign will inadvertently accelerate the aging of the population. Without a parallel effort to improve the quality of life, it may also extend the frequency and duration of frailty and disability at older ages. Society will soon be forced to realize that death is no longer its major adversary. The rising threat from the disabling diseases that accompany most people into advanced old age is already evident.

There is every reason for optimism that breakthroughs in molecular biology will permit the average life span to be modified. Just how far life span could be extended by slowing the rate of senescence is the subject of much speculation and debate. No one has yet demonstrated that human senescence can be modified by any means.

It is also unclear how those breakthroughs might influence the quality of life. If slowing the rate of senescence postpones all the physiological parameters of aging, then youth could be prolonged and disability compressed into a short time before death. If only some



PATTERNS OF DEATH AND DISABILITY are shifting as an epidemiologic transition occurs in the aging population. Because of healthier ways of life and medical interventions, people are surviving longer with heart disease, stroke and cancer. Yet because of their extended survival, they may suffer longer from the nonfatal but highly disabling illnesses associated with old age.

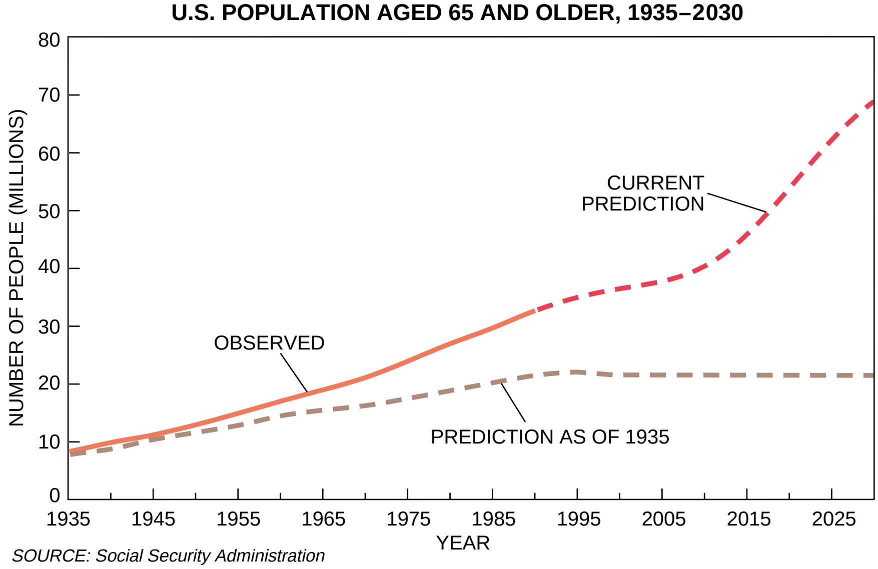
parameters of aging are amenable to modification, however, then the added years may become an extension of disabled life in old age.

We can identify with certainty some of the social problems that an aging population will face. Two of the most difficult will be the financial integrity of age-based entitlement programs, such as Social Security and Medicare, and the funding of health care. Social security programs in the U.S. and other countries were created when the age structures were still pyramidal and life expectancies were less than 60 years. The populations receiving benefits from those programs are much larger—and living considerably longer—than was anticipated at their inception. Given that the demographic momentum for larger and longer-lived older populations already exists, it is inescapable that such programs cannot survive in their present form much beyond the second decade of the next century.

Because declining mortality allows most people to survive past the age of 65, Medicare will need to cover tens of millions of people in the U.S. Many of them will need coverage for several decades. Medicare has few effective restraints on the use of expensive acute care, which is critical for treating many fatal illnesses. Yet it covers almost none of the expense of chronic long-term care—the need for which will grow as rapidly as the population ages. As a result, the cost of the Medicare program (like that of health care in general) will escalate swiftly, eroding the political will for systemic reforms that include long-term care. Can we continue to invest in ever more costly health care programs that are not designed to handle the unique demands of a growing and longer-lived aging population?

If during the next century life expectancy increases even marginally above the current estimates, the size of the beneficiary populations for age-entitlement programs will be two to five times greater than is already anticipated. That change would result in extreme financial hardship.

In the developed nations the demographic evolution of the age structure is beneficial in the short run: the coffers of the entitlement programs are swelling with the tax dollars from an unusually large cohort of working-age people. It would nonetheless be unwise to let that temporary condition lull us into complacency. When the age structure in those nations becomes rectilinear, the ratio of beneficiaries to taxpayers will mushroom, and surpluses in entitlement programs will vanish.



STRAINS ON SOCIAL PROGRAMS, such as Social Security and Medicare, will continue to emerge as the population ages and life expectancy increases. The number of beneficiaries in the Social Security program, for example, is growing much faster than was anticipated when the program was first conceived decades ago.

The financial integrity of age-entitlement programs has already been jeopardized in some countries. The worst problems will arise globally just after the year 2010, when the generation of baby boomers reaches entitlement age. The certainty of the demographic evolution of population aging will soon force governments to restructure all their entitlement programs.

The demographic evolution of the age structure will have an impact on many aspects of human society, including the job market, housing and transportation, energy costs, patterns of retirement, and nursing home and hospice care, to mention only a few. For example, if current trends toward early retirement persist, future retirees will draw benefits from age-entitlement programs for 30 years or more and spend up to one third of their lives in retirement. Thus, the current patterns of work and retirement will not be financially supportable in the future. Social structures have simply not evolved with the same rapidity as age structures. The rise in life expectancy is therefore a triumph for society, but many policy experts view it as an impending disaster.

Although we have emphasized the dark side of aging—frailty and disability—it is also true that the demographic evolution of the age structure will generate a large healthy, older population. All older people, both the healthy and the sick, will need the chance to contribute meaningfully to society. Achieving that end will require an economy that provides ample, flexible opportunities for experienced and skilled older persons, as well as modifications in the physical infrastructures of society. Changes in attitudes about aging will be essential.

The medical establishment is continuing to wage war against death. Researchers in the field of molecular biology are still searching for ways to slow the basic rate of aging. Those efforts lead us to believe that the aging of the population will also continue and perhaps even accelerate. Everybody wants to live longer, and medicine has helped that dream come true. Only now is society beginning to comprehend what it has set in motion by modifying the natural selection forces that have shaped the evolution of human aging.

FURTHER READING

IN SEARCH OF METHUSELAH: ESTIMATING THE UPPER LIMITS TO HUMAN LONGEVITY. S. J. Olshansky, B. A. Carnes and C. Cassel in *Science,* Vol. 250, pages 634–640; November 2, 1990.

EVOLUTION OF SENESCENCE: LATE SURVIVAL SACRIFICED FOR REPRODUCTION. T. B. L. Kirkwood and M. R. Rose in *Philosophical Transactions of the Royal Society of London,* Series B, Vol. 332, No. 1262, pages 15–24; April 29, 1991.

LIVING LONGER AND DOING WORSE? PRESENT AND FUTURE TRENDS IN THE HEALTH OF THE ELDERLY. Special issue of *Journal of Aging and Health,* Vol. 3, No. 2; May 1991.

THE OLDEST OLD. Edited by Richard Suzman, David Willis and Kenneth Manton. Oxford University Press, 1992.

AN AGING WORLD II. K. Kinsella and C. M. Taeuber. Center for International Research, U.S. Bureau of the Census, 1993.

Cavity Quantum Electrodynamics

Atoms and photons in small cavities behave completely unlike those in free space. Their quirks illustrate some of the principles of quantum physics and make possible the development of new sensors

by Serge Haroche and Jean-Michel Raimond

Fleeting, spontaneous transitions are ubiquitous in the quantum world. Once they are under way, they seem as uncontrollable and as irreversible as the explosion of fireworks. Excited atoms, for example, discharge their excess energy in the form of photons that escape to infinity at the speed of light. Yet during the past decade, this inevitability has begun to yield. Atomic physicists have created devices that can slow spontaneous transitions, halt them, accelerate them or even reverse them entirely.

Recent advances in the fabrication of small superconducting cavities and other microscopic structures as well as novel techniques for laser manipulation of atoms make such feats possible. By placing an atom in a small box with reflecting walls that constrain the wavelength of any photons it emits or absorbs—and thus the changes in state that it may undergo—investigators can cause single atoms to emit photons ahead of schedule, stay in an excited state indefinitely or block the passage of a laser beam. With further refinement of this technology, cavity quantum electrodynamic (QED) phenomena may find use in the generation and precise measurement of electromagnetic fields consisting of only a handful of photons. Cavity QED processes engender an intimate correlation between the states of the atom and those of the field, and so their study provides new insights into quantum aspects of the interaction between light and matter.

SERGE HAROCHE and JEAN-MICHEL RAIMOND work in a team of about a dozen researchers and students in the physics department of the École Normale Supérieure (ENS) in Paris. They have been studying the behavior of atoms in cavities for about 10 years. Haroche received his doctorate from ENS in 1971; he has been a professor of physics at Paris VI University since 1975. He has also been teaching and doing research at Yale University since 1984. In 1991 he became a member of the newly created Institut Universitaire de France. Raimond is also an alumnus of ENS; he earned his doctorate in 1984 working in Haroche's research group and is also a professor of physics at Paris VI University.

To understand the interaction between an excited atom and a cavity, one must keep in mind two kinds of physics: the classical and the quantum. The emission of light by an atom bridges both worlds. Light waves are moving oscillations of electric and magnetic fields. In this respect, they represent a classical event. But light can also be described in terms of photons, discretely emitted quanta of energy. Sometimes the classical model is best, and sometimes the quantum one offers more understanding.

When an electron in an atom jumps from a high energy level to a lower one, the atom emits a photon that carries away the difference in energy between the two levels. This photon typically has a wavelength of a micron or less, corresponding to a frequency of a few hundred terahertz and an energy of about one electron volt. Any given excited state has a natural lifetime—similar to the half-life of a radioactive element—that determines the odds that the excited atom will emit a photon during a given time interval. The probability that an atom will remain excited decreases along an exponential curve: to one half after one tick of the internal clock, one quarter after two ticks, one eighth after three and so on.

In classical terms, the outermost electron in an excited atom is the equivalent of a small antenna, oscillating at frequencies corresponding to the energy of transitions to less excited states, and the photon is simply the antenna's radiated field. When an atom absorbs light and jumps to a higher energy level, it acts as a receiving antenna instead.

If the antenna is inside a reflecting cavity, however, its behavior changes—as anyone knows who has tried to listen to a radio broadcast while driving through a tunnel. As the car and its receiving antenna pass underground, they enter a region where the long wavelengths of the radio waves are cut off. The incident waves interfere destructively with those that bounce off the steel-reinforced concrete walls of the tunnel. In fact, the radio waves cannot propagate unless the tunnel walls are separated by more than half a wavelength. This is the minimal width that permits a standing wave with at least one crest, or field maximum, to build up—just as the vibration of a violin string reaches a maximum at the middle of the string and vanishes at the ends. What is true for reception also holds for emission: a confined antenna cannot broadcast at long wavelengths.

An excited atom in a small cavity is precisely such an antenna, albeit a microscopic one. If the cavity is small enough, the atom will be unable to radiate because the wavelength of the oscillating field it would "like" to produce

CAVITY QED apparatus in the authors' laboratory contains an excitation zone for preparing a beam of atoms in highly excited states (*left*) and a housing surrounding a superconducting niobium cavity (*center*). Ionization detectors (*right*) sense the state of atoms after they have passed through the cavity. The red laser beam traces the line of the infrared laser used to excite the atoms; the blue beam marks the path of the atoms themselves. When in use, the entire apparatus is enclosed in a liquid-helium cryostat that cools it to less than one kelvin.

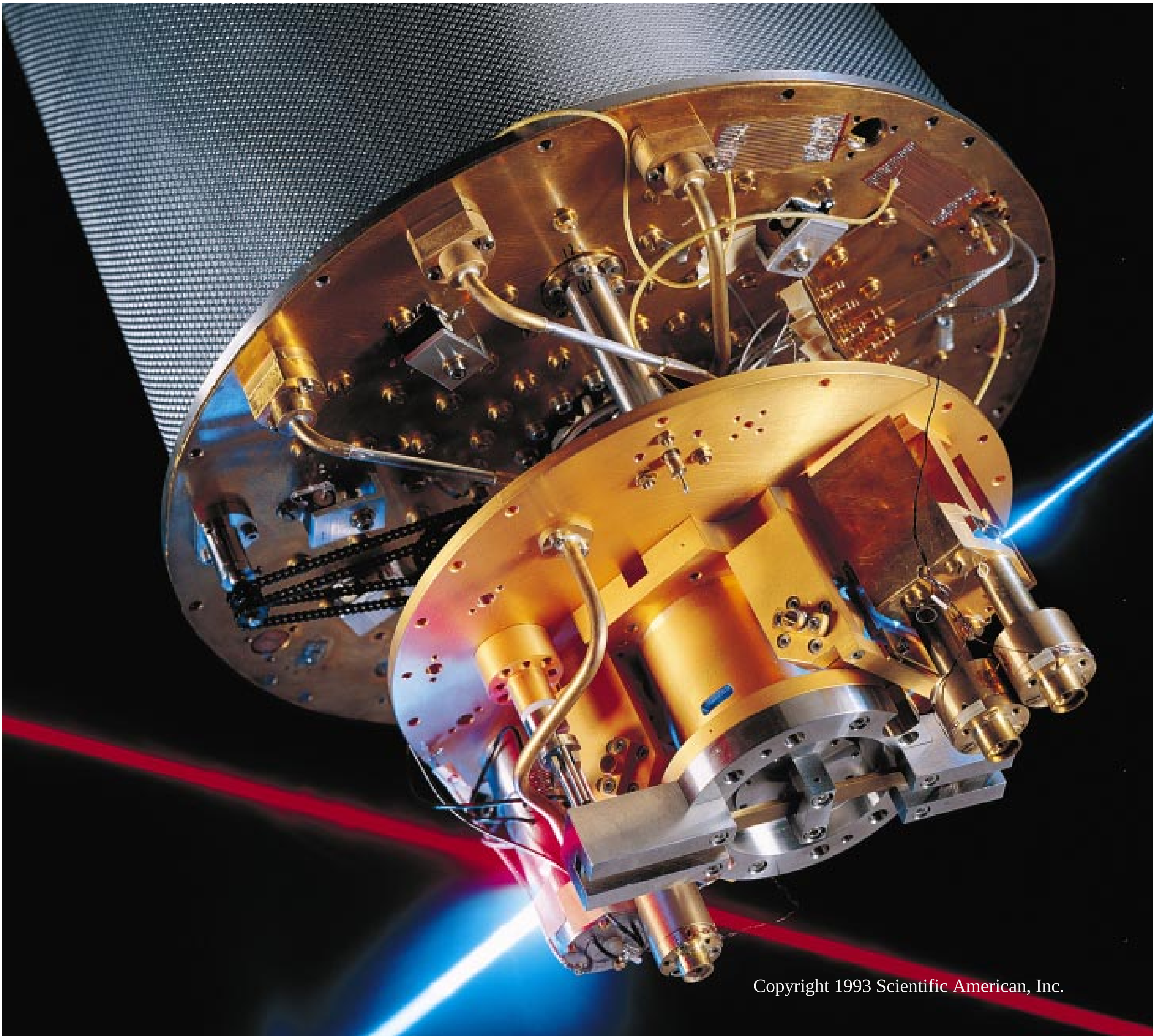
cannot fit within the boundaries. As long as the atom cannot emit a photon, it must remain in the same energy level; the excited state acquires an infinite lifetime.

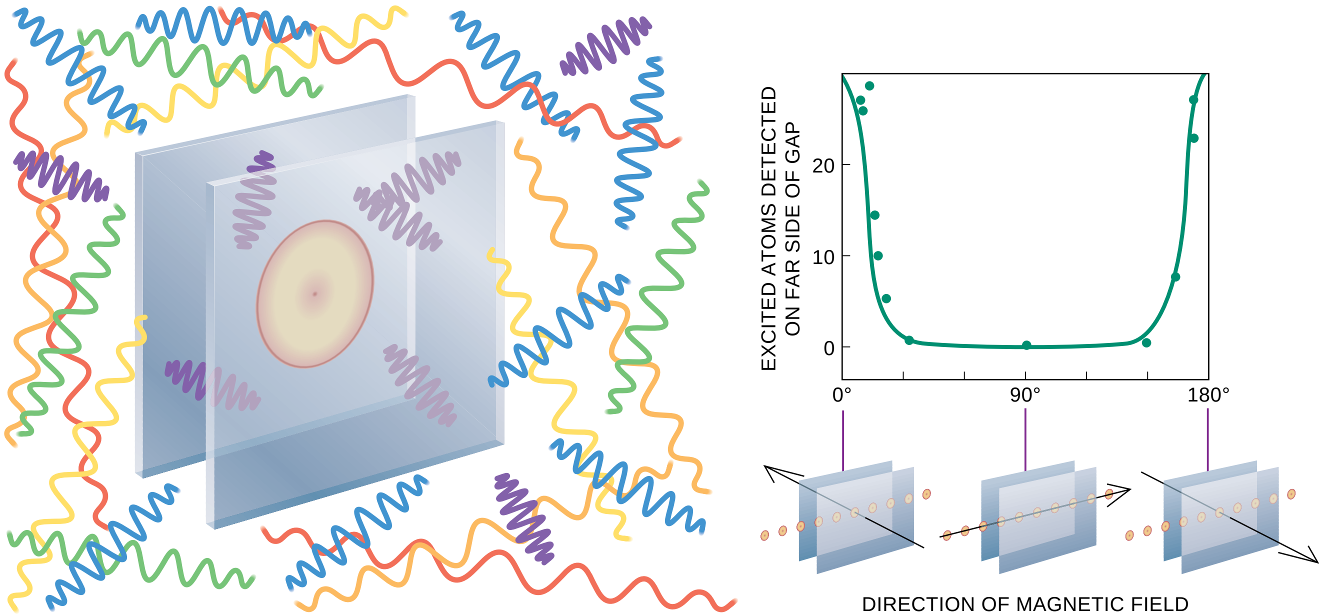
In 1985 research groups at the University of Washington and at the Massachusetts Institute of Technology demonstrated suppressed emission. The group in Seattle inhibited the radiation of a single electron inside an electromagnetic trap, whereas the M.I.T. group studied excited atoms confined between two metallic plates about a quarter of a millimeter apart. The atoms remained in the same state without radiating as long as they were between the plates.

Millimeter-scale structures are much too wide to alter the behavior of conventionally excited atoms emitting micron or submicron radiation; consequently, the M.I.T. experimenters had to work with atoms in special states known as Rydberg states. An atom in a Rydberg state has almost enough energy to lose an electron completely. Because this outermost electron is bound only weakly, it can assume any of a great number of closely spaced energy levels, and the photons it emits while jumping from one to another have wavelengths ranging from a fraction of a millimeter to a few centimeters. Rydberg atoms are prepared by irradiating ground-state atoms with laser light of appropriate wavelengths and are widely used in cavity QED experiments.

The suppression of spontaneous emission at an optical frequency requires much smaller cavities. In 1986 one of us (Haroche), along with other physicists at Yale University, made a micron-wide structure by stacking two optically flat mirrors separated by extremely thin metallic spacers. The workers sent atoms through this passage, thereby preventing them from radiating for as long as 13 times the normal excited-state lifetime. Researchers at the University of Rome used similar micron-wide gaps to inhibit emission by excited dye molecules.

The experiments performed on atoms between two flat mirrors have an interesting twist. Such a structure, with no sidewalls, constrains the wavelength only of photons whose polarization is parallel to the mirrors. As a result, emission is inhibited only if the atomic dipole antenna oscillates along the





EXCITED ATOM between two mirrors (*left*) cannot emit a photon. The atom is sensitive to long-wavelength vacuum fluctuations whose polarization is parallel to the mirrors, but the narrow cavity prevents such fluctuations. Atoms passing through a micron-wide gap between mirrors have remained in the excited state for 13 natural lifetimes. Subjecting the atoms to a magnetic field causes their dipole axes to precess and changes the transmission of excited atoms through the gap (*right*). When the field is parallel to the mirrors, the atom rotates out of the plane of the mirrors and can quickly lose its excitation.

plane of the mirrors. (It was essential, for example, to prepare the excited atoms with this dipole orientation in the M.I.T. and Yale spontaneous-emission inhibition experiments.) The Yale researchers demonstrated these polarization-dependent effects by rotating the atomic dipole between the mirrors with the help of a magnetic field. When the dipole orientation was tilted with respect to the mirrors' plane, the excited-state lifetime dropped substantially.

Suppressed emission also takes place in solid-state cavities—tiny regions of semiconductor bounded by layers of disparate substances. Solid-state physicists routinely produce structures of submicron dimensions by means of molecular-beam epitaxy, in which materials are built up one atomic layer at a time. Devices built to take advantage of cavity QED phenomena could engender a new generation of light emitters [see "Microlasers," by Jack L. Jewell, James P. Harbison and Axel Scherer; SCIENTIFIC AMERICAN, November 1991].

These experiments indicate a counterintuitive phenomenon that might be called "no-photon interference." In short, the cavity prevents an atom from emitting a photon because that photon would have interfered destructively with itself had it ever existed. But this begs a philosophical question: How can the photon "know," even before being emitted, whether the cavity is the right or wrong size?

Part of the answer lies in yet another odd result of quantum mechanics. A cavity with no photon is in its lowest-energy state, the so-called ground state, but it is not really empty. The Heisenberg uncertainty principle sets a lower limit on the product of the electric and magnetic fields inside the cavity (or anywhere else for that matter) and thus prevents them from simultaneously vanishing. This so-called vacuum field exhibits intrinsic fluctuations at all frequencies, from long radio waves down to visible, ultraviolet and gamma radiation, and is a crucial concept in theoretical physics. Indeed, spontaneous emission of a photon by an excited atom is in a sense induced by vacuum fluctuations.

The no-photon interference effect arises because the fluctuations of the vacuum field, like the oscillations of more actual electromagnetic waves, are constrained by the cavity walls. In a small box, boundary conditions forbid long wavelengths—there can be no vacuum fluctuations at low frequencies. An excited atom that would ordinarily emit a low-frequency photon cannot do so, because there are no vacuum fluctuations to stimulate its emission by oscillating in phase with it.

Small cavities suppress atomic transitions; slightly larger ones, however, can enhance them. When the size of a cavity surrounding an excited atom is increased to the point where it matches the wavelength of the photon that the atom would naturally emit, vacuum-field fluctuations at that wavelength flood the cavity and become stronger than they would be in free space. This state of affairs encourages emission; the lifetime of the excited state becomes much shorter than it would naturally be. We observed this emission enhancement with Rydberg atoms at the École Normale Supérieure (ENS) in Paris in one of the first cavity QED experiments, in 1983.

If the resonant cavity has absorbing walls or allows photons to escape, the emission is not essentially different from spontaneous radiation in free space—it just proceeds much faster. If the cavity walls are very good reflectors and the cavity is closed, however, novel effects occur. These effects, which depend on intimate long-term interactions between the excited atom and the cavity, are the basis for a series of new devices that can make sensitive measurements of quantum phenomena.

Instead of simply emitting a photon and going on its way, an excited atom in such a resonant cavity oscillates back and forth between its excited and unexcited states. The emitted photon remains in the box in the vicinity of the atom and is promptly reabsorbed. The atom-cavity system oscillates between two states, one consisting of an excited atom and no photon, and the other of a de-excited atom and a photon trapped in the cavity. The frequency of this oscillation depends on the transition energy, on the size of the atomic dipole and on the size of the cavity.

This atom-photon exchange has a deep analogue in classical physics. If two identical pendulums are coupled by a weak spring and one of them is set in motion, the other will soon start swinging while the first gradually comes to rest. At this point, the first pendulum starts swinging again, commencing an ideally endless exchange of energy. A state in which one pendulum is excited and the other is at rest is clearly not stationary, because energy moves continuously from one pendulum to the other. The system does have two steady states, however: one in which the pendulums swing in phase with each other, and the other in which they swing alternatively toward and away from each other. The system's oscillation in each of these "eigenmodes" differs because of the additional force imposed by the coupling—the pendulums oscillate slightly slower in phase and slightly faster out of phase. Furthermore, the magnitude of the frequency difference between the two eigenmodes is precisely equal to the rate at which the two pendulums exchange their energy in the nonstationary states.

Researchers at the California Institute of Technology recently observed this "mode splitting" in an atom-cavity system. They transmitted a weak laser beam through a cavity made of two spherical mirrors while a beam of cesium atoms also crossed the cavity. The atomic beam was so tenuous that there was at most one atom at a time in the cavity. Although the cavity was not closed, the rate at which it exchanged photons with each atom exceeded the rate at which the atoms emitted photons that escaped the cavity; consequently, the physics was fundamentally the same as that in a closed resonator.

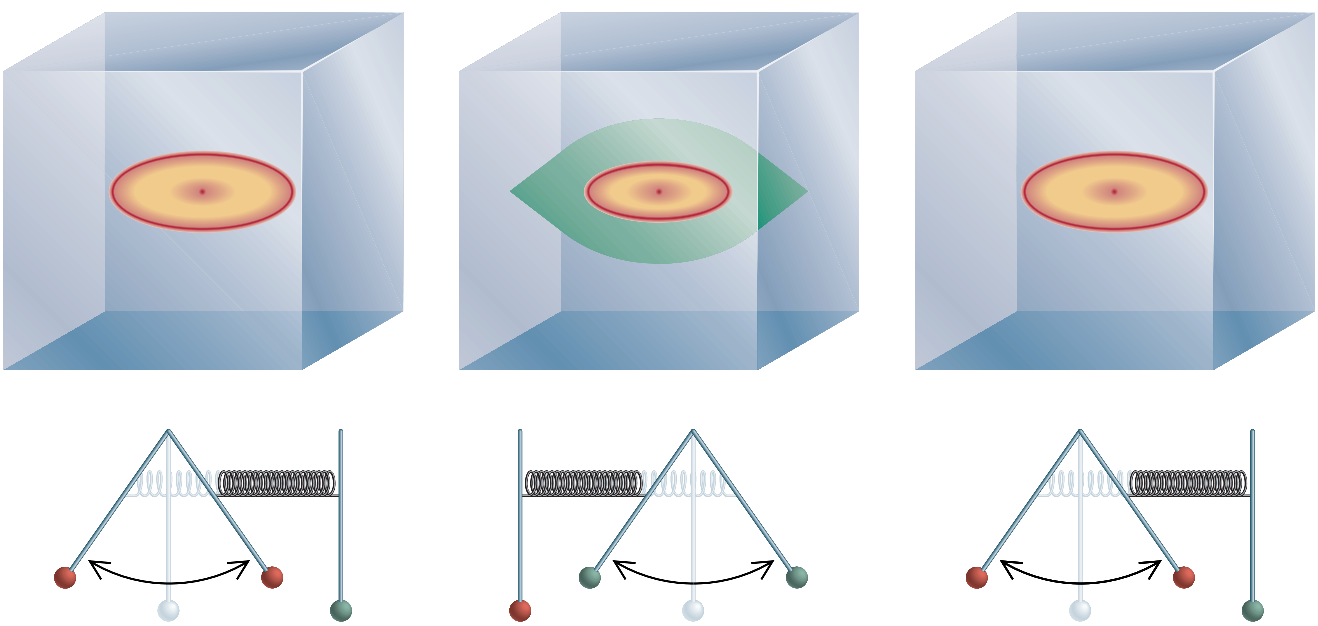
The spacing between the mirrors was an integral multiple of the wavelength of the transition between the first excited state of cesium and its ground state. Experimenters varied the wavelength (and hence frequency) of the laser and recorded its transmission across the cavity. When the cavity was empty, the transmission reached a sharp maximum at the resonant frequency of the cavity. When the resonator contained one atom on average, however, a symmetrical double peak appeared; its valley matched the position of the previous single peak. The frequency splitting, about six megahertz, marked the rate of energy exchange between the atom and a single photon in the cavity.

This apparatus is extremely sensitive: when the laser is tuned to the cavity's resonant frequency, the passage of a single atom lowers transmission significantly. This phenomenon can be used to count atoms in the same way one currently counts cars or people intercepting an infrared light in front of a photodetector.

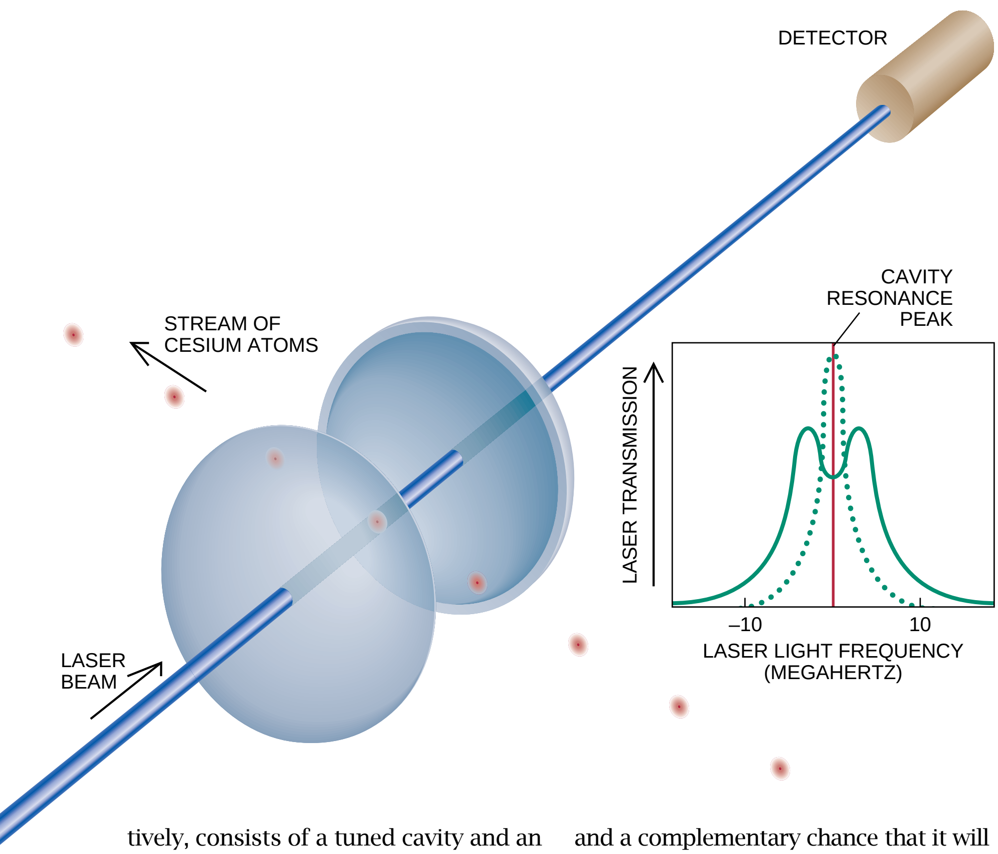
Although simple in principle, such an experiment is technically demanding. The cavity must be as small as possible because the frequency splitting is proportional to the vacuum-field amplitude, which is inversely proportional to the square root of the box's volume. At the same time, the mirrors must be very good reflectors so that the photon remains trapped for at least as long as it takes the atom and cavity to exchange a photon. The group at Caltech used mirrors that were coated to achieve 99.996 percent reflectivity, separated by about a millimeter. In such a trap, a photon could bounce back and forth about 100,000 times over the course of a quarter of a microsecond before being transmitted through the mirrors.

Experimenters have been able to achieve even longer storage times—as great as several hundred milliseconds—by means of superconducting niobium cavities cooled to temperatures of about one kelvin or less. These cavities are ideal for trapping the photons emitted by Rydberg atoms, which typically range in wavelength from a few millimeters to a few centimeters (corresponding to frequencies between 10 and 100 gigahertz). In a recent experiment in our laboratory at ENS, we excited rubidium atoms with lasers and sent them across a superconducting cylindrical cavity tuned to a transition connecting the excited state to another Rydberg level 68 gigahertz higher in energy. We observed a mode splitting of about 100 kilohertz when the cavity contained two or three atoms at the same time.

There is a striking similarity between the single atom-cavity system and a laser or a maser. Either device, which emits photons in the optical and microwave domain, respec-



ATOM IN A CAVITY with highly reflective walls can be modeled by two weakly coupled pendulums. The system oscillates between two states. In one, the atom is excited, but there is no photon in the cavity (*left* and *right*). In the other, the atom is de-excited, and the cavity contains a photon (*center*). The atom and the cavity continually exchange energy.



LASER BEAM TRANSMISSION through a cavity made of two closely spaced spherical mirrors is altered by the passage of individual atoms. When the cavity is empty, transmission peaks at a frequency set by the cavity dimensions (*dotted curve*). When an atom resonant with the cavity enters, however, the atom and cavity form a coupled-oscillator system. Transmission peaks at two separate frequencies corresponding to the "eigenmodes" of the atom-cavity system. The distance between the peaks marks the frequency at which the atom and cavity exchange energy.

tively, consists of a tuned cavity and an atomic medium that can undergo transitions whose wavelength matches the length of the cavity. When energy is supplied to the medium, the radiation field inside the cavity builds up to a point where all the excited atoms undergo stimulated emission and give out their photons in phase. A maser usually contains a very large number of atoms, collectively coupled to the radiation field in a large, resonating structure. In contrast, the cavity QED experiments operate on only a single atom at a time in a very small box. Nevertheless, the principles of operation are the same.

Indeed, in 1984 physicists at the Max Planck Institute for Quantum Optics in Garching, Germany, succeeded in operating a "micromaser" containing only one atom. To start up the micromaser, Rydberg atoms are sent one at a time through a superconducting cavity. These atoms are prepared in a state whose favored transition matches the resonant frequency of the cavity (between 20 and 70 gigahertz). In the Garching micromaser the atoms all had nearly the same velocity, so they spent the same time inside the cavity.

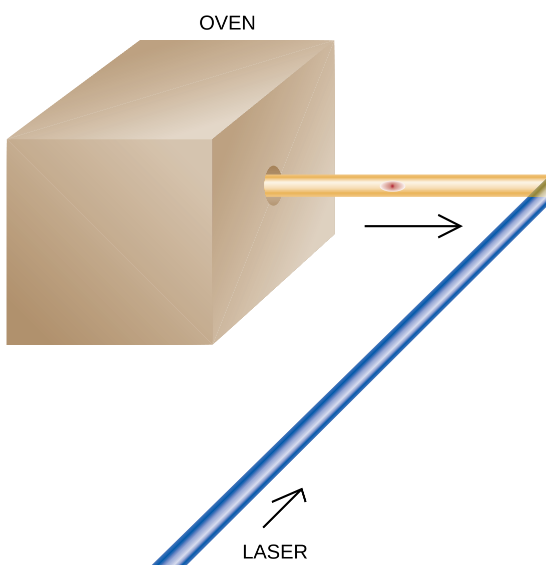
This apparatus is simply another realization of the atom-cavity coupled oscillator; if an atom were to remain inside the cavity indefinitely, it would exchange a photon with the cavity at some characteristic rate. Instead, depending on the atom's speed, there is some fixed chance that an atom will exit unchanged and a complementary chance that it will leave a photon behind.

If the cavity remains empty after the first atom, the next one faces an identical chance of exiting the cavity in the same state in which it entered. Eventually, however, an atom deposits a photon; then the next atom in line encounters sharply altered odds that it will emit energy. The rate at which atom and field exchange energy depends on the number of photons already present—the more photons, the faster the atom is stimulated to exchange additional energy with the field. Soon the cavity contains two photons, modifying the odds for subsequent emission even further, then three and so on at a rate that depends at each step on the number of previously deposited photons.

In fact, of course, the photon number does not increase without limit as atoms keep crossing the resonator. Because the walls are not perfect reflectors, the more photons there are, the greater becomes the chance that one of them will be absorbed. Eventually this loss catches up to the gain caused by atomic injection.

About 100,000 atoms per second can pass through a typical micromaser (each remaining perhaps 10 microseconds); meanwhile the photon lifetime within the cavity is typically about 10 milliseconds. Consequently, such a device running in steady state contains about 1,000 microwave photons. Each of them carries an energy of about 0.0001 electron volt; thus, the total radiation stored in the cavity does not exceed one tenth of one electron volt. This amount is much smaller than the electronic excitation energy stored in a single Rydberg atom, which is on the order of four electron volts.

Although it would be difficult to measure such a tiny field directly, the atoms passing through the resonator provide a very simple, elegant way to monitor the maser. The transition rate from one Rydberg state to the other depends on the photon number in the cavity, and experimenters need only measure the fraction of atoms leaving the maser in each state. The populations of the two levels can be determined by ionizing the atoms in two small detectors, each consisting of plates with an electric field across them. The first detector operates at a low field to ionize atoms in the higher-energy state; the second operates at a slightly higher field to



ionize atoms in the lower-lying state (those that have left a photon behind in the cavity).

With its tiny radiation output and its drastic operational requirements, the micromaser is certainly not a machine that could be taken off a shelf and switched on by pushing a knob. It is nevertheless an ideal system to illustrate and test some of the principles of quantum physics. The buildup of photons in the cavity, for example, is a probabilistic quantum phenomenon—each atom in effect rolls a die to determine whether it will emit a photon—and measurements of micromaser operation match theoretical predictions.

An intriguing variation of the micromaser is the two-photon maser source. Such a device was operated for the first time five years ago by our group at ENS. Atoms pass through a cavity tuned to half the frequency of a transition between two Rydberg levels. Under the influence of the cavity radiation, each atom is stimulated to emit a pair of identical photons, each bringing half the energy required for the atomic transition. The maser field builds up as a result of the emission of successive photon pairs.

The presence of an intermediate energy level near the midpoint between the initial and the final levels of the transition helps the two-photon process along. Loosely speaking, an atom goes from its initial level to its final one via a "virtual" transition during which it jumps down to the middle level while emitting the first photon; it then jumps down again while emitting the second photon. The intermediate step is virtual because the energy of the emitted photons, whose frequency is set by the cavity, does not match the energy differences between the intermediate level and either of its neighbors. How can such a paradoxical situation exist? The Heisenberg uncertainty principle permits the atom briefly to borrow enough energy to emit a photon whose energy exceeds the difference between the top level and the middle one, provided that this loan is paid back during the emission of the second photon.

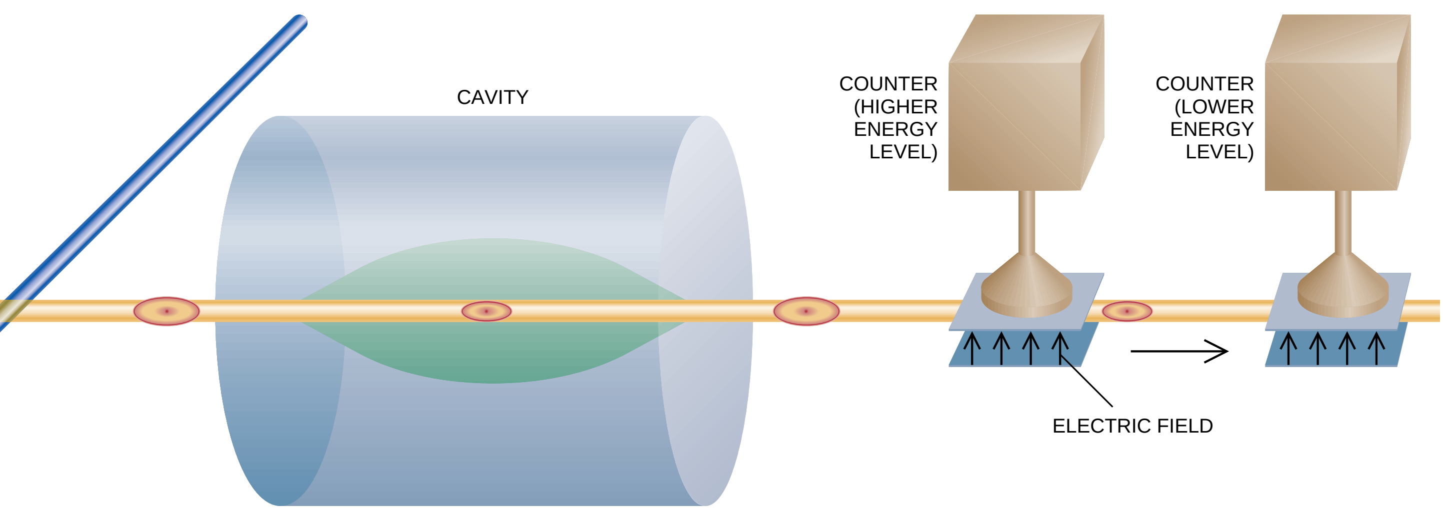
Like all such quantum transactions, the term of the energy loan is very short. Its maximum duration is inversely proportional to the amount of borrowed energy. For a mismatch of a few billionths of an electron volt, the loan typically lasts a few nanoseconds. Because larger loans are increasingly unlikely, the probability of the two-photon process is inversely proportional to this mismatch.

The micromaser cavity makes two-photon operation possible in two ways. It inhibits single-photon transitions that are not resonant with the cavity, and it strongly enhances the emission of photon pairs. Without the cavity, Rydberg atoms in the upper level would radiate a single photon and jump down to the intermediate level. This process would deplete the upper level before two-photon emission could build up.

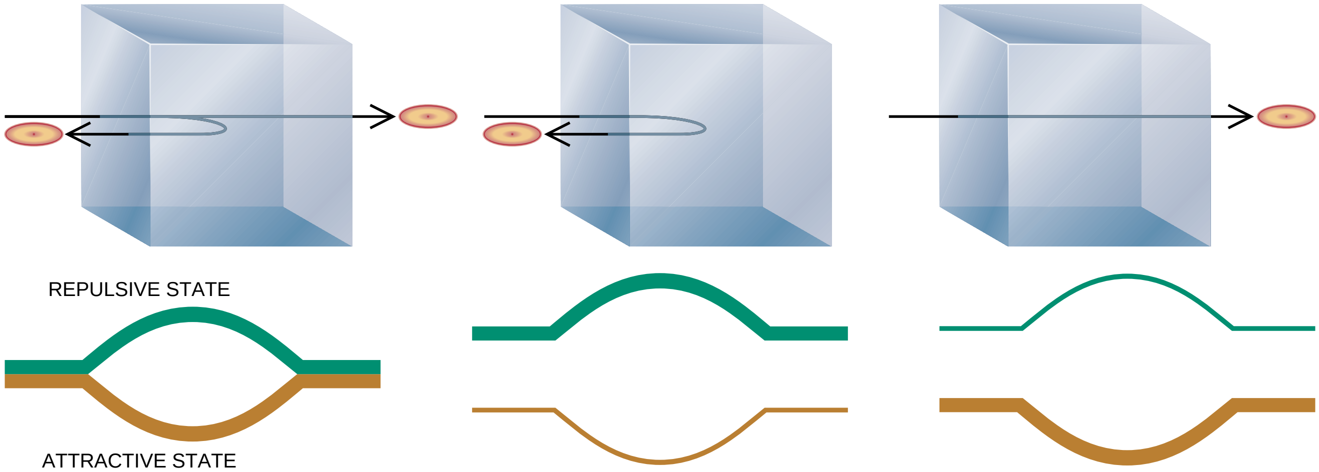
Although the basic principle of a two-photon micromaser is the same as that of its simple one-photon cousin, the way in which it starts up and operates differs significantly. A strong fluctuation, corresponding to the unlikely emission of several photon pairs in close succession, is required to trigger the system; as a result, the field builds up only after a period of "lethargy." Once this fluctuation has occurred, the field in the cavity is relatively strong and stimulates emission by subsequent atoms, causing the device to reach full power (about 10^{-18} watt) rapidly. A two-photon laser system recently developed by a group at Oregon State University operates along a different scheme but displays essentially the same metastable behavior.

The success of micromasers and other similar devices has prompted cavity QED researchers to conceive new experiments, some of which would have been dismissed as pure science fiction only a few years ago. Perhaps the most remarkable of these as yet hypothetical experiments are those that deal with the forces experienced by an atom in a cavity containing only a vacuum or a small field made of a few photons.

The first thought experiment starts with a single atom and an empty cavity tuned to a transition between two of the atom's states. This coupled-oscillator system has two nonstationary states: one corresponds to an excited atom in an empty cavity, the other to a de-excited atom with one photon. The system also has two stationary states, obtained by addition or subtraction of the nonstationary ones—addition of the nonstationary states corresponds to the in-phase oscillation mode of the two-pendulum model, and subtraction of the states corresponds to the



MICROMASER uses an atomic beam and a superconducting cavity to produce coherent microwave radiation. A laser beam (*left*) strikes atoms coming out of an oven and excites them into high-energy Rydberg states. The atoms pass one at a time through a cavity tuned to the frequency of a transition to a lower-energy state; the field builds up as successive atoms interact with the cavity and deposit photons in it. The micromaser field can be inferred from the readings of counters that monitor the number of atoms leaving the cavity in either the higher- or lower-energy state.



EMPTY CAVITY can repel or attract slow-moving, excited atoms. The strength of the coupling between an atom and a tuned cavity typically vanishes at the walls and reaches a maximum in the center. (Curves at the bottom show the energy of the atom-cavity system as a function of the atom's position within the cavity.) The change in energy results in a force on atoms moving through the cavity. If the cavity wavelength matches the atomic transition exactly, this force can be either attractive or repulsive (*left*). If the atomic transition has a slightly higher frequency than the resonant frequency of the cavity, the force will be repulsive (*center*); if the transition has a lower frequency, the force will be attractive (*right*).

out-of-phase mode. These stationary states differ in energy by a factor equal to Planck's constant, *h,* times the exchange frequency between the atom and the cavity.

This exchange frequency is proportional to the amplitude of the cavity's resonant vacuum field. Typically this field vanishes at the walls and near the ports by which the atom enters and leaves the cavity. It reaches a maximum at the cavity center. As a result, the atom-cavity coupling (and thus the energy difference between the system's two stationary states) is zero when the atom enters and leaves the cavity and goes to a maximum when the atom reaches the middle of the cavity.

The fundamental laws of mechanics say, however, that for a change in the relative position of two objects to lead to a change in energy, a force must be exerted between these objects. In other words, the atom experiences a push or a pull, albeit an infinitesimal one, as it moves through the empty cavity. If the system is prepared in the higher-energy state, its energy reaches a maximum at the center—the atom is repelled. If the system is in the lower-energy state, the interaction attracts the atom to the cavity center. These forces have been predicted independently by our group and by a group at Garching and the University of New Mexico.

For Rydberg atoms in a microwave cavity with a typical exchange frequency of 100 kilohertz, the potential energy difference is about one ten-billionth of an electron volt. This corresponds to a temperature of a few microkelvins and to the kinetic energy of an atom moving with a velocity of a few centimeters per second. If the speed of the incoming atom is less than this critical value, the potential barrier caused by the atom-cavity interaction will reflect the atom back, or, conversely, the potential well will be deep enough to trap it near the cavity center. Atoms in such slow motion can now be produced by laser cooling [see "Laser Trapping of Neutral Particles," by Steven Chu; SCIENTIFIC AMERICAN, February 1992]; these tiny forces may yet be observed.

If a very slow moving, excited atom is sent into a resonant, empty cavity, these forces result in a kind of atomic beam splitter. The nonstationary initial state of the system consists of the sum of the repelling and attractive states—a superposition of the two stationary atom-cavity wave functions. Half corresponds to an atom reflected back at the cavity entrance, and the other half corresponds to an atom passing through; either outcome occurs with equal probability.

To prepare a pure attractive or repelling state, one should detune the cavity slightly from the atomic transition. When the transition is a bit more energetic than the photon that the cavity can sustain, the state with an excited atom and no photon has a little more energy than the one with a de-excited atom and one photon. When the atom enters the cavity, the exchange coupling works to separate the two states, so that the state with an excited atom and no photon branches unambiguously into the higher-energy steady state, in which the atom is repelled. The same trick just as easily makes an attractive state if the cavity photon energy is slightly higher than the atomic transition.

This evolution of the atom-cavity system relies on the so-called adiabatic theorem, which says that if a quantum system's rate of change is slow enough, the system will continuously follow the state it is initially prepared in, provided the energy of that state does not coincide at any time with that of another state. This adiabaticity criterion is certainly met for the very slow atoms considered here.

These atom-cavity forces persist as long as the atom remains in its Rydberg state and the photon is not absorbed by the cavity walls. This state of affairs can typically last up to a fraction of a second, long enough for the atom to travel through the centimeter-size cavity.

The forces between atom and cavity are strange and ghostly indeed. The cavity is initially empty, and so in some way the force comes from the vacuum field, which suggests that it is obtained for nothing. Of course, that is not strictly true, because if the cavity is empty, the atom has to be initially excited, and some price is paid after all.

The force can also be attributed to the exchange of a photon between the atom and the cavity. Such a view is analogous to the way that electric forces between two charged particles are ascribed to the exchange of photons or the forces between two atoms in a molecule to the exchange of electrons.

Another interpretation of the atom-cavity vacuum attraction and repulsion, based on a microscopic analysis, shows that these phenomena are in fact not essentially different from the electrostatic forces whose demonstration was a society game in the 18th-century French court. If one charges a needle and brings small pieces of paper into its vicinity, the pieces stick to the metal. The strong electric field at the tip polarizes the pieces, pulling their electrons onto one side and leaving a net positive charge on the other, essentially making small electric dipoles. The attraction between the needle and the charges on the near side of the paper exceeds the repulsion between the needle and those on the far side, creating a net attractive force.

The atom and the cavity contain the same ingredients, albeit at a quantum level. The vacuum field bounded by the cavity walls polarizes the Rydberg atom, and the spatial variations of the field produce a net force. The atomic dipole and the vacuum field are oscillating quantities, however, and their respective oscillations must maintain a constant relative phase if a net force is to continue for any length of time. As it turns out, the photon exchange process does in fact lock the atomic dipole and the vacuum fluctuations.

The tiny force experienced by the atom is enhanced by adding photons to the cavity. The atom-cavity exchange frequency increases with the field intensity, so that each photon adds a discrete quantum of height to the potential barrier in the repelling state and a discrete quantum of depth to the potential well in the attractive state. As a result, it should be possible to infer the number of photons inside the cavity by measuring the time an atom with a known velocity takes to cross it or, equivalently, by detecting the atom's position downstream of the cavity at a given time.

One could inject perhaps a dozen or so photons into a cavity and then launch through it, one by one, Rydberg atoms whose velocity is fixed at about a meter per second. The kinetic energy of these atoms would be greater than the atom-cavity potential energy, and they would pass through the cavity after experiencing a slight positive or negative delay, depending on the sign of the atom-cavity detuning. To detect the atom's position after it has passed through the cavity, researchers could fire an array of field ionization detectors simultaneously some time after the launch of each atom. A spatial resolution of a few microns should be good enough to count the number of photons in the cavity.

Before measurement, of course, the photon number is not merely a classically unknown quantity. It also usually contains an inherent quantum uncertainty. The cavity generally contains a field whose description is a quantum wave function assigning a complex amplitude to each possible number of photons. The probability that the cavity stores a given number of photons is the squared modulus of the corresponding complex amplitude.

The laws of quantum mechanics say that the firing of the detector that registers an atom's position after it has crossed the cavity collapses the ambiguous photon-number wave function to a single value. Any subsequent atom used to measure this number will register the same value. If the experiment is repeated from scratch many times, with the same initial field in the cavity, the statistical distribution of photons will be revealed by the ensemble of individual measurements. In any given run, however, the photon number will remain constant, once pinned down.

This method for measuring the number of photons in the cavity realizes the remarkable feat of observation known as quantum nondemolition. Not only does the technique determine perfectly the number of photons in the cavity, but it also leaves that number unchanged for further readings.

Although this characteristic seems to be merely what one would ask of any measurement, it is impossible to attain by conventional means. The ordinary way to measure this field is to couple the cavity to some kind of photodetector, transforming the photons into electrons and counting them. The absorption of photons is also a quantum event, ruled by chance; thus, the detector adds its own noise to the measured intensity. Furthermore, each measurement requires absorbing photons; thus, the field irreversibly loses energy. Repeating such a procedure therefore results in a different, lower reading each time. In the nondemolition experiment, in contrast, the slightly nonresonant atoms interact with the cavity field without permanently exchanging energy.

Quantum optics groups around the world have discussed various versions of quantum nondemolition experiments for several years, and recently they have begun reducing theory to practice. Direct measurement of an atom's delay is conceptually simple but not very sensitive. More promising variants are based on interference effects involving atoms passing through the cavity—like photons, atoms can behave like waves. They can even interfere with themselves. The so-called de Broglie wavelength of an atom is inversely proportional to velocity; a rubidium atom traveling 100 meters per second, for example, has a wavelength of 0.45 angstrom.

If an atom is slowed while traversing the cavity, its phase will be shifted by an angle proportional to the delay. A delay that holds an atom back by a mere 0.22 angstrom, or one half of a de Broglie wavelength, will replace a crest of the matter wave by a trough. This shift can readily be detected by atomic interferometry.

If one prepares the atom itself in a superposition of two states, one of which is delayed by the cavity while the other is unaffected, then the atomic wave packet itself will be split into two parts. As these two parts interfere with each other, the resulting signal yields a measurement of the phase shift of the matter wave and hence of the photon number in the cavity. Precisely this experiment is now under way at our laboratory in Paris, using Rydberg atoms that are coupled to a superconducting cavity in an apparatus known as a Ramsey interferometer.

Such an apparatus has many potential uses. Because the passing atoms can monitor the number of photons in a cavity without perturbing it, one can witness the natural death of photons in real time. If a photon disappears in the cavity walls, that disappearance would register immediately in the atomic interference pattern. Such experiments should provide more tests of quantum theory and may open the way to a new generation of sensors in the optical and microwave domains.

FURTHER READING

RADIATIVE PROPERTIES OF RYDBERG STATES IN RESONANT CAVITIES. S. Haroche and J.-M. Raimond in *Advances in Atomic and Molecular Physics,* Vol. 20, pages 350–409; 1985.

THE SINGLE ATOM MASER AND THE QUANTUM ELECTRODYNAMICS IN A CAVITY. H. Walther in *Physica Scripta,* Vol. T23, pages 165–169; 1988.

CAVITY QUANTUM ELECTRODYNAMICS. S. Haroche and D. Kleppner in *Physics Today,* Vol. 42, No. 1, pages 24–30; January 1989.

CAVITY QUANTUM ELECTRODYNAMICS. E. A. Hinds in *Advances in Atomic, Molecular, and Optical Physics,* Vol. 28, pages 237–289; 1991.

CAVITY QUANTUM OPTICS AND THE QUANTUM MEASUREMENT PROCESS. P. Meystre in *Progress in Optics*, Vol. 30. Edited by E. Wolf. Elsevier Science Publishers, 1992.

CAVITY QUANTUM ELECTRODYNAMICS. S. Haroche in *Fundamental Systems in Quantum Optics*. Proceedings of Les Houches Summer School, Session LIII. Edited by J. Dalibard, J.-M. Raimond and J. Zinn-Justin. North-Holland, 1992.

Listening with Two Ears

Studies of barn owls offer insight into just how the brain combines acoustic signals from two sides of the head into a single spatial perception

by Masakazu Konishi

Why do people have two ears? We can, after all, make sense of sounds quite well with a single ear. One task, however, requires input from both organs: pinpointing the exact direction from which a sound, such as the cry of a baby or the growl of a dog, is emanating. In a process called binaural fusion, the brain compares information received from each ear and then translates the differences into a unified perception of a single sound issuing from a specific region of space.

Extensive research has shown that the spatial cues extracted by the human brain are differences in the arrival time and the intensity, or force, of sound waves reaching the ears from a given spot. Differences arise because of the distance between the ears. When a sound comes from a point directly in front of us, the waves reach both ears at the same time and exert equal force on the receptive surfaces that relay information to the brain. But if a sound emanates from, say, left of center, the waves will reach the right ear slightly after the left. They will also be somewhat less intense at the right because, as they travel to the far ear, some fraction of the waves will be absorbed or deflected by the head.

The brain's use of disparities in timing and intensity becomes especially obvious when tones are delivered separately to each ear through a headset. Instead of perceiving two distinct signals, we hear one signal—a phantom—originating from somewhere inside or outside the head. If the stimuli fed to the ears are equally intense (equally loud) and are conveyed simultaneously, we perceive one sound arising from the middle of the head. If the volume is lowered in just one ear or if delivery to that ear is delayed, the source seems to move in the direction of the opposite ear.

This much has long been known. What is less clear is how the brain manages to detect variances in timing and intensity and how it combines the resulting information into a unified spatial perception. My colleagues and I at the California Institute of Technology have been exploring this question for more than 15 years by studying the behavior and brain of the barn owl (*Tyto alba*). Recently we have uncovered almost every step of the computational process in these animals. (The only other sensory system that is as completely defined belongs to a fish.) We find that the owl brain combines aural signals relating to location not all at once but through an amazing series of steps. Information about timing and intensity is processed separately in parallel pathways that converge only late in those pathways. It is highly probable that humans and other mammals achieve binaural fusion in much the same manner.

MASAKAZU KONISHI has been Bing Professor of Behavioral Biology at the California Institute of Technology since 1980. He earned a doctorate in zoology from the University of California, Berkeley, in 1963. Three years later he joined the faculty of Princeton University, where he studied hearing and vocalization in songbirds as well as sound localization in owls. Konishi moved to Caltech as a professor in 1975. In his free time, he enjoys hiking, skiing and visiting England to attend sheep-dog competitions.

I first thought of examining the neural basis of sound location in owls in 1963, when I heard Roger S. Payne, now at the Whale Conservation Institute in Lincoln, Mass., report that the barn owl can catch a mouse readily in darkness, solely by relying on acoustic cues. I had recently earned a doctorate in zoology and wanted to know more about how animals identify the position of a sound source, but I had yet to choose a species to study. Three years later, at Princeton University, I observed the exquisite aural abilities of barn owls for myself after I obtained three of them from a bird-watcher. When I watched one of the owls through an infrared-sensitive video camera in a totally dark room, I was impressed by the speed and accuracy with which it turned its head toward a noise. I concluded that the head-turning response might help uncover whether such animals use binaural fusion in locating sound. If they did, studies of their brain could help elucidate how such fusion is accomplished.

As I had anticipated, the head-turning response did prove extremely useful to me and my postdoctoral fellows, particularly after I established a laboratory at Caltech in 1975. In some of our earliest research there, Eric I. Knudsen, now at Stanford University, and I obtained indirect evidence that barn owls, like humans, must merge information from the two ears to locate a sound. When one ear was plugged, the animals turned the head in response to noise from a loudspeaker, but they did not center on the speaker [see "The Hearing of the Barn Owl," by Eric I. Knudsen; SCIENTIFIC AMERICAN, December 1981].

In the early 1980s Andrew Moiseff and I additionally showed that the barn owl extracts directional information from disparities in the timing and the intensity of signals reaching the two ears—technically called interaural time differences and interaural intensity differences. As part of that effort, we measured the differences that arose as we moved a speaker across the surface of an imaginary globe around an owl's head. Microphones we had placed in the ears relayed the signals reaching each ear to a device that measured arrival time and volume. When we eased the speaker from the midline of the face (zero angle) 90 degrees to the left or

BARN OWL PINPOINTS PREY in the dark by listening. It determines the appropriate trajectory in which to fly by comparing differences in the timing and the intensity of sounds reaching its two ears. An infrared strobe flashing five times per second caught this barn owl in action in the author's laboratory.

right, the difference in arrival time at the two ears increased systematically. Those results resembled the findings of human studies.

In contrast to human findings, the difference in intensity did not vary appreciably as the speaker was moved horizontally. But it did increase as the speaker was moved up or down from eye level—at least when the sound included waves of frequencies higher than three kilohertz, or 3,000 cycles per second. Payne, who had seen the same intensity changes in earlier studies, has attributed them, apparently correctly, to an asymmetry in the placement of the owl's ears. The left ear is higher than eye level but points downward, whereas the right ear is lower but points upward. The net result is that the left ear is more sensitive to sounds coming

from below, and the right is more sensitive to sounds from above.

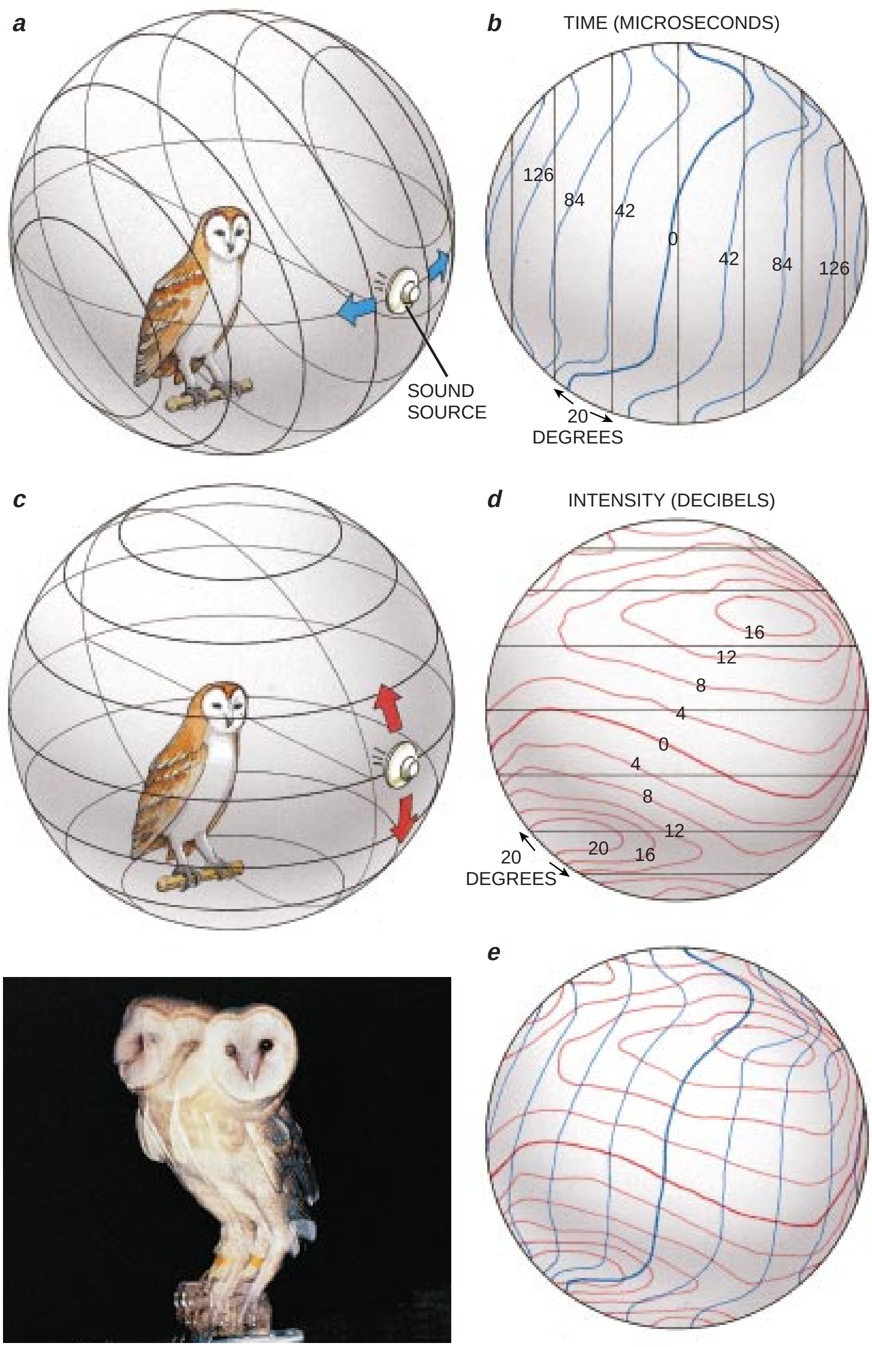
Satisfied that arrival time and intensity often differ for the two ears, we could go on to determine whether the owl actually uses specific combinations of disparities in locating sound sources. We intended to put a standard headset on tame animals and to convey a noise separately to each ear, varying the difference in delivery time or volume, or both. We would then see whether particular combinations of time and intensity differences caused the animals to turn the head reliably in specific directions. Unfortunately, we did not receive cooperation from our subjects. When we tried to affix the earphones, each owl we approached shook its head and backed off. We managed to proceed only after we acquired tiny earphones that could be inserted into the owls' ear canal.

We also had to devise a way to measure the direction of head turning, determining both the horizontal and vertical components of the response to each set of stimuli. We solved the problem mainly by applying the search-coil technique that Gary G. Blasdel, now at Harvard Medical School, had designed a few years earlier. We fit two small coils of copper wire, arranged perpendicularly to each other, on an owl's head. We positioned the owl between two big coils carrying electric current. As the head moved, the large coils induced currents in the small ones. Variations in the flow of current in the smaller coils revealed both the horizontal and vertical angles of the head turning.

Sure enough, the owl responded rapidly to signals from the earphones, just as if it had heard noise arising from outside the head. When the sound in one ear preceded that in the other ear, the head turned in the direction of the leading ear. More precisely, if we held the volume constant but issued the sound to one ear slightly before the other ear, the owl turned its head mostly in the horizontal direction. The longer we delayed delivering the sound to the second ear, the further the head turned.

Similarly, if we varied intensity but held timing constant, the owl tended to move its head up or down. If we issued sounds so that both the delivery time and the intensity of signals to the left ear differed from those of the right, the owl moved its head horizontally and vertically. Indeed, combinations of interaural timing and intensity differences that mimicked the combinations generated from a speaker at particular sites caused the animal to turn toward exactly those same sites. We could therefore be confident that the owl brain does fuse timing and intensity data to determine the horizontal and vertical coordinates of a sound source. The process by which barn owls calculate distance is less clear.

To learn how the brain carries out binaural fusion, we had to examine the brain itself. Our research plan built on work Knudsen and I had completed several years earlier. We had identified cells that are now known to be critical to sound location. Called space-specific neurons, they react only



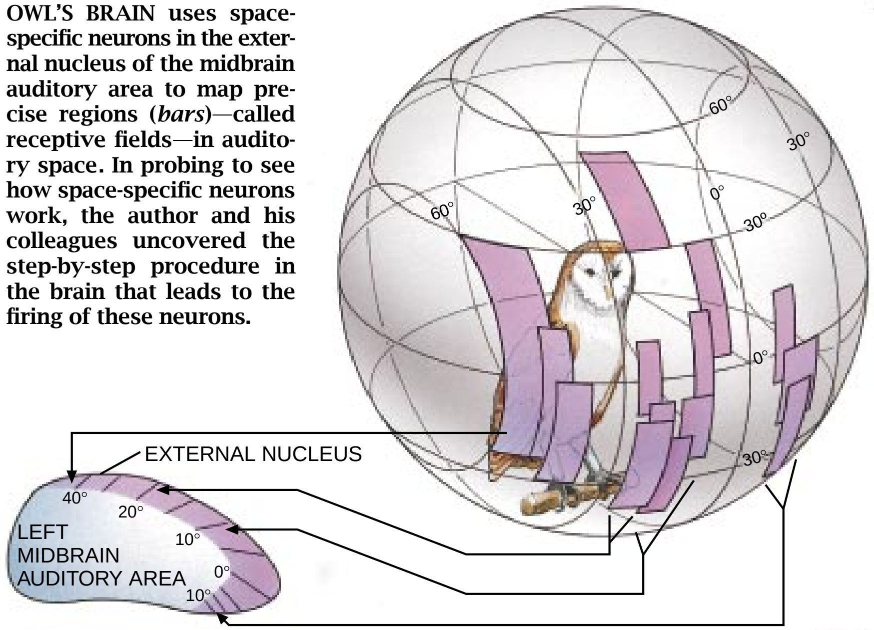
DIFFERENCES IN TIMING AND INTENSITY at which a sound reaches an owl's two ears vary as the source of the sound moves along the surface of an imaginary globe around the owl's head. Differences in timing locate the sound in the horizontal plane (*a*); the difference increases 42 microseconds every 20 degrees a sound source moves (*b*). Differences in intensity locate the sound vertically (*c*). Sound from above eye level is more intense in the right ear, by the decibel levels shown (*d*); from below eye level, it is more intense in the left ear. Differences vary with frequency; they were measured for six kilohertz in this case. Combining the two graphs (*e*) defines each location in space. When an owl is exposed to a particular pair of differences, it quickly turns its head in a predictable direction (*photograph*).

to acoustic stimuli originating from specific receptive fields, or restricted areas in space [*see illustration at right*]. These neurons reside in a region of the brain called the external nucleus, which is situated within the auditory area of the midbrain (the equivalent of the mammalian inferior colliculus). Collectively, the space-specific neurons in the left external nucleus form a map of primarily the right side of auditory space (the broad region in space from which sounds can be detected), and those of the right external nucleus form a map of primarily the left half of auditory space, although there is some overlap.

We identified the space-specific cells by resting a microelectrode, which resembles a sewing needle, on single neurons in the brain of an anesthetized animal. As we held the electrode in place, we maneuvered a speaker across the surface of our imaginary globe around the owl's head. Certain neurons fired impulses only if the noise emanated from a particular receptive field. For instance, in an owl facing forward, one space-specific neuron might respond only if a speaker were placed within a receptive field extending roughly 20 degrees to the left of the owl's line of sight and some 15 degrees above or below it. A different neuron would fire when the speaker was transferred elsewhere on the globe.

But how did these neurons obtain directional information? Did they process the relevant cues themselves? Or were the cues extracted and combined to some extent at one or more lower way stations (relay centers) in the brain [*see illustration on page 72*], after which the results were simply fed upward?

Moiseff and I intended to answer these questions by carrying out experiments in which we would deliver sounds through earphones. But first we had to be certain that signals able to excite particular space-specific neurons truly mimicked the interaural time and intensity differences that caused the neurons to fire under more natural conditions—namely, when a sound emanated from a spot in the neuron's receptive field. A series of tests gave us the encouragement we needed. In these studies, we issued sounds through the earphones and monitored the response of individual neurons by again holding a microelectrode on or near the cells. As we hoped, we found that cells responded to specific combinations of signals. Further, the sets of timing and intensity differences that triggered strong firing by space-specific neurons corresponded exactly to the combinations that caused an owl to turn its head toward a spot in the neuron's receptive field. This congruence affirmed that our proposed approach was sensible.



OWL'S BRAIN uses space-specific neurons in the external nucleus of the midbrain auditory area to map precise regions (*bars*)—called receptive fields—in auditory space. In probing to see how space-specific neurons work, the author and his colleagues uncovered the step-by-step procedure in the brain that leads to the firing of these neurons.

In our initial efforts to trace the steps by which the brain circuitry accomplishes binaural fusion, Moiseff and I tried to find neurons sensitive to interaural timing or intensity differences in the way stations that relay signals from the auditory nerve up to the midbrain. These preliminary investigations, completed in 1983, suggested that certain stations are sensitive only to timing cues, whereas others are sensitive solely to intensity cues. The brain, it seemed, functioned like a parallel computer, processing information about timing and intensity through separate circuits.

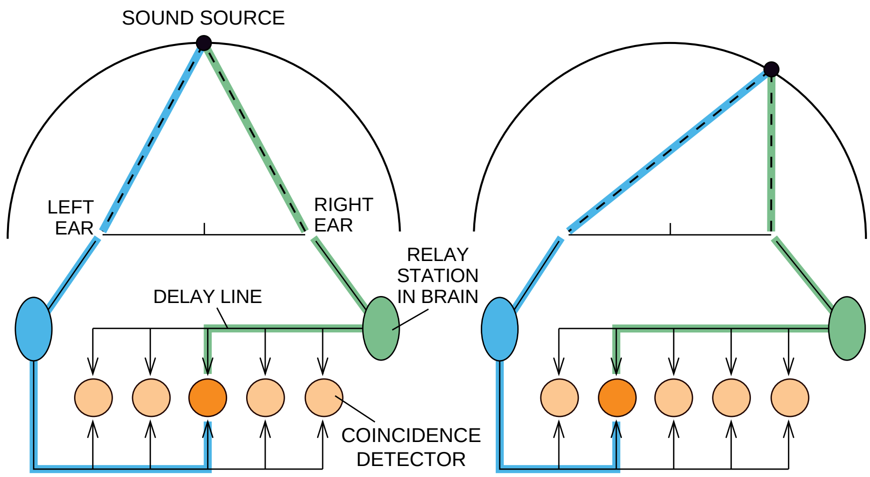
Such clues led us to seek further evidence of parallel processing. Joined by Terry T. Takahashi, now at the University of Oregon, we began by examining the functioning of the lowest way stations in the brain—the cochlear nuclei. Each cerebral hemisphere has two: the magnocellular nucleus and the angular nucleus. In owls, as in other birds, each fiber of the auditory nerve—that is, each signal-conveying axon projecting from a neuron in the ear—divides into two branches after leaving the ear. One branch enters the magnocellular nucleus; the other enters the angular nucleus.

We wondered how the space-specific neurons would behave if we prevented nerve cells from firing in one of the two cochlear nuclei. We therefore injected a minute amount of a local anesthetic into either the magnocellular or angular nucleus. The results were dramatic: the drug in the magnocellular nucleus altered the response of space-specific neurons to interaural time differences without affecting the response to intensity differences. The converse occurred when the angular nucleus received the drug. Evidently, timing and intensity are indeed processed separately, at least at the lowest way stations of the brain; the magnocellular neurons convey timing data, and the angular neurons convey intensity data.

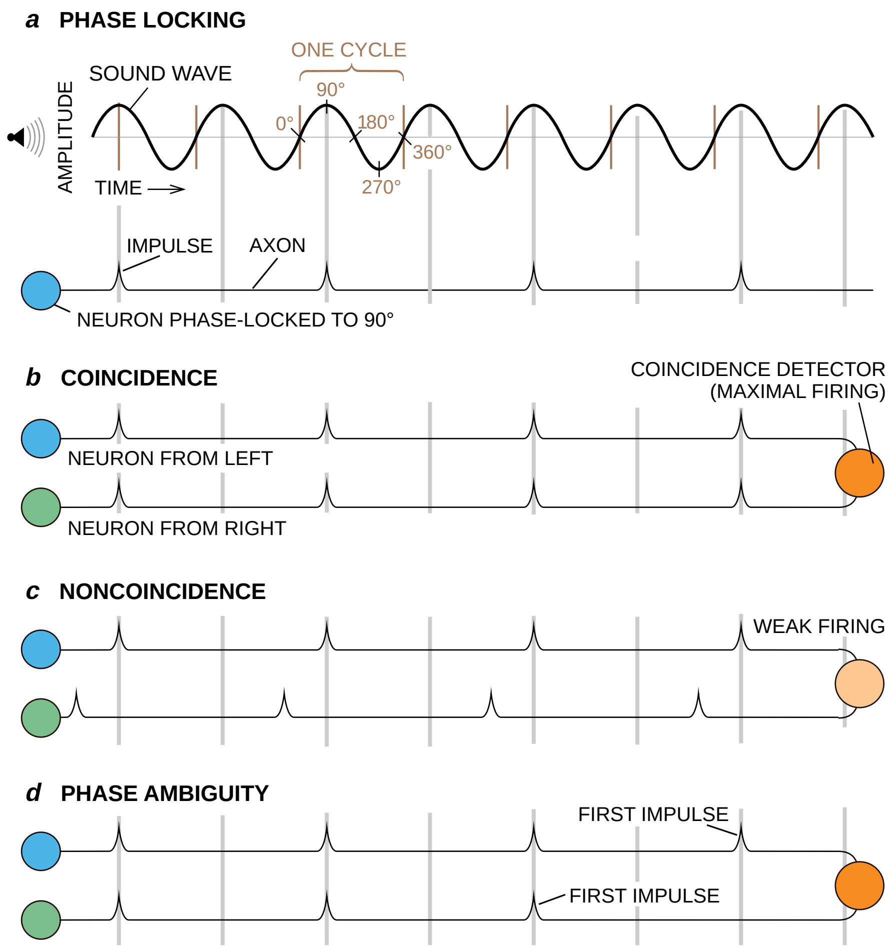
These exciting results spurred me to ask Takahashi to map the trajectories of the neurons that connect way stations in the auditory system. His work eventually revealed that two separate pathways extend from the cochlear nuclei to the midbrain. The anatomic evidence, then, added further support to the parallel-processing model.

While Takahashi was conducting his mapping research, W. E. Sullivan and I explored the ways magnocellular and angular nuclei extract timing and intensity information from signals arriving from the auditory nerve. To understand our discoveries, one needs to be aware that most sounds in nature are made up of several waves, each having a different frequency. When the waves reach a receptive surface in the ear, known as the basilar membrane, the membrane begins to vibrate, but not uniformly. Different parts of the membrane vibrate maximally in response to particular frequencies. In turn, neurons that are connected to the maximally vibrating areas (and thus are "tuned" to specific frequencies) become excited. These neurons propagate impulses along the auditory nerve to the brain.

We and others find that the intensity of a sound wave of a given frequency is conveyed to the brain from the ear by the firing rate of auditory neurons tuned to that frequency. This much makes



MODEL CIRCUIT for detection of interaural time differences was suggested in 1948. The coincidence detectors receive inputs from both ears. They fire only when impulses from the two sides arrive simultaneously through fibers that serve as delay lines. The detector that responds (*darkly colored circle*) changes as a sound source moves from directly in front of an individual (*left*) to the side (*right*). The owl brain operates in much the way the model proposed.



SOUND WAVE OF A SINGLE FREQUENCY causes neurons sensitive to it to fire trains of impulses at a particular phase angle (*a*). Coincidence detectors in the owl's brain fire most strongly when impulses generated at the same phase angle reach the detectors simultaneously (*far right in b*). Detectors can also fire, but more weakly, when impulse trains reaching them are slightly asynchronous (*c*). In what is called phase ambiguity, peak firing can occur if a sound to one ear is delayed or advanced by a full cycle from another delivery time that yields coincidence (*d*).

intuitive sense. Our next result is less obvious. Neurons of the auditory nerve also exhibit what is called phase locking: they fire at characteristic points, or phase angles, along the sound wave [*see bottom illustration on this page*]. That is, a neuron tuned to one frequency will tend to fire, for example, when the wave is at baseline (zero degrees), although it does not necessarily fire every time the wave reaches that position. A neuron tuned to a different frequency will tend to fire at a different phase angle, such as when a wave is cresting (at the point called 90 degrees, which is a quarter of the way through a full 360-degree wave cycle), or reaches some other specific point. In both ears, impulses produced by neurons tuned to the same frequency will lock to the same phase angle. But, depending on when the signals reach the ears, the train of impulses generated in one ear may be delayed relative to the impulse train generated in the opposite ear.

It turns out that cells of the magnocellular nucleus exhibit phase locking. But they are insensitive to intensity; changes in the volume of a tone do not affect the rate of firing. In contrast, few angular neurons show phase locking, but they respond distinctly to changes in intensity. These and other results indicate that the owl depends on trains of phase-locked impulses relayed from the magnocellular nucleus for measuring interaural time differences, and the animal relies on the rate of impulses fired by the angular nucleus for gauging interaural intensity differences. Overall, then, our analyses of the lowest way stations of the brain established that the cochlear nuclei serve as filters that pass along information about timing or intensity, but not both.

We then proceeded to explore higher regions, pursuing how the brain handles timing data in particular. Other studies, which will be discussed, addressed intensity. We learned that when phase-locked impulses induced by sound waves of a single frequency (a pure tone) leave the magnocellular nucleus on each side of the brain, they travel to a second way station: the laminar nucleus. Impulses from each ear are transmitted to the nucleus on both the opposite and the same side of the head. The laminar nucleus is, therefore, the first place where the information from both ears comes together in one place.

The general problem of how the brain combines timing data has been a subject of speculation for decades. The late Lloyd A. Jeffress put forth a reasonable model in 1948, while spending a sabbatical leave at Caltech. Jeffress proposed that the nerve fibers carrying time-related signals from the ears (called delay lines) vary in how rapidly they deliver signals to way stations in the brain. They ultimately converge at neurons (known as coincidence detectors) that fire only when impulses from the two sides arrive simultaneously.

Signals reaching the ears at different times would attain coincidence—arrive at coincidence detectors in unison—if the sum of a sound wave's transit time to an ear and the travel time of impulses emanating from that ear to a coincidence detector were equal for the two sides of the head. Consider a sound that reached the left ear five microseconds before it reached the right ear. Impulses from the two ears would meet simultaneously at a coincidence detector in, say, the right hemisphere if the delay lines from the left ear (the near ear) prolonged the transit time of impulses from that ear to a coincidence detector by five microseconds over the time it would take impulses to traverse fibers from the right ear [*see top illustration on opposite page*].

Since 1948, physiological studies examining neuronal firing in dogs and cats and anatomic studies of chicken brains have suggested that the brain does in fact measure interaural time differences by means of delay lines and coincidence detection. In 1986 Catherine E. Carr, now at the University of Maryland, and I demonstrated in the barn owl that nerve fibers from magnocellular neurons serve as delay lines and neurons of the laminar nucleus serve as coincidence detectors.

But the owl's detection circuit, like those of mammals that have been examined, differs somewhat from the Jeffress model. Neurons of the laminar nucleus respond most strongly to coincidence brought about by particular time differences. Yet they also respond, albeit less strongly, to signals that miss perfect coincidence. The number of impulses declines gradually as the interaural time difference increases or decreases from the value that produces coincidence—that is, until the waves reaching one ear are 180 degrees (a full half cycle) out of phase from the position that would bring about coincidence. At that point, firing virtually ceases. (The neurons also respond, at an intermediate level, to signals delivered to just one ear.)

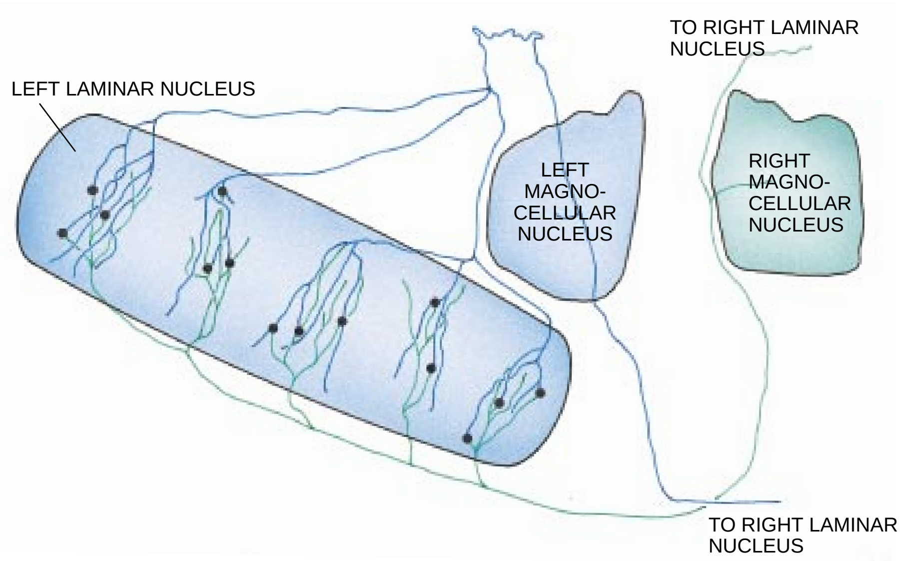
In a way, then, coincidence detectors, by virtue of the delay lines feeding them, can be said to be maximally sensitive to specific time differences. They are not, however, totally selective as to when they produce a peak response. They can be induced to fire with rising strength as the phase difference increases beyond 180 degrees from the value that produces coincidence. When the displacement reaches a full 360 degrees, the arrival time of sound waves at one ear is delayed by the time it takes for a sound wave to complete a full cycle. In that situation, and at every 360-degree difference, coincidence detectors will repeatedly be hit by a series of synchronous impulses and will fire maximally. Thus, the same cell can react to more than one time difference.

Fortunately for the owl, some mechanism resolves such "phase ambiguity" at higher stages, thereby preventing confusion. How this resolution is achieved remains obscure. Another mystery engages us as well: the owl can detect interaural time differences as short as 10 microseconds (10 millionths of a second). Yet a single impulse persists considerably longer than that, on the order of 1,000 microseconds. We are seeking an explanation for this apparent paradox.

The rest of the pathway for time detection is more straightforward. After a coincidence detector in the laminar nucleus on one side of the brain determines the interaural time difference produced by a sound of a given frequency, it simply passes the result upward to higher stations, including to the core region of the midbrain auditory area on the opposite side of the head. Consequently, the higher areas inherit from the laminar nucleus not only selectivity for frequency and interaural time differences but also phase ambiguity. The information in the core, in turn, is passed to a surrounding area—known as the shell of the midbrain auditory area—on the reverse side of the brain, where it is finally combined with information about intensity.

My colleagues and I understand less about the operation of the intensity pathway that converges with the time pathway in the shell. But we have made good progress. Unlike the magnocellular nucleus, which projects up only one stage, to the laminar nucleus, the intensity-detecting angular nucleus projects directly to many higher stations (except the external nucleus). Among them is the posterior lateral lemniscal nucleus.

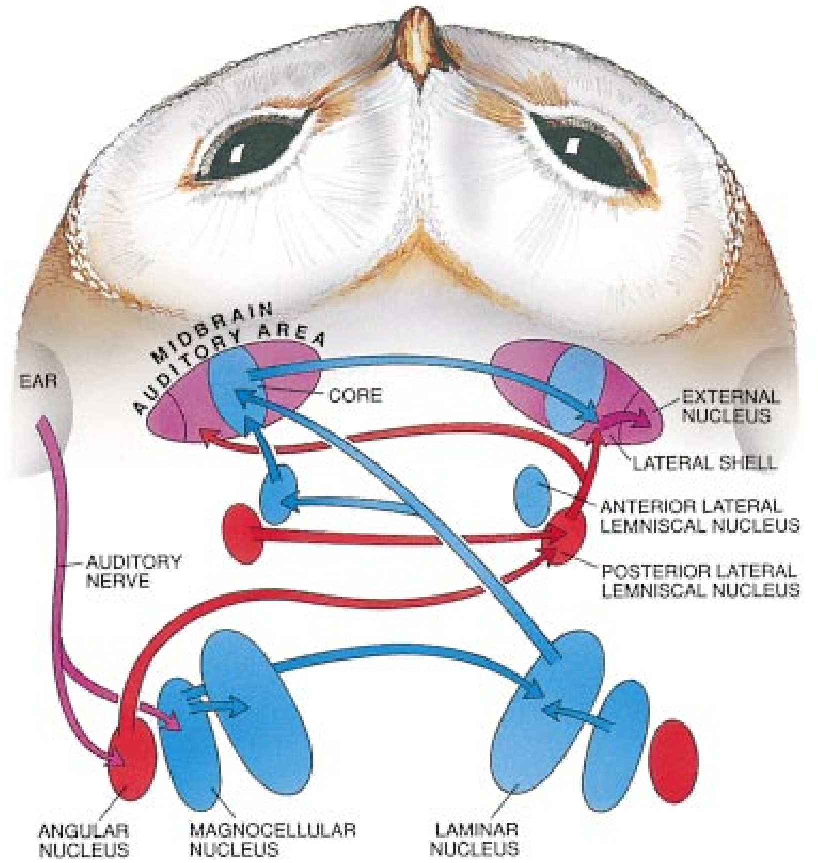
The posterior lemniscal nucleus on one side of the head receives direct input only from the angular nucleus on the opposite side. It nonetheless manages to discern intensity differences between the two ears. Indeed, it is the lowest station in the brain to do so. The lemniscal area can detect such differences because its twin (in the opposite cerebral hemisphere) sends it information from the other angular nucleus. In essence, neurons of the lemniscal nucleus on one side of the head receive excitatory signals from the ear on the opposite side, and they receive inhibitory signals from the ear on the same side. The balance between excitatory and in-



FIBERS FROM THE MAGNOCELLULAR NUCLEUS serve as delay lines, and neurons in the laminar nucleus act as coincidence detectors in the owl's brain. When impulses traveling through the left (*blue*) and right *(green)* fibers reach laminar neurons (*black dots*) simultaneously, the neurons fire strongly.

The Auditory Circuit

Parallel pathways in the barn owl's brain separately process the timing (*blue*) and the intensity (*red*) of sounds reaching the ears (*diagram and flow chart*). The simplified diagram depicts the pathways only for the left ear except where input from the right ear joins that pathway; brain structures are not drawn to scale. Processing begins as the magnocellular nucleus separates out information about time and as the angular nucleus extracts information about intensity from signals delivered by the auditory nerve. The time pathway goes to the laminar nucleus, which receives input from both the right and the left magnocellular nuclei. Neurons of the laminar nucleus are connected to two higher stations: the anterior lateral lemniscal nucleus and the core of the midbrain auditory area. Meanwhile information about intensity travels from the angular nucleus to the posterior lateral lemniscal nucleus, where information from the two ears comes together. The time and intensity pathways finally join in the lateral shell of the midbrain auditory area. They project from there to the external nucleus, which houses the space-specific neurons and is the final station in processing the acoustic cues for locating a sound. If viewed in terms of an algorithm (*far right*), a set of step-by-step procedures for solving a problem, these neurons are at the top of the hierarchy: they represent the final results of all computations that take place in the network.



hibitory signals determines the rate at which the lemniscal neurons fire.

We have observed, too, that neurons of the posterior lemniscal nucleus vary systematically in the intensity differences that cause them to fire most strongly. Geoffrey A. Manley and Christine Köppl of the Technical University of Munich showed in my laboratory that neurons at the bottom of the left nucleus respond maximally when sound is much louder in the left ear and that those at the top of the nucleus fire most strongly when sound is louder in the right ear. Similarly, neurons at the bottom of the right posterior nucleus respond most strongly when sound is much louder in the right ear, and those at the top of the nucleus prefer louder sound in the left ear. This arrangement clearly enables space-specific neurons to determine that a noise is coming from above or below eye level. The process by which space-specific neurons convert signals from the posterior lemniscal nucleus into vertical coordinates remains to be established, however.

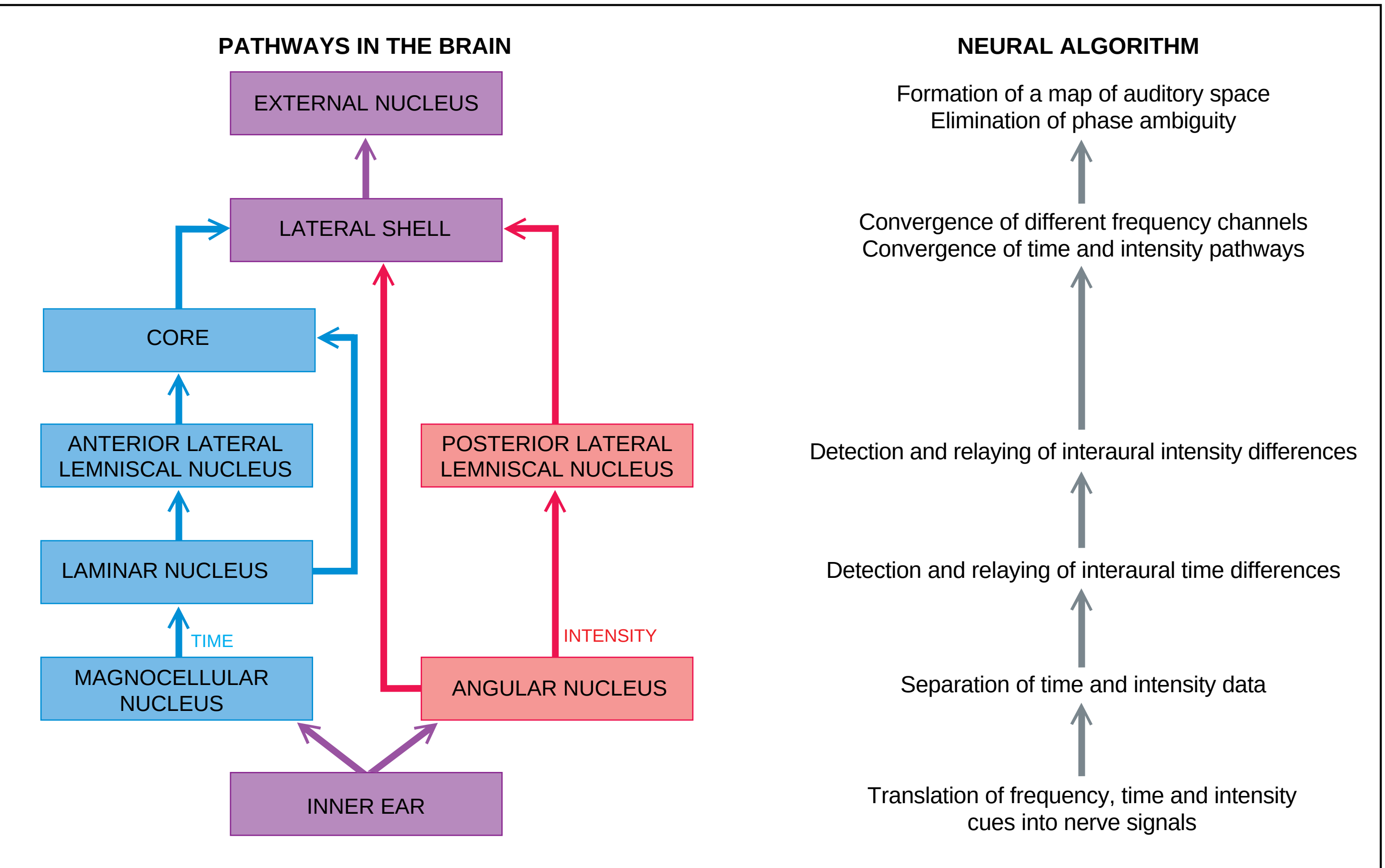
The next higher station is the lateral shell of the midbrain auditory area; neurons from the posterior lemniscal nucleus on each side of the brain send signals to the shell in both hemispheres. In the shell, most neurons respond strongly to both interaural intensity and interaural timing differences generated by sounds within a narrow range of frequencies. This station does not provide the owl with sufficient information to ensure accurate sound location, however, because phase ambiguity persists.

The ambiguity disappears only at the level of the external nucleus, home of the space-specific neurons. These neurons are broadly tuned to frequency, receiving timing and intensity data from many frequency channels. This convergence somehow supplies the input needed for the brain to select the correct coordinates of a sound source. The selectivity of space-specific neurons, then, results from the parallel processing of time and intensity data and from the combination of the results in the shell and in the external nucleus itself.

We have not yet resolved the number of space-specific neurons that must fire in order for an owl to turn its head toward a sound source. Nevertheless, we know that individual neurons can carry the needed spatial data. This fact belies the view of some researchers that single neurons cannot represent such complex information and that perceptions arise only when whole groups of cells that reveal nothing on their own fire impulses collectively in a particular pattern.

Together our neurological explorations have elucidated much of the algorithm, or step-by-step protocol, by which the owl brain achieves binaural fusion. Presumably, we humans follow essentially the same algorithm (although some of the processing stations might differ). Recall, for example, that several lines of evidence suggest mammals rely on delay lines and coincidence detection in locating sounds.

We can extrapolate even further. The only other neural algorithm for a sensory task that has been deciphered in equal detail is one followed by electricity-emitting fish of the genus *Eigenmannia*. Walter F. Heiligenberg of the University of California at San Diego and his associates have worked out the rules enabling members of this species to determine whether their electric waves are of higher or lower frequency than those of other *Eigenmannia* in the immediate vicinity. (In response, a fish might alter the frequency of the wave it emits.) *Eigenmannia* rely on parallel pathways to process separate sensory information. Also, relevant information is processed in steps; the parallel pathways converge



at a high station; and neurons at the top of the hierarchy respond selectively to precise combinations of cues. The fish algorithm is thus remarkably similar to that of the barn owl, even though the problems that are solved, the sensory systems involved, the sites of processing in the brain and the species are different. The similarities suggest that brains follow certain general rules for information processing that are common to different sensory systems and species.

Carver A. Mead, here at Caltech, thinks the owl algorithm may also teach something to designers of analog silicon chips, otherwise known as VLSI (Very Large Scale Integrated) circuits. In 1988 he and John Lazzaro, then his graduate student, constructed an "owl chip" that reproduces the steps through which the barn owl measures interaural time differences. The model, about 73 square millimeters in area, contains only 64 auditory nerve fibers in each ear (many fewer than truly exist) and some 21,000 delay lines. (It also has 200,000 transistors, mainly to regulate the delay lines.) Even in its pared-down version, the electronic nervous system takes up much more space and energy than does the biological system. Historically, engineers have constructed chips according to principles drawn from electronics, physics and chemistry. The economy of the biological circuit suggests that natural principles may help engineers build analog chips that consume less energy and take up less space than usual.

My laboratory's research into the owl brain is by no means finished. Beyond filling in some of the gaps in our knowledge of binaural fusion, we hope to begin addressing other problems. For example, Alvin M. Liberman of Haskins Laboratories in New Haven, Conn., has proposed that the human brain processes speech sounds separately from nonspeech sounds. By the same token, we can ask whether the owl separately processes signals for sound location and other acoustic information. Some brain stations that participate in spatial orientation may also take part in other sensory activities, such as making owls selectively attuned to the calls of mates and chicks. How does the owl, using one set of neurons, sort out the algorithms for different sensory tasks? By solving such riddles for the owl, we should begin to answer some of the big questions that relate to more complex brains and, perhaps, to all brains.

FURTHER READING

A NEURAL MAP OF AUDITORY SPACE IN THE OWL. E. I. Knudsen and M. Konishi in *Science*, Vol. 200, pages 795–797; May 19, 1978.

NEURONAL AND BEHAVIORAL SENSITIVITY TO BINAURAL TIME DIFFERENCES IN THE OWL. Andrew Moiseff and Masakazu Konishi in *Journal of Neuroscience*, Vol. 1, No. 1, pages 40–48; January 1981.

NEUROPHYSIOLOGICAL AND ANATOMICAL SUBSTRATES OF SOUND LOCALIZATION IN THE OWL. M. Konishi, T. T. Takahashi, H. Wagner, W. E. Sullivan and C. E. Carr in *Auditory Function: Neurobiological Bases of Hearing*. Edited by G. M. Edelman, W. E. Gall and W. M. Cowan. John Wiley & Sons, 1988.

A CIRCUIT FOR DETECTION OF INTERAURAL TIME DIFFERENCES IN THE BRAIN STEM OF THE BARN OWL. C. E. Carr and M. Konishi in *Journal of Neuroscience*, Vol. 10, No. 10, pages 3227–3246; October 1990.

THE NEURAL ALGORITHM FOR SOUND LOCALIZATION IN THE OWL. M. Konishi in *The Harvey Lectures*, Series 86, pages 47–64; 1992.

Catalysis on Surfaces

Scientists can now observe how solids interact with individual molecules to speed reactions. Information about these catalysts is being used to improve everything from materials synthesis to pollution control

by Cynthia M. Friend

When atoms on the surface of a solid interact with molecules of a gas or liquid, they may alter the structure of the molecules ever so slightly, thereby promoting unusual chemical reactions. Indeed, by investigating the interactions between molecules and the surfaces of solids, researchers have learned to synthesize a myriad of novel substances, develop chemical processes of unprecedented efficiency and remove pollutants from the environment. Yet the study of chemistry on solid surfaces will have the most impact, in my opinion, on the technology of catalysts—substances that increase the rates of desirable chemical reactions at the expense of others.

The rise of surface chemistry is a recent event. Some 15 years ago the subject was just emerging as a branch of science, and only a few methods were available for the analysis of complex surface reactions. In recent years, I have watched the field advance rapidly as techniques were devised to create detailed pictures of surface reactions. By using the tools of surface chemistry, researchers can now view the action of a catalyst at the molecular level.

To illustrate how surface chemistry has advanced our understanding of catalytic reactions, I have focused on three cases that are simple yet have great scientific, technological and social significance. The first is the synthesis of ammonia from air—a process that made possible the large-scale production of fertilizers and explosives. Although chemists invented the commercial process for ammonia synthesis in the early part of this century, they did not understand the nuances of the catalytic reaction until the techniques of surface chemistry were applied several years ago.

The second case is the breakdown of nitric oxide in the exhaust of automobiles. By reducing the emissions of nitric oxide from car exhaust, chemists have helped alleviate the problem of acid rain and mitigate the effects of other harmful pollutants in the atmosphere. The final case is the removal of sulfur from fossil fuels, a process that also has implications for the atmosphere. The desulfurization of fuels involves particularly complicated surface reactions, and the investigation of such phenomena represents the frontier in surface chemistry.

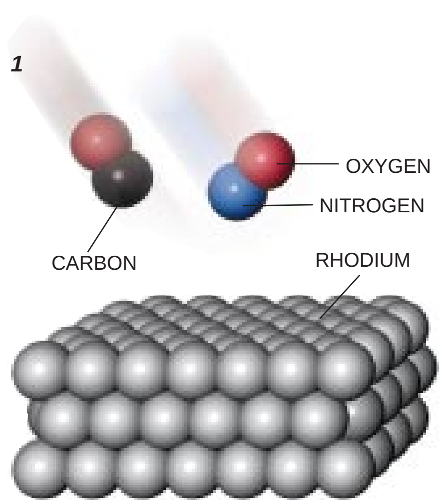
In 1909 Fritz Haber, a German chemist, found an efficient way to synthesize ammonia from nitrogen and hydrogen gas. Five years later commercial production began at Badische Anilin- & Soda-Fabrik (BASF)—an event that later proved to be a bane and a boon to all humanity. The Haber process provided Germany with the main ingredient for manufacturing explosives, such as nitroglycerine, for use in World War I. Previously, Germany had made explosives from saltpeter mined in Chile. If the supply had been blocked and the Haber process had not been invented, some historians argue, World War I probably would have been shortened.

The Haber process also revolutionized agriculture by fostering the production of mass quantities of cheap fertilizer. As a direct consequence, farmers achieved higher crop yields than ever before and greatly increased the global food supply.

To produce ammonia efficiently, Haber added iron as a catalyst to a mixture of nitrogen and hydrogen gas and then subjected the reaction vessel to pressures of some 100 atmospheres and to temperatures around 500 degrees Celsius. As nitrogen and hydrogen circulated over the catalyst, ammonia condensed out of the mixture.

Haber's major contribution was the identification of a good catalyst for the synthesis of ammonia. He tested more than 1,000 materials before settling on iron. Today manufacturers of ammonia use a catalyst consisting of iron, potassium and calcium, which performs better than iron alone.

CYNTHIA M. FRIEND is professor of chemistry at Harvard University and a trustee of Radcliffe College. After earning her Ph.D. in chemistry from the University of California, Berkeley, in 1981, she conducted research in surface chemistry at Stanford University. Her work has earned her the Garvin Medal of the American Chemical Society and the Presidential Young Investigator Award. For the past decade at Harvard, Friend has been a pioneer in applying surface chemistry techniques to investigate catalysis.



CATALYST made of the metal rhodium is used in the exhaust systems of automobiles to help transform nitric oxide (NO) and carbon monoxide (CO) into harmless gases. The tools of surface chemistry enabled researchers, in 1986, to learn how the rhodium catalyst works at the molecular level. Carbon monoxide (*1*) binds to the rhodium surface (*2*). When nitric oxide does the same, it dissociates into oxygen and nitrogen (*3*). The bound oxygen reacts with the CO to form carbon dioxide (CO_2) (*4*). If another CO and NO land close to the remaining bound nitrogen (*5–8*), the rhodium catalyst promotes the formation of a second carbon dioxide molecule and a nitrogen molecule (N_2).

Iron, like all catalysts, increases the rate of a desirable reaction—the synthesis of ammonia, in this case—while simultaneously diminishing the chance of undesirable reactions—for instance, the recombination of nitrogen atoms to form nitrogen gas. The catalytic action occurs when reactant molecules temporarily bind to the surface of the solid. This binding alters the forces between atoms, thereby changing the energetic requirements for the reaction. Consequently, in the presence of a catalyst, some processes will be favored, and others will not. Catalysts are never consumed or produced during the reaction.

For decades after the commercial process for ammonia synthesis was invented, researchers did not understand at the molecular level how the iron worked as a catalyst. The tools of surface chemistry have recently made it possible, however, to confirm that the primary function of the catalyst is to facilitate the dissociation of nitrogen gas (N_2). The bond that holds the N_2 molecule together is very strong and therefore will not break unless the molecule contains a substantial amount of energy (941,000 joules per mole of nitrogen, to be exact). As the energy content of an N_2 molecule increases, the average distance between the nitrogen atoms grows. The process is analogous to climbing a mountain. Just as a hiker must burn calories to get to the top of a mountain, the N_2 molecule must contain energy to surmount a high energy barrier. And just as a fast-moving hiker reaches the summit quickly, a molecule that has a lot of energy will more readily pass over the barrier. (The number of high-energy molecules in a gas increases in proportion to the temperature.)

The rate at which N_2 dissociates depends on the rate at which the molecules can surmount the barrier, and these rates, in turn, are related to the available energy, or temperature, of the molecules. Ammonia does not form readily, however, if nitrogen and hydrogen molecules are heated simply to the point at which they dissociate. At such temperatures, ammonia itself breaks up as quickly as it is formed.

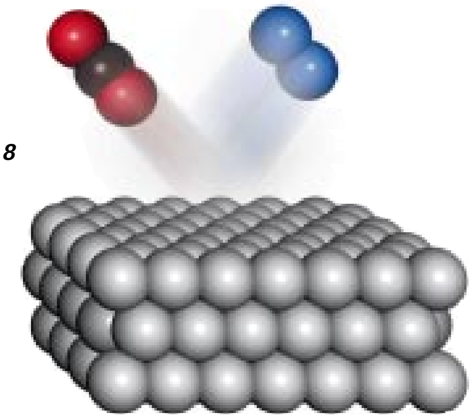
Yet if nitrogen molecules interact with iron atoms on a surface, they will dissociate at relatively low temperatures. Iron catalyzes the dissociation of N_2; in other words, it lowers the energy barrier so that the two nitrogen atoms will separate easily. The iron atoms accomplish this by donating electrons to the nitrogen molecule. As a result, the iron atoms form a chemical bond with the nitrogen molecule, and in a reciprocal manner, the bond between the two nitrogen atoms in the molecule is weakened. The weak nitrogen-nitrogen bond corresponds to a low energy barrier and facilitates N_2 dissociation. This process is typical of how catalysts work.

In 1984 Gerhard Ertl, Michael Grunze and Min-Chi Tsai and their co-workers, then at the University of Munich, obtained the first direct evidence for the weakening of the nitrogen-nitrogen bond. Their work involved the examination of samples made from single crystals of iron under vacuum conditions—that is, an environment free from extraneous gases. Unlike the atoms in ordinary iron particles, those in a single crystal form a perfectly ordered lattice. Because some chemical reactions are extremely sensitive to the arrangement of atoms on the surface, the use of single crystals permits the methodical study of these effects. Researchers can even vary the surface arrangement by cutting crystals at various angles.

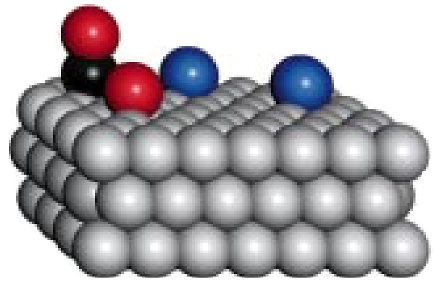
To observe the reactions, the Munich group used two of the most important tools in surface chemistry: x-ray photoelectron spectroscopy and high-resolution electron energy loss (HREEL) spectroscopy. By probing the forces between atoms, these techniques let chemists "see" the changes that take place when a molecule interacts with the surface.

X-ray photoelectron spectroscopy, which was first developed in the early 1960s, enables researchers to measure the energy of strongly bound elec-

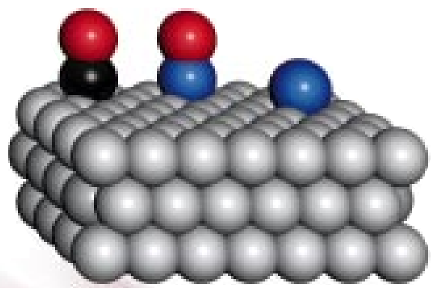
8



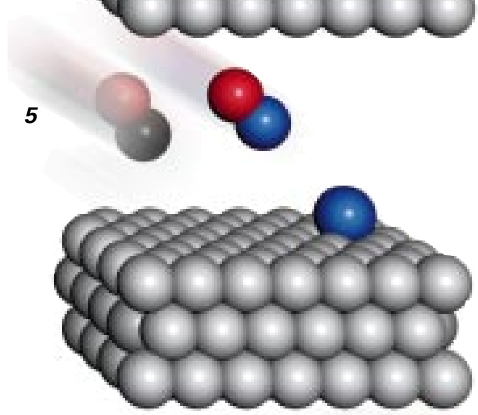
7



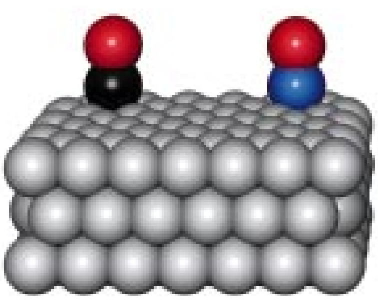
6



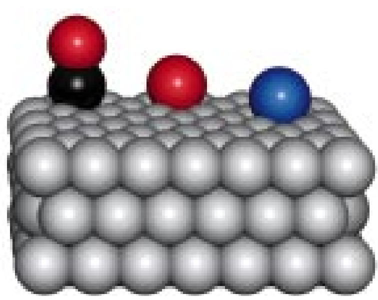
5



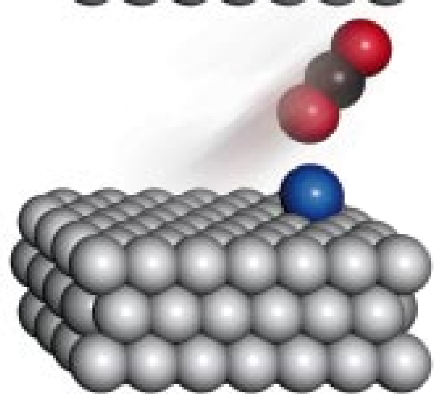
2



3



4



trons in molecules. In this method, x-rays hit the experimental sample, thereby knocking electrons from the atoms of the surface and from any molecules bound to the surface. The energy of these electrons is sensitive to the local chemical environment, and thus the photoelectron technique yields information about bonding (for instance, the breaking of the nitrogen-nitrogen bond in N_2).

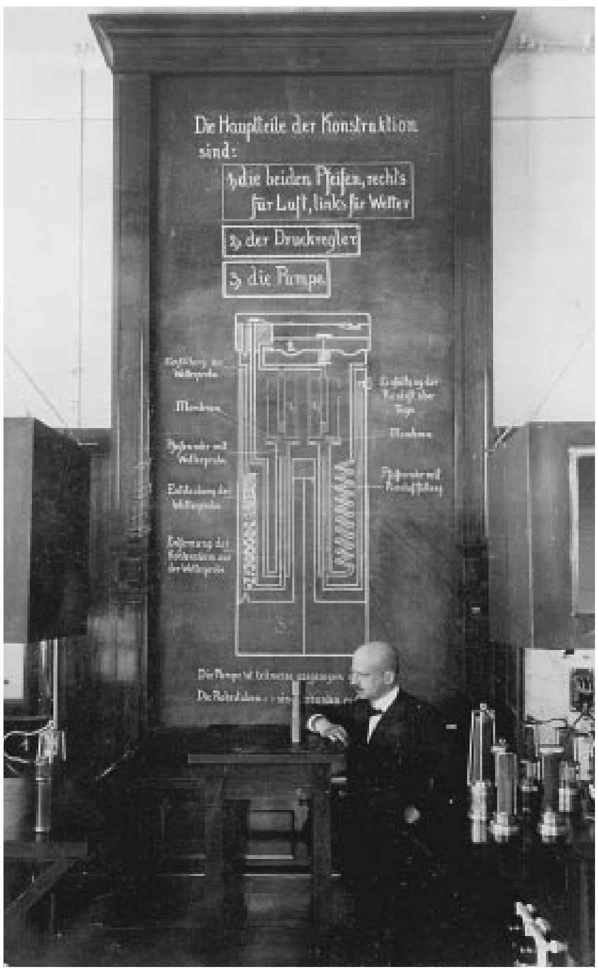
HREEL spectroscopy, which was invented in the late 1970s and refined in the mid-1980s, probes the forces between atoms in molecules by measuring the energies of molecular vibrations. The atoms in a molecule can be envisioned as balls held together by springs. The springs represent the chemical bonds; the balls are the atomic nuclei. The strength of the bond is related to the stiffness of the spring and therefore depends on the energy required to stretch the atoms apart or to change the relative angle of the bond.

The results of HREEL spectroscopy are rather straightforward to interpret in the case of a nitrogen molecule. The only way the nitrogen molecule can vibrate is through the stretching and compression of the nitrogen-nitrogen bond; a nitrogen molecule has a single vibrational mode, to use the jargon of chemistry. If a nitrogen molecule is bombarded with electrons, it will vibrate and then rob them of a characteristic amount of energy. These electrons can be detected to reveal the energy of the vibrations and thereby the strength of the spring, or bond, that holds the molecule together. The same technique can be used for nitrogen molecules that bond to the surface, but the results are bit more complicated. A nitrogen molecule bound to a surface has more than one vibrational mode because there are additional "springs" associated with the metal-nitrogen bonds.

HREEL spectroscopy enabled the Munich group to demonstrate that when an N_2 molecule is adsorbed on an iron surface, the energy required to stretch the nitrogen atoms apart is decreased dramatically compared with the energy needed to lengthen the bond of a nitrogen molecule floating freely.

Besides weakening the nitrogen-nitrogen bond, the techniques of surface chemistry have shown that iron performs two other important functions in the synthesis of ammonia. First, when hydrogen molecules (H_2) interact with the iron surface, the hydrogen-hydrogen bond is weakened. Hydrogen atoms are thus easily freed and bind to the surface. Second, the iron atoms confine the hydrogen and nitrogen atoms on the surface so that they may react to form NH, NH_2 and finally NH_3, ammonia, the desired product.

The chance that a bond will form between the nitrogen and hydrogen atoms depends on how well the nitrogen attaches to the iron. In fact, the iron-nitrogen bond is sufficiently strong so that nitrogen does not recombine to form N_2, and it is sufficiently weak so that the nitrogen can combine with hydrogen. Several metals are more effective than iron in dissociating nitrogen molecules into their component atoms, but typically the atoms are bound so strongly to the metal that they do not combine with hydrogen to form ammonia. These insights and others are helping researchers to figure out why one catalyst is better than another and ultimately may lead to the discovery of practical catalysts that will improve the commercial synthesis of ammonia and enhance the efficiency of related industrial processes.



FRITZ HABER developed, in 1909, the first commercial process for synthesizing ammonia. One of his important insights was that iron catalyzes the reaction that transforms hydrogen and nitrogen gas from ammonia.

When applying the results of surface studies to commercial catalysis, investigators face a difficult problem. The tools of surface chemistry require vacuum conditions, whereas most practical catalysts must work in a high-pressure environment.

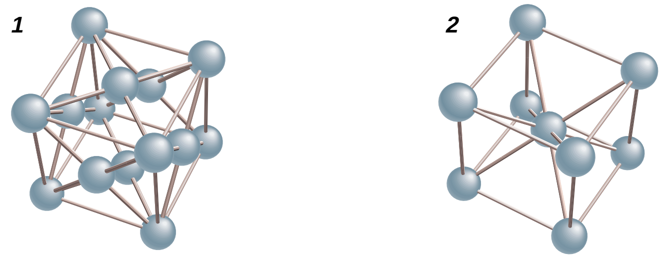
Indeed, surface chemists have an extreme definition of clean. By their standards, the surface of a sample must remain relatively free of contaminants for at least a few hours. It must therefore be stored under ultrahigh vacuum conditions—that is, pressures 10 trillion times lower than that of the atmosphere. The extremely low pressures enable chemists to study surfaces of known composition and to introduce reactants in a purposeful way. Such studies could never be done, say, in the open air because a perfectly clean metal surface would be covered with gas molecules in about a billionth of second. Even if the pressure were reduced to a billionth of an atmosphere, the surface would be covered in about one second, insufficient time to perform an experiment.

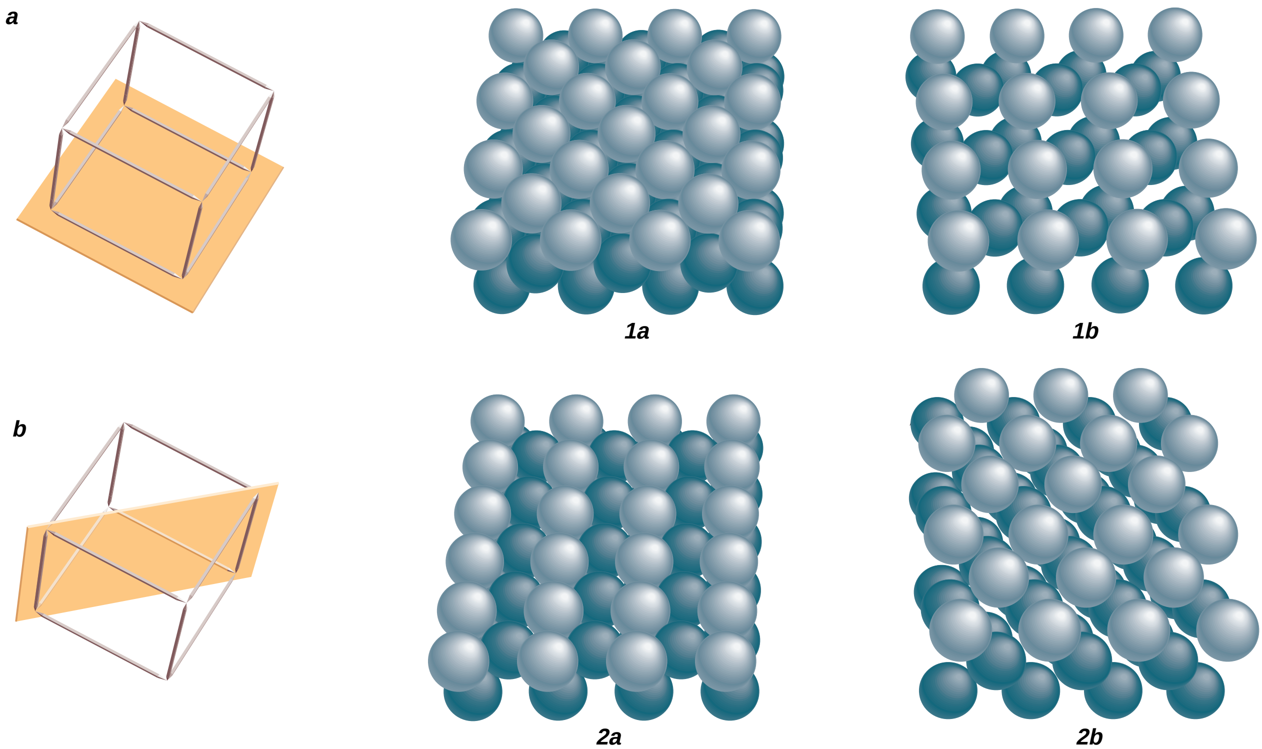
Researchers need to be careful about how they apply the findings of surface chemistry to high-pressure catalytic reactions. The great disparity in pressure means a vast difference in the number of gas molecules hitting a catalyst at any one time, and therefore the kinetics of the laboratory reaction may differ from the dynamics of the high-pressure reaction. Although chemists could not initially figure out how to compensate for the so-called pressure gap, they recently learned to extrapolate the behavior of high-pressure catalytic reactions from idealized surface studies.

An elegant example of how the pressure gap can be bridged is the story of a set of surface chemistry experiments aimed at improving the technology of catalytic converters for automobiles. The primary function of the converters is to remove nitric oxide (NO) and carbon monoxide (CO) from automo-

Surface Geometry

Arrangement of atoms on the surface of a catalyst depends on how the atoms are packed together and how the material is cut. Two typical atomic lattices are the face-centered cubic (*1*) and the body-centered cubic (*2*). Different cuts (*a*, *b*) through such lattices yield different surface arrangements (*1a*, *1b*, *2a*, *2b*).





bile exhaust. Nitric oxide reacts rapidly with air to form nitrogen-oxygen compounds (NO_x), which are harmful to the environment, notably because they contribute to acid rain. Carbon monoxide is also extremely toxic to most forms of life. By transforming NO and CO from car exhaust into less harmful products, catalytic converters have dramatically reduced the levels of these contaminants. The technology has been very successful, yet chemists continue to search for better catalysts, realizing that small improvements can mean tremendous benefits for the entire environment.

A typical catalytic converter consists of particles of platinum (Pt) and rhodium (Rh) deposited on a ceramic honeycomb. The platinum and rhodium particles catalyze the reactions that remove NO_x, CO and uncombusted hydrocarbons from car exhaust. The ceramic honeycomb and small particle size serve the dual function of maximizing the exposure of the metals to the exhaust fumes and minimizing the quantity of platinum and rhodium—two very expensive metals.

In the mid-1980s researchers at General Motors and elsewhere set out to investigate how rhodium interacts with the nitric oxide and carbon monoxide from car exhausts. To do so, they studied reactions of NO and CO on single crystals of rhodium. Employing HREEL spectroscopy and other methods, the workers identified the key steps in the breakdown of nitric oxide, and they determined how the arrangement of rhodium atoms on the surface influenced catalysis.

Yet it was not clear at first whether these results were applicable to the technology of catalytic converters: the experiments were conducted under vacuum conditions, whereas the rhodium particles used in catalytic converters are exposed to gases at high pressure. To demonstrate that practical information could be gained from the surface chemistry studies, the GM workers tested the rate of NO reduction on a rhodium surface in an environment that replicated the pressure conditions in a catalytic converter. By analyzing the action of rhodium in vacuum conditions and under high pressure, the researchers devised a mathematical model of the catalytic process. The model has allowed the results of the surface studies to be used in determining how new kinds of catalytic materials will operate under high pressures.

The GM group also found that the dissociation of NO was sensitive to the arrangement of atoms on the rhodium surface. They arrived at this conclusion by using infrared spectroscopy. This technique is similar to HREEL spectroscopy, but it uses infrared radiation, instead of electrons, to cause molecular vibrations. Infrared spectroscopy has an advantage, however: it can be used both under vacuum conditions and at high pressures. Researchers at GM were therefore able to study, at high and low pressure, how nitric oxide interacts with irregular particles of rhodium and how it binds to a rhodium surface in which the atoms are arranged in a hexagonal pattern. Infrared spectroscopy identified differences in the vibrations of NO on the two types of surfaces, proving that surface structure influences bonding and ultimately the catalytic process.

To study, in even finer detail, how surface structure affects such reactions as the dissociation of nitric oxide on rhodium, chemists have recently begun to turn to the scanning tunneling microscope. In 1986 Gerd K. Binnig and Heinrich Rohrer won the Nobel Prize for inventing the device because it revolutionized the study of surface structure. To produce atomic-scale images of a surface, a scanning tunneling microscope positions a fine metal stylus only a few angstroms above a sample and then moves the stylus across the surface. The microscope senses electrons as they pass between the stylus and the surface. Because the flow of electrons is related to the height of the atoms on the surface, the information can be translated into images.

Scanning tunneling microscopy is just beginning to provide insights into how surface structure influences catalysis. It should permit chemists to identify the surface structures that optimize the performance of catalysts.

Surface chemistry has also helped researchers understand the catalytic process for removing sulfur from fossil fuels. Traces of sulfur in fossil fuels harm the environment in two ways. First, when fuel is burned in an engine, some sulfur reacts with air to form sulfur-oxygen compounds, which contribute to acid rain. Second, sulfur sticks to the platinum and rhodium in catalytic converters, thereby shutting down their activity and indirectly increasing the emission of NO and CO.

Sulfur is removed from petrochemicals at refineries as crude oil is transformed into such useful hydrocarbons as octane. Ideally, the desulfurization process should extract all the sulfur without destroying the valuable hydrocarbons. The process therefore requires a catalyst that encourages desulfurization but discourages the breakdown of pure hydrocarbons. At present, the best catalyst for desulfurization is a mixture of molybdenum, cobalt and sulfur itself. Because the material has a very complicated structure, chemists have had difficulty figuring out how petrochemicals interact with the catalyst. It is not clear whether all components of the catalyst are effective in promoting desulfurization. For example, one component, cobalt sulfide, is not thought to be the active catalytic material.

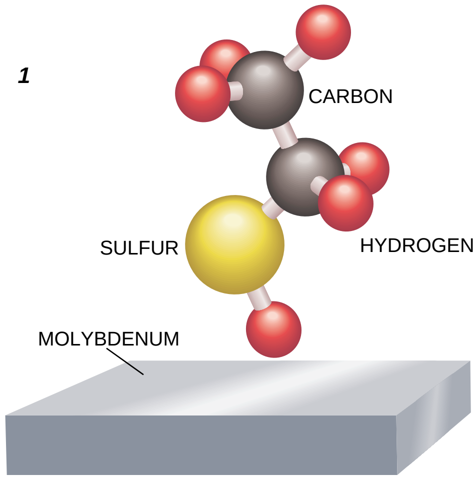
My colleagues and I at Harvard University have investigated the role of molybdenum in desulfurization. We are interested in the question of how petrochemicals interact with molybdenum in various forms and in combination with cobalt and sulfur. Our goal is to construct a general model that can be used to predict the products and rates of reactions for all the common sulfur-containing hydrocarbons.

During the past eight years, we formulated such a general model. It describes how thiols, a major class of sulfur-containing petrochemicals, interact with various molybdenum surfaces. To be precise, thiols consist of a hydrogen atom attached to a sulfur atom, which in turn is bonded to some combination of carbon and hydrogen. To study the interaction between thiols and molybdenum catalysts, we used crystals of pure molybdenum so that the atoms on the surface would form a highly ordered pattern. The regular structure limits the number of different kinds of sites available for bonding. By introducing various components to the molybdenum surface in a systematic way, we could then infer the role of each component in desulfurization.

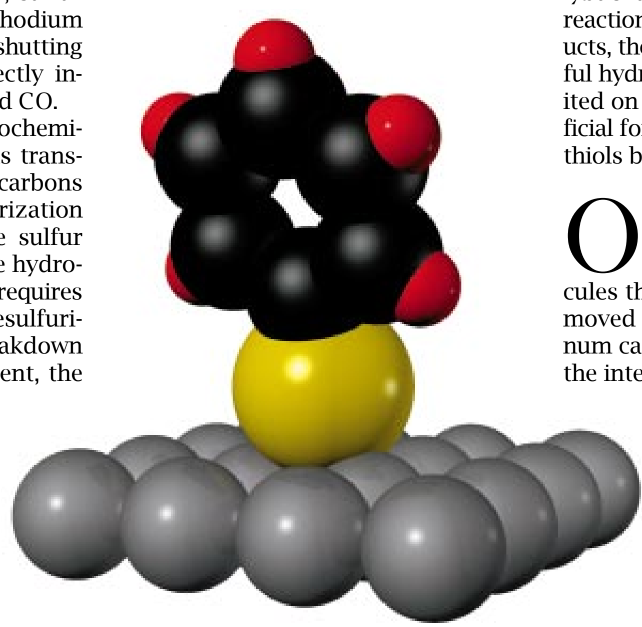
Many years ago chemists made an astonishing discovery about molybdenum-induced desulfurization. The performance of pure molybdenum is actually enhanced when it is contaminated with sulfur, whereas most metals lose their ability to catalyze reactions when sulfur bonds to them. To study the role of surface sulfur, we compared how thiols react on clean molybdenum surfaces with how they perform on molybdenum covered with an ordered array of sulfur. Thiols interact with molybdenum catalysts to produce hydrocarbons, hydrogen gas, surface carbon and sulfur. For example, ethanethiol ($CH_3(CH_2)SH$) is broken down into sulfur and one of two hydrocarbons: ethane (CH_3CH_3) and ethene (CH_2CH_2). Ideally, the catalyst should promote only the production of hydrocarbons and the removal of sulfur; it should discourage the synthesis of surface carbon and hydrogen gas, which have little value.

The molybdenum surface coated with sulfur removes sulfur from thiols more slowly than the clean surface can, but at the same time, the sulfur on the molybdenum surface decreases the rate of reactions that lead to undesirable products, thereby increasing the yield of useful hydrocarbons. Indeed, sulfur deposited on a molybdenum surface is beneficial for the desulfurization not only of thiols but also of other petrochemicals.

One major objective of our research has been to discover the important intermediate molecules that are produced as sulfur is removed from thiols using a molybdenum catalyst. We attempted to identify the intermediates by applying a combination of x-ray photoelectron and HREEL spectroscopies. Our analysis was complicated by the fact that thiols consist of a large number of atoms and may react with the surface in many different ways. Furthermore, the large size of the reactants made it difficult to interpret spectroscopic results; molecules composed of many atoms have many vibra-

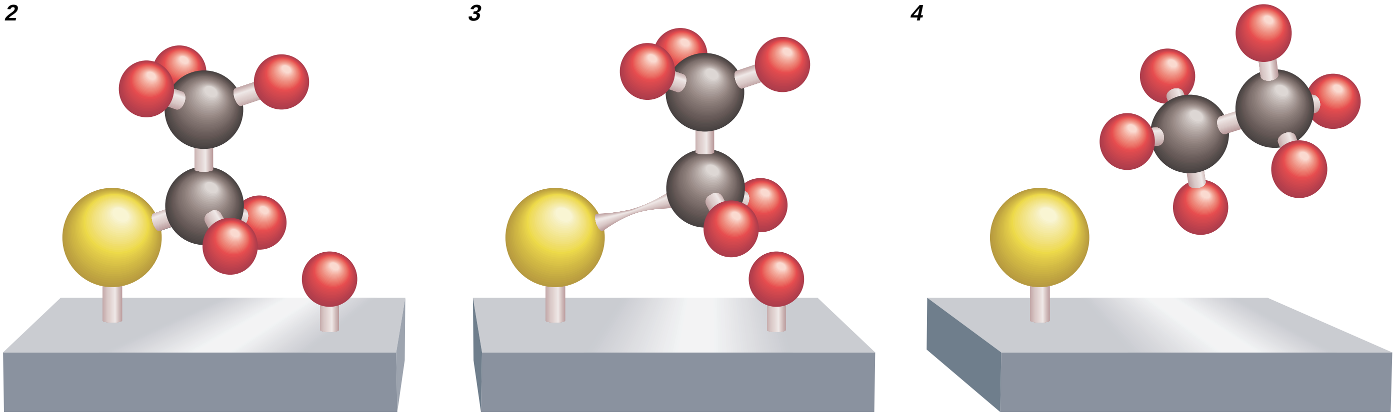


MOLYBDENUM catalyzes a reaction that removes sulfur from fuels. When a mol-



SULFUR-CONTAINING MOLECULE binds to a molybdenum surface. The orientation of the molecule on the surface was deduced using electron spectroscopy.

2 3 4



ecule (*1*) based on sulfur, carbon and hydrogen interacts with molybdenum, it loses a hydrogen and binds to the surface (*2*). The bond between carbon and sulfur then weakens (*3*), and the remaining atoms may recombine to form ethane (*4*).

tional and electronic states. Nevertheless, we were able to distinguish several of the steps in desulfurization of thiols on molybdenum.

A detailed analysis of spectroscopic data can reveal the identity of an intermediate molecule on the surface. The data can also provide information about how the bonds within an intermediate change as the molecule interacts with the surface. In some cases, we can even deduce the orientation of the intermediates from spectroscopic studies. Such structural information is exceedingly important in creating theoretical models of the bonding of thiols and their associated intermediates to surfaces. Infrared spectroscopy is particularly versatile as a structural probe because the vibrations within a molecule that are detected in an infrared experiment are sensitive to the symmetry of the molecule.

Such successes have encouraged surface chemists to extend their experimental reach by inventing additional tools for determining molecular structure on surfaces. Moreover, this effort is one of the frontiers in surface chemistry. Several of the emerging techniques require radiation from a huge electron accelerator known as a synchrotron. By accelerating electrons around a circular track to speeds approaching that of light, a synchrotron can generate powerful beams of x-rays and other kinds of radiation. To utilize the x-ray beam for spectroscopy, researchers must design their experiments and measurement apparatus so that they interface with the x-ray source.

During the past few years, we have used synchrotron, infrared and other techniques to figure out what the stages of the desulfurization are and what steps are crucial in setting the overall rate of the reaction. The first step is the cleavage of the sulfur-hydrogen bond. We found that this step occurs very rapidly and is favored because both sulfur and hydrogen form strong bonds to the molybdenum surface. In subsequent steps, the carbon-sulfur bond must break, and one carbon-hydrogen bond is either formed or broken to yield the hydrocarbon products from the thiol. For example, ethanethiol bonds to the catalyst, breaking the sulfur-hydrogen bond and becoming ethyl thiolate ($CH_3(CH_2)S$). The carbon-sulfur bond is broken next. Finally, the formation of a carbon-hydrogen bond yields ethane, whereas the breaking of such a bond leads to ethene.

The intermediates that form as ethyl thiolate becomes ethane or ethene are too short-lived to be detected by means of surface spectroscopy. Therefore, indirect probes must be used. These probes indicate that the breaking of the carbon-sulfur bond is the rate-limiting step in the desulfurization of ethyl thiolate. More important, we have found evidence that for any thiol molecule, the strength of the carbon-sulfur bond determines the overall rate of the desulfurization reaction.

By using our general model and information about the strength of the carbon-sulfur bond, we have been able to predict the rates of reaction and the types of products formed during the desulfurization of thiols on molybdenum surfaces. Surface-bound carbon and gaseous hydrogen—the undesirable products—are formed at a rate that depends largely on whether a carbon-hydrogen bond can be broken before the breaking of the carbon-sulfur bond. According to fundamental principles of chemistry, therefore, the fraction of thiol intermediates that lead to useful hydrocarbons is proportional to the rate of carbon-sulfur bond breaking relative to the rate of carbon-hydrogen bond breaking. Hence, we predicted that thiols with low carbon-sulfur bond strengths—that is, high rates of bond breaking—would yield a large fraction of hydrocarbons. This conclusion has been borne out by all our experiments.

The desulfurization of fossil fuels, the removal of nitric oxide from car exhaust and the synthesis of ammonia are just three of the many areas that have benefited from surface chemistry. Yet researchers need to continue to make technological and conceptual advances if we are truly to understand how surfaces influence the bonding and structure of complex intermediates. The next decade promises major advances in experimental and theoretical methods that should give a tremendous boost to our knowledge of the complicated catalytic reactions that occur on surfaces.

FURTHER READING

π-BONDED N_2 ON FE(111): THE PRECURSOR FOR DISSOCIATION. M. Grunze, M. Golze, W. Hirschwald, H.-J. Freund, H. Pulm, U. Seip, M. C. Tsai, G. Ertl, J. Küppers in *Physical Review Letters,* Vol. 53, No. 8, pages 850–853; August 20, 1984.

MOLECULAR TRANSFORMATIONS ON SINGLE CRYSTAL METAL SURFACES. R. J. Madix in *Science,* Vol. 233, pages 1159–1166; September 12, 1986.

MECHANISM OF THE NITRIC OXIDE–CARBON MONOXIDE–OXYGEN REACTION OVER A SINGLE CRYSTAL RHODIUM CATALYST. Galen B. Fisher et al. in *Proceedings of the 9th International Congress on Catalysis,* Vol. 3: *Characterization and Metal Catalysts.* Edited by M. J. Phillips and M. Ternan. Chemical Institute of Canada, 1988.

SURFACE CHEMISTRY. John T. Yates, Jr., in *Chemical & Engineering News,* Vol. 70, No. 13, pages 22–35; March 30, 1992.

The Reproductive Behavior of the Stickleback

To reproduce, this tiny fish engages in behaviors not commonly associated with such animals, including luring intruders away and cannibalizing another's eggs

by Gerard J. FitzGerald

It was a sunny spring afternoon, and I was lying on my stomach peering into one of the myriad tide pools that dot a vast stretch of salt marsh along the southern shore of the Saint Lawrence estuary. Concentrating on my observations, I did not hear the approach of a local resident, who suddenly appeared beside me. Apart from the two of us, there was no one for miles around. In the distinctive Québec French of the region, he queried, "*Qu'est-ce que tu fais là?*" I replied, "*Je regarde les poissons.*" The poor man beat a hasty retreat, no doubt wondering what this crazy "Anglais" was doing watching inedible fish by himself in the middle of a mosquito-infested marsh.

The subject of my curiosity was the three-spine stickleback, *Gasterosteus aculeatus,* a small temperate-zone fish about the length of one's middle finger. The name comes from the three sharp dorsal spines that help to protect the fish from its numerous predators.

Although the fish may seem rather insignificant to the casual observer, they are in fact a source of endless fascination to behavioral ecologists. Older readers of this magazine may have read an article by Nobel Prize–winner Nikolaas Tinbergen [see "The Curious Behavior of the Stickleback," by N. Tinbergen; SCIENTIFIC AMERICAN, December 1952]. He described how males establish territories, build nests and court females to obtain eggs. As a boy, Tinbergen had watched sticklebacks in the streams of his native Holland, often bringing them home in jam jars. After becoming a lecturer at the University of Leiden, Tinbergen decided to use the fish in laboratory experiments. In his words: "I soon discovered that in choosing these former pets I had struck oil."

After studying this little fish for some 15 years, I now know what the great ethologist meant. The stickleback's behavior is much richer than simple stimulus and response. Females choose a male based on a variety of visual cues, such as his color and the quality of his nest site. Some males are cuckholders, fertilizing the eggs in another male's nest. Females themselves are more than just egg layers; they try to enhance their reproductive success by eating the eggs of other females. Such complex behavior, coloration, widespread distribution throughout the world and ease of study in the laboratory and in the field make the stickleback an excellent creature by which to determine how various behaviors and physical characteristics enable an animal to adapt to its surroundings. The reproductive behavior of the stickleback offers an ideal testing ground for concepts in adaptive evolution and sexual selection.

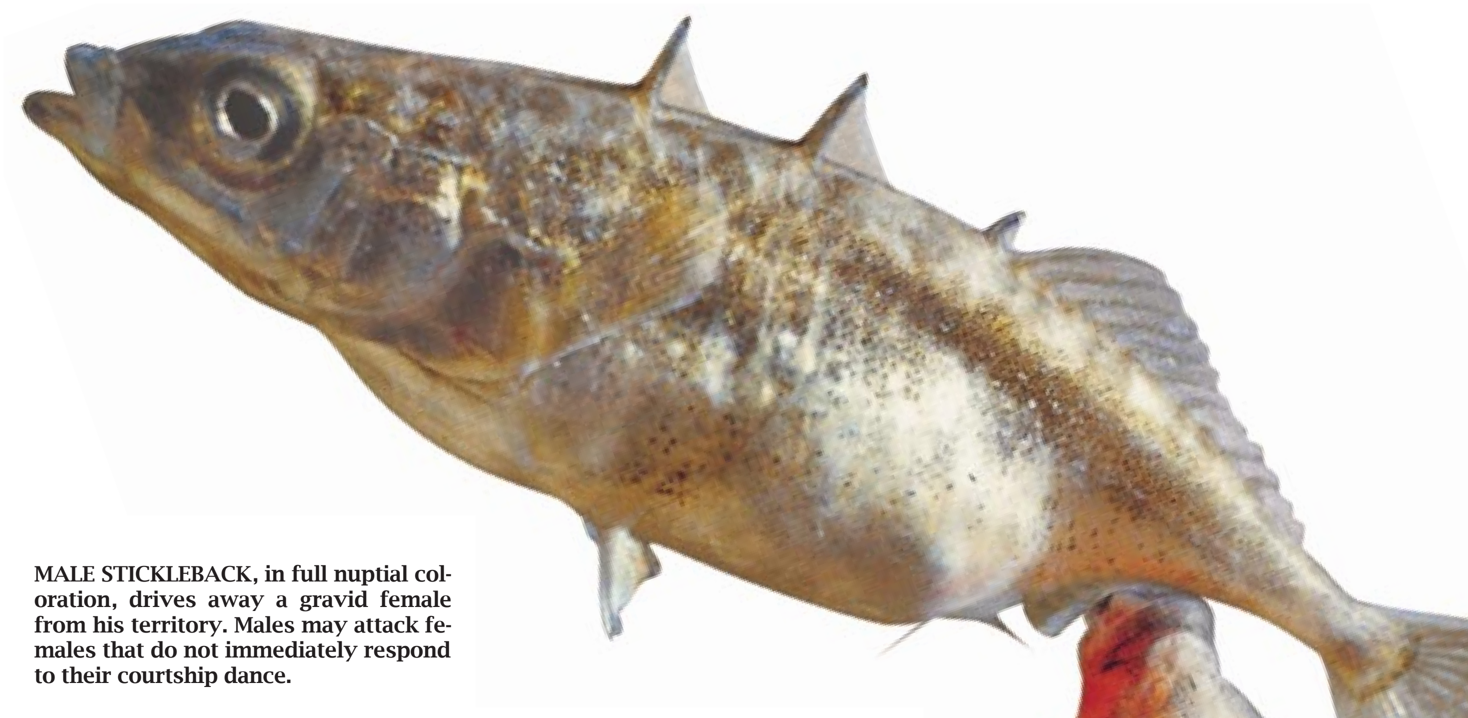
Several events inspired me to follow Tinbergen and others in deciding to concentrate my research on three-spine sticklebacks. When I collected nests in tide pools, I found wide variation in the number of eggs in the nests. Females were apparently rejecting some suitors. I noticed that after being courted by some males and entering their nests, females frequently backed out. Shortly thereafter, the same females often spawned with a rival male. Furthermore, I discovered additional behavior patterns that had not been previously reported by laboratory scientists. Clearly, stickleback reproductive behavior was more complex than the earlier laboratory experiments suggested.

Like most young Ph.D. ethologists in the mid-1970s, I was influenced by the intellectual revolution sweeping the field of animal behavior. Many researchers were switching from studies on the proximate basis of behavior, in which the stickleback played a key role, to studies on ultimate explanations of behavior. How does a behavior pattern contribute to survival value and fitness? I began work on the three-spine stickleback to address questions about the adaptive significance of its behavior.

Many of the new insights about stickleback behavior arose because of field studies conducted during the past 15 years. Most work had been done in the laboratory, and whether those findings applied to field observations had been debatable. When I was a doctoral student under the supervision of Miles H. A. Keenleyside at the University of Western Ontario, I had studied the parental behavior of tropical fishes in the laboratory. I was frequently bothered by questions about the relevance of my laboratory findings. I had had no opportunity to observe my study animal in the field. Not surprisingly, I desperately wanted to watch the fish in their natural habitat.

Shortly after I began my career at Laval University in 1976, one of my colleagues told me that large numbers of sticklebacks were breeding in a salt marsh near Isle Verte, only a few hours' drive east of Québec City. This site proved to be a superb natural laboratory. It contains hundreds of shallow,

GERARD J. FITZGERALD is *professeur titulaire* of biology at Laval University in Québec City, where he teaches behavioral ecology. He received an undergraduate degree in psychology and biology from Memorial University in Newfoundland, a master's degree in ecology from McGill University and a doctorate in zoology from the University of Western Ontario. FitzGerald's current interests focus on the tactics and strategies that animals living in harsh environments use to maximize their reproductive success. He has written more than 80 articles, mostly on sticklebacks.



MALE STICKLEBACK, in full nuptial coloration, drives away a gravid female from his territory. Males may attack females that do not immediately respond to their courtship dance.

brackish tide pools, and its only fish are three species of sticklebacks. Best of all, the fish are easy to see and to catch, allowing observations of undisturbed behavior and manipulative experiments in the field.

The sticklebacks enter the Isle Verte marsh in early spring, soon after the ice melts. With the spring tides, shoals of up to 1,000 individuals move from their offshore overwintering areas in the Saint Lawrence estuary to breed in the pools. After the tides recede, the fish are trapped in the pools until the next flooding, usually some two or three weeks later. The young of that year and any surviving adults must return to the estuary with an outgoing flood before the pools freeze in November.

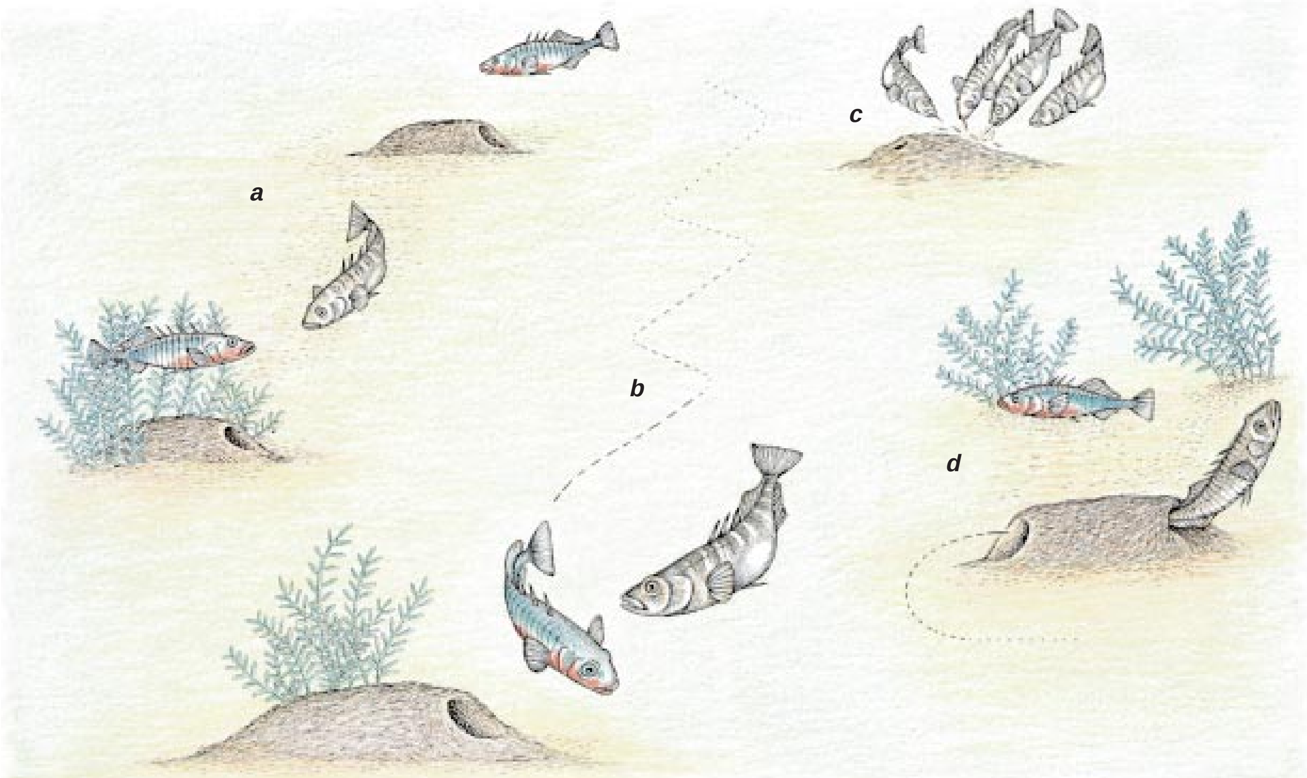
Not long after their arrival at the breeding grounds, the males separate from the shoals and compete to establish territories in the tide pools. Inside the territory the male builds a mound by cementing a mix of sand, mud, filamentous algae or small pieces of dead vegetation together with a glue secreted from its kidney. Next it bores a tunnel by wriggling through this mound of material. The tunnel is the nest.

After constructing the nest, the male, now sporting its nuptial coloration of a bright, red body and blue eyes, is ready to court females, which are a shiny silver color. The male attracts the female with what might be the most famous courtship dance in ethology. He swims in a series of zigzags: first a sideways lunge toward her, then one away from her. If sufficiently enchanted, the female follows the male as he turns and swims rapidly toward the nest. There the male first makes a series of rapid thrusts with his snout into the entrance before turning on his side and raising his dorsal spines. Next the female propels herself forward into the nest with strong tail beats.

After the female enters the nest, the male prods her at the base of the tail, an action that causes her to lay eggs. A fecund female may produce up to 800 eggs, although the average clutch is about 350 eggs. As soon as the eggs are laid, she leaves the nest. The male then quickly enters to fertilize the eggs. The entire process takes only about a minute. He then chases the female away and begins hunting for new mates. The male also guards the eggs and ventilates them with the movements of his pectoral fins. Without ventilation, the eggs will quickly suffocate. Males also remove any dead or diseased eggs they find by eating them. A male's other main responsibility is to defend the eggs against predators, which at Isle Verte are other sticklebacks.

After the eggs hatch (about a week after fertilization), the fathers guard the free-swimming fry. The males care for them for up to two weeks, chasing off any intruder that comes too close and swimming after any young that stray too far from the school. Errant fry are taken into the mouth and spit back into the school. The fry remain together in a school until they leave the pools at about 45 days of age.

Only about half to two thirds of the males that enter the marsh at Isle Verte are able to obtain a territory. The ability



REPRODUCTIVE BEHAVIORS of sticklebacks are diverse. Females at the Isle Verte marsh seem to choose a male based on the nest site, preferring nests hidden in cover (*a*). The male will dance in a zigzag fashion (*b*), enticing the female to lay eggs in his nest. Females try to increase their reproductive success by raiding nests and cannibalizing the eggs competitors may have laid (*c*). Males unsuccessful in establishing a nest site or in courting a female may resort to cuckholding (*d*). They drop their colors and sneak into another male's nest, inseminating the eggs present there.

to establish and defend a nesting area clearly contributes to a male's reproductive success. In many species, bigger males are better fighters, so we suspected that body size might play a role in competition for territories. Because fish keep growing throughout their lives, body size and age are closely correlated. We discovered that generally the larger, two-year-old males secure more of the territories than the smaller, one-year-olds.

Possessing territory, however, does not guarantee reproductive success. Some males get many more eggs than others. Most males receive only a single clutch, whereas some "superstuds" receive up to 10 clutches. Furthermore, I always found a sizable proportion of empty nests, indicating that no female had spawned with their owners.

Why are so many males unable to attract females to spawn in their nests? The answer depends on how females select mates. The picture of stickleback sexual behavior that I have drawn suggests that it is purely instinctive and automatic. This "textbook" account, while accurate for situations where a single male and female are put together in a tank, is incomplete. It implies that the sexes work in harmony to reproduce. It assumes that any healthy, sexually motivated male in nuptial coloration is attractive to the female and that any nonred fish with a swollen belly is attractive to the male. (Tinbergen had described how a male patrolling its territory would court another male if it were swollen with food and hence resembled a female.) The classic ethological viewpoint is that the stickleback responds to a "sign stimulus," that is, to a few characteristics of an object rather than to the object as a whole. The redness of the fish counts, not the fish itself.

But the stickleback is much more sophisticated than classic thinking suggests. Automatic responses simply would not be able to cope with the many complexities females face in choosing a mate. In a tide pool of 30 square meters, there may be 30 or more males nesting. The female must leave her precious eggs in the nest of only one of these males. Then it is he alone who will attempt to raise her young. A bad decision may result in the loss of her entire season's reproductive effort. Ideally, a female will choose a male that is likely to escape predation by birds, is healthy enough to undergo the rigors of parental care and is aggressive or skillful enough to defend her young from predation.

A female could base her choice on several possibilities—for instance, the size of the male, some aspect of the male's behavior or the nest site. Unlike the case for many other animals, size seems to matter little to female sticklebacks at Isle Verte. When Laval graduate students France Dufresne and Simon Lachance and I compared the mating success of the two groups by counting eggs in nests, we found no difference in mating success between large and small males. Laboratory work by Laval postdoctoral fellow Michèle de Fraipont also revealed that females showed no preference for size.

These observations contradict findings from other stickleback populations. For instance, William J. Rowland of Indiana University had found that females prefer to spend time near bigger models of male sticklebacks. Other studies found that in many populations, the nests of larger males generally harbored more eggs. Investigators are currently trying to determine the conditions under which females use size as a criterion in mate choice. At Isle Verte, they clearly base their selection on other factors.

Another criterion that females might use is the male's red nuptial coloration. Manfred Milinski and Theo C. M. Bakker of the University of Bern reported that female sticklebacks preferred the redder of two males when given a choice in the laboratory. Milinski and Bakker argued that females preferred the more brightly colored male because the intensity indicated that he harbored few parasites and was in better physical condition than the duller male. When Milinski and Bakker experimentally infected some brightly colored males with parasites in the laboratory, the fish became less intensely colored, lost body mass and consequently became less attractive to females.

The coloration may also correlate with the male's feeding abilities. The red comes from substances called carotenoids. Sticklebacks cannot manufacture these compounds; they must procure them from certain foods. If these foods are scarce, only the best foragers will be red. Thus, an intense nuptial coloration would indicate to the female that the male is in good physical condition.

But findings in the laboratory may not apply under natural conditions. We suspected that bright coloration might make the males in the tide pools unattractive to females. Brightly colored males could be more vulnerable than dull ones to bird predators such as herons and gulls, which are common at our site. Some years ago Frederick G. Whoriskey, now at McGill University, and I estimated that approximately 30 percent of all adult sticklebacks entering the marsh were killed by birds. Most of these victims were males.

To see if Isle Verte females preferred to spawn with the most brightly colored males, Laval graduate students Marc Fournier and Julie Morrissette and I evaluated the degree of nuptial coloration and then counted the numbers of eggs in the nests of 100 randomly chosen males. For the most extensively colored males—scarlet red from the tip of the snout to the anal pore—82 percent of the nests harbored eggs. Only 57 percent of the nests of the less colorful males had eggs. Furthermore, the bright males received about three times as many eggs as their duller rivals.

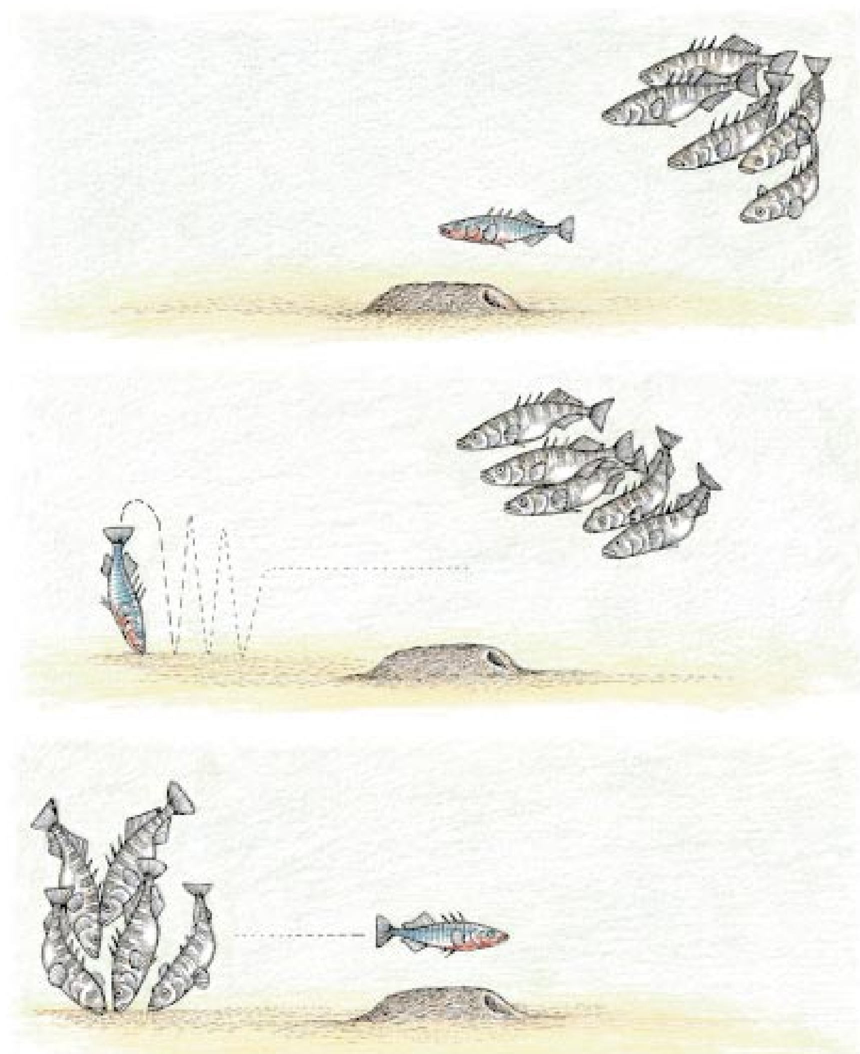
These field results seem to confirm the Swiss laboratory findings. But in behavioral studies, things are rarely as simple as they appear. Although the brightest males obtain the most eggs, the Isle Verte females apparently do not actually select these males because of their redness. In laboratory tests we discovered that, unlike the Swiss sticklebacks, Québec females showed no preference for the brighter of two males.

Another reason must exist to explain why the most brightly colored males procure the most eggs. Perhaps these males are those with the best nest sites. In the tide pools, some males nest in the open areas toward the center, whereas other males nest closer to the banks. These latter nests are often well hidden in a layer of filamentous green algae.

When we studied the characteristics of the nest site, we found that the best predictor of a male's mating success was the degree of cover around his nest—the more cover, the more eggs. We also found a strong positive correlation between a male's coloration and the amount of algal cover around his nest. Taken together, our laboratory and field results suggest that females prefer certain males because they nest in cover and not necessarily because they are more brightly colored than others. It is likely that males nesting in algal cover are less likely to be captured by birds than males that have nests in the open. Nests in cover are also more likely to go undetected by raiding sticklebacks. Support for this hypothesis came from R. Craig Sargent, then at the State University of New York at Stony Brook. He found that males with hidden nests mated more frequently and were more likely to hatch their eggs than males nesting in the open.

The hypothesis would also explain our previous laboratory finding that females did not discriminate based on color. The nest sites in the laboratory were similar, so females may have been indifferent to the males themselves.

If, for Isle Verte sticklebacks, nest sites are more important than physical characteristics, why should a male fish be red? Using sticklebacks from the Netherlands, Bakker showed a strong link between the brightness of the



DISTRACTION DISPLAY is a tactic a male may use to protect his nest from raiding females (*top*). The guardian male swims away from the nest, poking his snout into the ground several times (*middle*). This action resembles a feeding movement that lures the attacking females away from the nest (*bottom*).

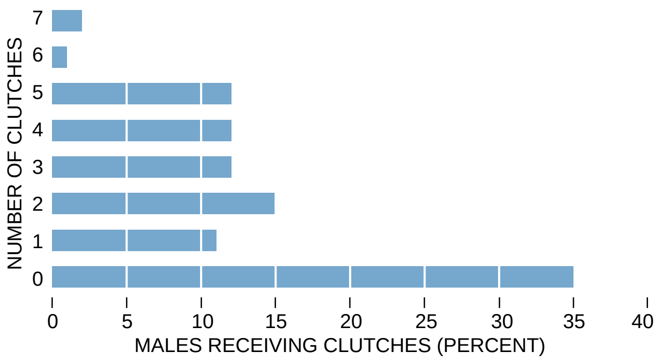
male’s nuptial coloration and his ability to dominate other males. It may be that the brightest males have an advantage in competition for the best nest sites at Isle Verte.

Although a female may choose to spawn in a given male’s nest, she actually has no control over which male fertilizes her eggs. A rival male may fertilize some or all of her eggs before the male that guards the nest does so. The interlopers are called “sneakers” because of their deceptive ways. When they see a territorial male courting a female, they drop their red coloration and become much drabber, more like a female. This metamorphosis enables them to approach the nest without eliciting the resident’s attacks. When close enough, the sneakers dart into the nest and release sperm over the eggs.

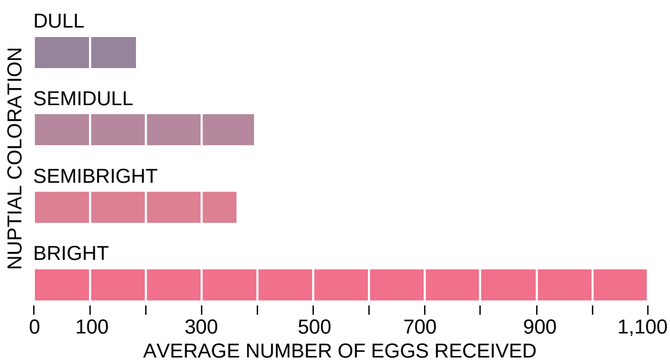
Using genetic testing, Urs Kühnlein of McGill, Laval graduate student Ciro Rico and I found that about 20 percent of the nests at Isle Verte contain eggs that had been fertilized by sneakers. In some cases, several males besides the guardian had fertilized the eggs.

Any male is a potential sneaker, but most sneaking is done by the younger, nonterritorial males. These males may also have a specialized adaptation for sneaking. Helga Guderley, also of Laval, de Fraipont and I recently discovered that younger males produce more spermatozoids and have more mobile sperm than the older males. Although we have no direct evidence, we think the sperm of a younger male may fertilize more eggs than those of an older male. Such sperm competition would occur if both males release sperm over the eggs within a short period.

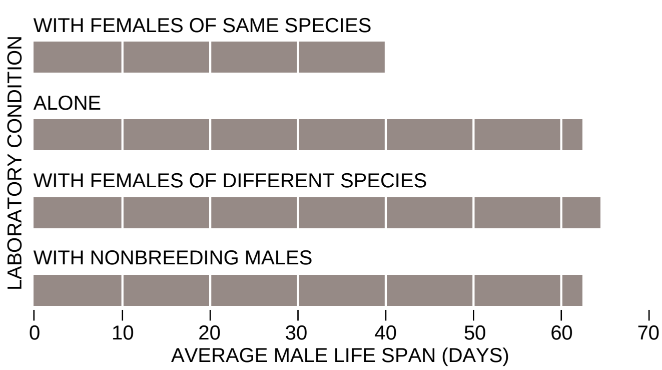
Another way sticklebacks increase their reproductive prowess is by eating



MATING SUCCESS varies greatly among males that have built nests. More than one third did not receive any eggs; others received many clutches. An average clutch is about 350 eggs.



NUPTIAL COLORATION may aid in courting females. The most brightly colored males received substantially more eggs from females than did their duller rivals.



FEMME FATALE EFFECT refers to the early death of males housed with females of their own species. Males survive longer if they are alone or with other types of fish.

the eggs of competitors. Since the early 1950s, fish ecologists have reported finding stickleback eggs in the stomachs of male sticklebacks. Cannibalism had been explained by the fact that males had fewer foraging opportunities than females did because of the need to remain near the young. Consequently, the males would eat some of their own eggs or steal them from rival males.

Yet the sticklebacks at Isle Verte presented a new conundrum. The females there are ferocious egg cannibals. In groups ranging from two to several hundred, they seek out nests and destroy all the eggs therein. We were perplexed to find so many cannibals in the tide pools because alternative food supplies—zooplankton and benthic invertebrates—are abundant. In addition, female three-spine sticklebacks generally ignored the eggs of the other stickleback species, a fact that piqued my curiosity.

I suspected that cannibalism, at least by the females, must occur for a reason other than nutrition. Of course, females do obtain a nourishing meal. But an alternative evolutionary hypothesis occurred to me after reading a study of infanticide in lions. After taking over a pride, a male lion kills any cubs he finds. The loss of cubs quickly brings the lionesses into heat, allowing the new male to impregnate them.

I reasoned that something similar might be occurring in the tide pools. Because male sticklebacks are the sole caregivers and because so many males appeared unacceptable to the females, I suspected that female sticklebacks might be competing for high-quality males. Whoriskey and I collaborated with William L. Vickery of the University of Québec to develop a mathematical model to explain female cannibalism. We assumed that the females kill a male's young in order to spawn with him at a later time. Any nutritional advantages are secondary.

Our model had two main predictions. First, females should spawn in nests with few eggs rather than in nests with many eggs. This hypothesis was based on work by Stephane Reebs, then an undergraduate student in my laboratory, who found that eggs laid in crowded nests had a low hatching rate. Recently my graduate student Jean-Claude Belles-Isles confirmed this finding. In the laboratory, he gave females a choice of spawning in nests containing one clutch of eggs or several clutches of eggs. They preferred to spawn in the nests with the fewest eggs.

A second prediction of the model was that females should not eat their own eggs and that they should spawn with males whose nests they had previously raided. This was verified by Nathalie van Havre and me in laboratory experiments. We also found that those females that had initiated the raids were the ones most likely to spawn when the males rebuilt their nests. Although several details of the model still need to be worked out, overall it gives a good account of why female three-spine sticklebacks are so cannibalistic.

Given the great risk presented by females to the eggs, we wondered whether the males have countertactics to minimize their egg losses. Indeed, they do. Males can usually repel attacks by one or even several females by vigorous bites, chases and threatening displays. But when females attack in larger groups, the male must resort to more subtle means to save his eggs. Like some species of birds that pretend to have broken wings to lure predators away from their nests, male three-spine sticklebacks use a decoy display. The male tricks the raiders away from his nest by "promises" of food elsewhere. He rolls on his side, swims up into the water column and out of his territory. He then pokes his snout into the substrate well away from his nest. This erratic swimming and snout poking closely resembles a feeding movement and is often sufficient to lure the raiders away from the nest.

Another tactic males employ to protect their eggs is to bury the nests with sand during the daytime. Nest burying, however, has a cost: the eggs may suffocate. Thus far we have seen nest burying only in the laboratory, where the sand is usually not as tightly packed as it is in the field. The behavior may be limited to populations with coarse substrates.

Males may also minimize the risk to any eggs by practicing "safe sex"—that is, courting mates only under a specific condition. Experiments by Belles-Isles, Dane Cloutier and I showed that males court solitary females. They remain immobile when groups of females are cruising near their nests. Not surprisingly, only the males that take mating precautions hatch any eggs.

Because they raid nests, female sticklebacks represent the greatest source of stress for the males (apart from birds). Males never seem to become accustomed, or habituated, to the presence of females. Except when engaging in sexual behavior, males always attack any female that crosses their territorial boundary. In contrast, territorial males rapidly habituate to the presence of males in adjacent territories. This tolerance toward known enemies has been called the "dear enemy effect" by Harvard University sociobiologist E. O. Wilson.

The stress has a so-called femme fatale effect on the male. The mere presence of females seems to shorten the life span of the male. De Fraipont, Guderley and I found that males kept in view of females through a glass partition died much sooner than males kept alone or with other stickleback species.

To those of us who spend warm spring days studying sticklebacks, these alluring animals always present a source of immense pleasure. But I am also asked whether sticklebacks have any practical importance. My usual answer is that they play a key role in the ecosystem because of their great abundance. They are an important forage food for bird and fish predators.

Not everyone shares my love of sticklebacks. Recently some concerned sport fishery managers complained that their lakes were "infested" with an introduced population of sticklebacks. They asked me to conduct a study to see whether the sticklebacks were competing for food with the indigenous brook char, an important sport fish. Although plentiful, the sticklebacks were feeding on a different prey. In fact, the sticklebacks were a food source for the char. The problem for fishermen is that the char are not responsive to lures if sticklebacks are abundant.

Regardless of their economic role, sticklebacks are exquisite models for studies of the adaptive significance of behavior. Still, many questions about their biology remain. Do territorial males have tactics to minimize sneaking? Does it matter to the females who really fertilizes their eggs, as long as a good parent takes care of them? What physiological mechanism causes the male's shortened life span in the femme fatale effect? These little fish will undoubtedly continue to fascinate biologists for many years to come.

FURTHER READING

THE ROLE OF CANNIBALISM IN THE REPRODUCTIVE ECOLOGY OF THE THREESPINE STICKLEBACK. G. J. FitzGerald in *Ethology*, Vol. 89, No. 3, pages 177–194; November 1991.

BEHAVIOURAL ECOLOGY OF STICKLEBACKS. G. J. FitzGerald and R. J. Wooton in *Behaviour of Teleost Fishes*. Edited by T. J. Pitcher. Chapman and Hall, 1992.

ECOLOGY OF THE THREESPINE STICKLEBACK ON THE BREEDING GROUNDS. F. G. Whoriskey and G. J. FitzGerald in *The Evolutionary Biology of the Threespine Stickleback*. Edited by M. A. Bell and S. A. Foster. Oxford University Press (in press).

The Evolution of Virulence

Human behavior appears to influence whether pathogens evolve into benign or harmful forms. Health policy should therefore include evolutionary considerations

by Paul W. Ewald

Some pathogens, such as those that cause cholera, smallpox, tuberculosis, malaria and AIDS, have quite severe effects. Others rarely inflict any damage beyond a cold or a sore throat. Recent studies suggest there are several evolutionary reasons for these varying levels of virulence. They include a pathogen's mode of transmission as well as its ability to survive outside a host organism for long periods. Human behavior may also play a significant, largely unrecognized role in the evolution of pathogens because it often determines the route and timing of transmission.

Understanding the forces that shape changes in virulence could become a powerful tool for medicine. By examining these variables, evolutionary biologists have already been able to predict patterns of morbidity and mortality in several diseases—including cholera, dysentery and AIDS. Using such an approach, medical scientists may be able to anticipate alternative evolutionary courses of pathogens and to tailor treatment and social behavior accordingly. They might even be able to transform virulent adversaries into mild ones.

Until recently, the understanding of how virulence evolves has generally been limited to one view. Most physicians and medical writers have concluded, unjustifiably, that the evolution of host-parasite relations ultimately leads to benign coexistence. Their opinion is based on the idea that parasites that do not harm their hosts have the best long-term chance of survival: they thrive because their hosts thrive.

Some biologists, however, have arrived at a different conclusion. Evolutionary theory holds that what is best for a species may differ from what is best for its component individuals and that what is best for these individuals is defined by which genes are passed along most successfully. Even if a pathogen reproduces so extensively that it causes its host to become gravely sick, its host-impairing instructions may still win out over the less damaging instructions of a less aggressive competitor. The more virulent pathogen would achieve this success if its increased replication led to a level of transmission into new hosts that exceeded the loss of transmission resulting from the host's illness or death. In this scenario, the extinction of the host may eventually cause the demise of any pathogen that relies exclusively on one host. But the possibility of extinction does not inhibit the spread of the more virulent genes within the population of pathogens.

This perspective suggests that a parasite's virulence may reflect its mode of transmission. If the illness of a host impairs transmission, evolutionary biologists would predict that parasites would evolve to have milder effects. In contrast, if the host's disability does not inhibit transmission, pathogens could gain a competitive advantage by reproducing more rapidly.

Consider, for instance, a pathogen that relies on the mobility of its host. Rhinoviruses, causes of the common cold, reproduce in the cells that line the nasal passages. The viruses are shed in nasal secretions that trickle out as a runny nose or blast out during a sneeze. A person's finger may wipe away the mucus and may then touch the fingers of another person in the course of a handshake or by way of a borrowed pencil. The exposed individual may later touch fingers to nose or inhale the contaminated air, planting rhinoviruses on fertile ground.

Whichever route is taken, the ability of the host to move is critical. If the pathogen reproduces so extensively that the host is too unwell to leave home, thousands of rhinoviruses shed that day will die of exposure. A few might squeak by if they infect a family member, but such transmission could hardly be considered a great accomplishment, especially if the new host also becomes bedridden. For rhinoviruses, a person's movement improves prospects for transmission. Accordingly, rhinovirus replication is restricted to cells scattered in a sea of uninfected nasal mucosa, and virulence remains slight.

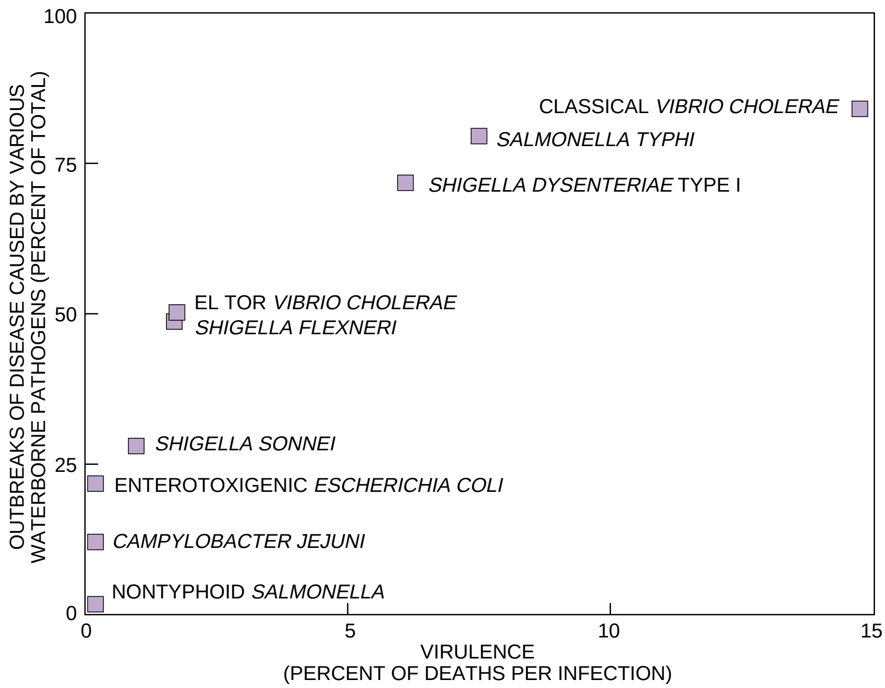
PAUL W. EWALD is associate professor and chair of the biology department at Amherst College. In 1980 he received his Ph.D. in zoology from the University of Washington, where he studied the evolutionary and ecological aspects of aggression in birds. Two years ago the Smithsonian Institution named Ewald the first George E. Burch Fellow of Theoretical Medicine and Affiliated Science. When not researching evolutionary epidemiology, Ewald continues to investigate animal behavior. He also works as an unpaid carpenter with the hope of nursing his colonial-period farmhouse into its second quarter-millennium.

UNPROTECTED WATERS, including wells (*above*) and rivers, such as the Yamuna in India (*right*), have been infamous disseminators of disease. Purification limits the spread of infection. It also may provide an unrecognized benefit: it may force pathogens to evolve into mild forms.

The opposite can occur when pathogens are transported by a vector, an organism that transmits an infectious agent. If a mosquito-borne pathogen reproduces so extensively that the host becomes immobilized, the pathogen can still pass along its genes. The vector transports pathogens from the incapacitated host to susceptible individuals. Transmission can even be facilitated when hosts are sick because they are then less able to prevent mosquitoes from biting. (Pathogens tend to treat their vectors kindly: a delirious mosquito could not effectively shuttle malaria-causing organisms to new hosts.)

Extensive multiplication and distribution in the vertebrate host are beneficial to those pathogens traveling by vector. Such agents tend to reach higher densities than less noxious pathogens do, increasing the probability that a biting insect will obtain an infective dose. Without systemic spread, a mosquito or another carrier would pick up the pathogen only if it bit near the site of the bite that originally infected the host.

Medical literature provides ample evidence of these trends. Pathogens that are transmitted by biting arthropods



PATHOGENS that are most frequently transported by water systems, such as classical *Vibrio cholerae,* are more virulent than those that are not, such as nontyphoid *Salmonella.* A waterway contaminated by wash or sewage can serve as a cultural vector, a transporter of pathogens that is created by human behavior.

are especially damaging to humans: they tend to cause death more frequently than do pathogens that are transmitted directly [*see illustration on page 90*]. The severity of diseases such as malaria, yellow fever, typhus and sleeping sickness can thus be explained as an evolutionary consequence of vector-borne transmission.

But there are exceptions to this rule. Some directly transmitted pathogens—including the smallpox virus and the tuberculosis bacterium—can often be lethal. One possible explanation for their virulence is the sit-and-wait hypothesis. A pathogen can be passed from an incapacitated host to a susceptible host in two ways. As already described, it can be transported by something that travels, such as an insect. Alternatively, if the pathogen is able to survive in the external environment for an extended time, it can just sit and wait for a host to happen along. Like their vector-borne counterparts, such enduring pathogens benefit from extensive multiplication inside their hosts, while losing little from the ensuing immobilization of their hosts. According to this argument, sit-and-wait pathogens would be particularly virulent.

Indeed, pathogens that are able to survive for long periods in the external environment are often more injurious. This trend was recently documented in a study of human respiratory pathogens I conducted with one of my students, Bruno Walther. Infectious agents that survive for weeks or years in the environment tend to cause death most frequently: for example, smallpox kills one in 10 people, and the virus that causes it can persist for more than a decade outside a host. The pathogens giving rise to tuberculosis and diphtheria can also live externally for weeks to months and are correspondingly noxious. Most other pathogens tend to survive for only hours or days without hosts; they tend to be less destructive, causing death in fewer than one in 10,000 infected people.

Human behavior may also affect virulence. As the emergence of antibiotic resistance shows, pathogens can evolve quickly in response to human interference. But modification of pathogens by people is not limited to the introduction of medical treatments. Some activities create cultural vectors—amalgams of culture, human behavior and the physical environment that allow pathogens to be transmitted from immobilized, infectious hosts to susceptible individuals.

For instance, the virulence of diarrheal pathogens should reflect their tendency to be effectively transmitted through drinking-water systems. Like pathogens transmitted by insects, those spread by water may have little to lose from the immobility of their hosts. A sick person can release infectious agents onto bed sheets or clothing that will be washed. If the contaminated washing water can reach unprotected drinking water, even an immobilized person with diarrhea can infect many others. I conducted a review of the literature on bacterial diarrheal diseases and found just such a correlation between waterborne transmission and mortality. The evidence supports the idea that such transmission leads to an increase in virulence. The severity of typhoid fever, cholera and the most dangerous forms of dysentery can therefore be explained as the evolutionary result of waterborne transmission.

It follows, then, that if transmission by water confers an advantage to highly virulent bacteria, the introduction of uncontaminated drinking supplies should reduce virulence, because pathogens would not gain much benefit by reproducing rapidly. In fact, this trend has been documented. As water supplies were purified in cholera-endemic regions in India during the 1950s and 1960s, the milder agent of cholera, the El Tor type of *Vibrio cholerae,* displaced the more dangerous form, classical *V. cholerae*. In Bangladesh, however, where the war of independence and extensive socioeconomic difficulties have delayed water purification, classical *V. cholerae* persists.

The three causes of bacillary dysentery have experienced similar transitions. In all countries with detailed records, the most virulent bacteria, *Shigella dysenteriae,* virtually disappeared as major improvements in water purification were regionally enacted, while the moderately virulent *S. flexneri* predominated. As these improvements progressed toward complete purification, *S. flexneri* was replaced by the mildest species, *S. sonnei*. In the same regions the agent of typhoid fever, *Salmonella typhi,* has also been replaced by the less debilitating species of *Salmonella*.

Hospitals may generate another form of cultural vector. Attendants inadvertently carry pathogens on their hands from patient to patient. They do so by direct contact or indirectly through the use of equipment. As with transmission by arthropods or water, attendant-borne transmission could promote virulent genotypes of pathogens—again because extensive replication in the infected host would enhance the infection of new hosts.

In such settings, newborn infants would be at risk of being infected by attendants. From the pathogen's point of view, a neonate is very different from a nurse. Neonates are highly vulnerable because they lack the acquired immunity of adults. A nurse typically touches each baby about 20 times a day and at-tends very ill babies more frequently. Intense diarrhea should facilitate transmission by attendants given that it is easier to avoid a well-formed stool than a diffuse fecal film. In addition, hospital strains of diarrhea-causing pathogens may persist and sometimes multiply, even after standard washing with disinfectants. Contaminated hands may accidentally touch objects before hand washing; these contaminated items can then recontaminate hands after washing.

Reports describing outbreaks of a common inhabitant of the human intestine, *Escherichia coli,* support this argument. When outbreaks of pathogenic *E. coli* were over in a week or so, few infected babies died. But in places where the incidents continued for weeks to months, about one of every 10 infected babies died [*see illustration on opposite page*]. At first glance, the association between lethality and the duration of outbreak may appear to be the result of poor hygiene, not of evolutionary forces. Hospitals that put less effort into curbing outbreaks might put less effort into cleanliness and therefore might have had more severe infections because patients received larger doses of the pathogen.

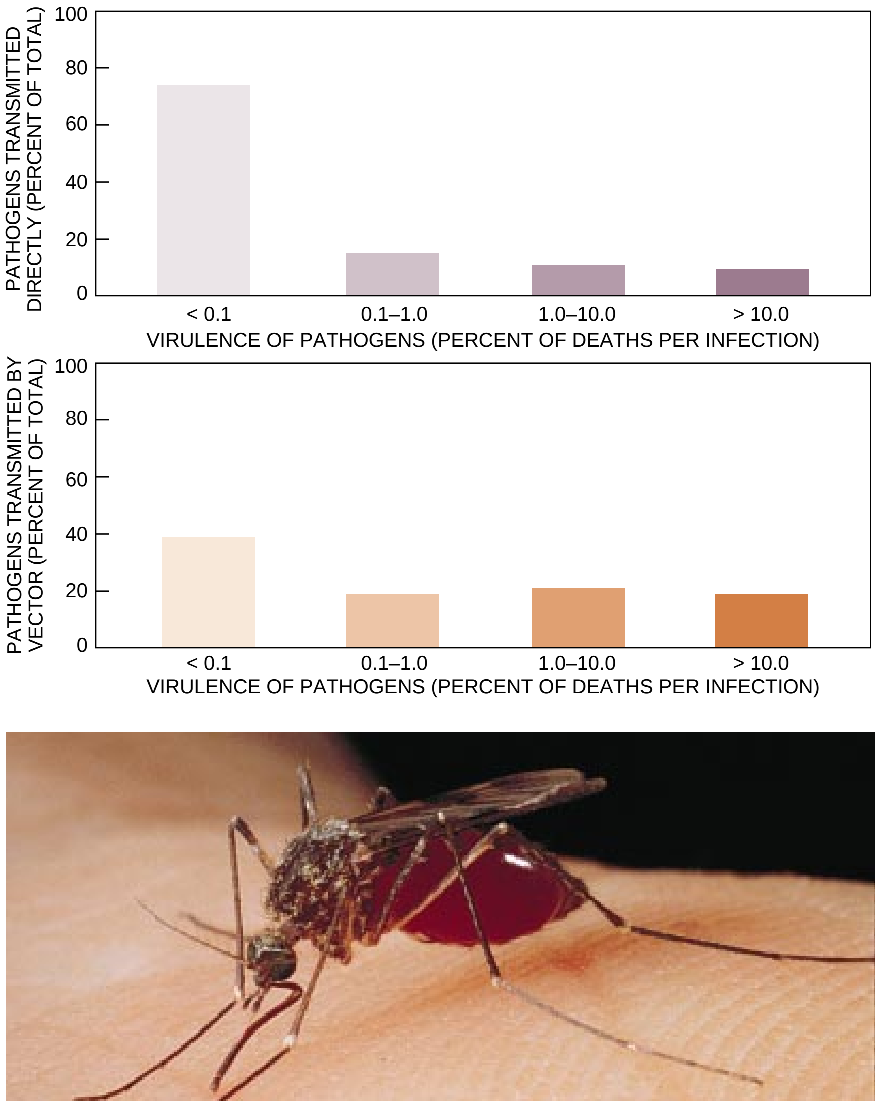
The details of the outbreaks, however, are inconsistent with such an interpretation. To quell the attacks of *E. coli* and other bacteria, hospital staff usually extensively disinfected the facility and the hands of attendants. During these campaigns, it stands to reason that the doses of *E. coli* must have been drastically reduced. Yet damaging infections persisted. The ability of pathogens to cause severe illness in the presence of such improvements in sanitation suggests that their virulence, rather than increased contamination, was to blame. To extinguish such ongoing outbreaks, the staff frequently closed the affected wards and doused them with disinfectants.

About one of every 20 hospitalized patients in the U.S. acquires an infection before leaving the hospital. The annual total approaches four million infections and thousands of deaths. Despite the magnitude of this problem, the possibility that attendant-borne transmission heightens virulence remains untested—except in the case of *E. coli.* Long-term studies that monitor frequencies of infection as attendant-borne transmission is suppressed need to be broadened to determine whether this intervention reduces virulence.

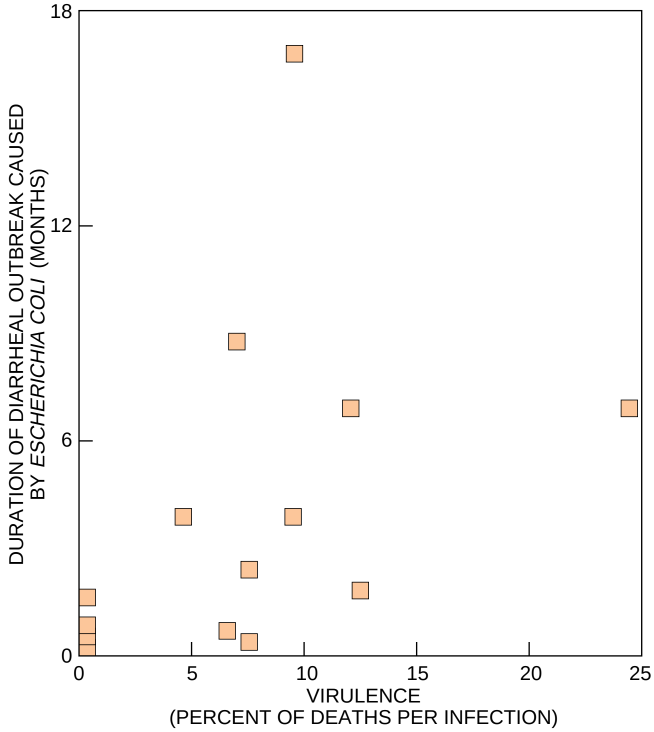
If an evolutionary understanding of virulence could help resolve vexing medical problems, its application to HIV would be extremely timely. Some 10 million people are infected with HIV, and one million have symptoms of AIDS or have already succumbed to it. By examining the conditions that influence the competitive success of sexually transmitted pathogens, we can begin to clarify the influences determining the virulence of HIV.

Consider a sexually transmitted pathogen in a population of relatively monogamous individuals. After it infects a couple, the pathogen would have to remain viable until one member engaged in sexual activity outside the pair. If such activity took place only, say, once every three years, the typical duration of infectiousness of directly transmitted pathogens would not allow the agent to survive—as the example of rhinoviruses suggests. Only the variants that had some way of extending their infectiousness would be transmitted venereally.

To extend infectiousness, pathogens must avoid being destroyed by the host's immune system. At the same time, however, they must maintain access to new hosts. One option is to remain latent inside long-lived cells. By merging with a chromosome in a host cell and by suppressing the creation of products that would trigger an immune response, a latent virus can wait out the time between new sexual partners. Indeed, retroviruses such as HIV can hide in long-lived white blood cells and be transmitted years after the initial infection. (Retroviruses operate by inserting their RNA into host cells; the RNA is then transcribed into DNA, which is incorporated into the host's genome.)



MODE OF TRANSMISSION reflects a pathogen's virulence. If the pathogen moves directly from host to host (*top*), its effects must be mild enough that they do not disable the host: if the host cannot move, the pathogen cannot be transmitted. But pathogens carried by a vector, such as a mosquito (*bottom*), need not spare the host: the vector can transmit the pathogen regardless of the host's state. Thus, vector-borne pathogens tend to be quite virulent (*middle*).



HOSPITAL ATTENDANTS, who may touch many babies (*left*), can inadvertently serve as vectors. As the chart shows, *Escherichia coli,* a diarrhea-causing organism, is particularly lethal when it is repeatedly transmitted in a hospital (*right*).

Of course, the benefit of latency would have to be weighed against the higher rate of reproduction that would occur in the absence of latency. HIV has the raw material for both strategies. It remains latent in most of the cells that are infected at any given time, but throughout each infection some actively reproducing viruses tend to be present. The rapid rate at which HIV mutates results in a great potential for quick changes in replication rates. In fact, nonlatent HIV regularly evolves an increased replication rate in the interval between the onset of infection and death.

If sexual activity increases so that people are involved with a different partner every week, the factors contributing to virulence shift. A sexually transmitted virus that actively reproduces in an infected person could soon infect several new hosts, whereas a virus that is latent for three years would propagate less extensively. Thus, increased rates of partner change could favor HIV variants that replicate rapidly. The pressures favoring the evolution of fast viral reproduction when sexual contact is high and of delayed viral reproduction when sexual contact is low provide the basis for a prediction: HIV virulence should be correlated with rates of sexual contact.

Data gathered over the past decade are consistent with this prediction. One form of evidence comes from the two types of HIV: HIV-1, the more rapidly lethal form, and HIV-2, which is associated with a longer symptom-free period and with lower frequencies of full-blown AIDS. During the early years of the AIDS pandemic, HIV-1 tended to occur in Central and East Africa and HIV-2 in West Africa. Interestingly, these different forms of infection reflect different social patterns. During the 1960s and 1970s, an economic crisis in Central and East Africa caused a migration away from rural areas. Men left to obtain industrial jobs in cities; large populations of men without families created, in turn, a market for sexual commerce and drew women into these urban areas.

In contrast, West African countries generally did not undergo massive movements of people or increases in sexual contact during the same period. According to the evolutionary theory of virulence, the virus endemic to West African countries (HIV-2) should not be as lethal as that in Central and East Africa (HIV-1). And such is, in fact, the case.

Differences in virulence can also be predicted within these two types of HIV. Although data are scarce, one difference seems to be emerging in the two most extensively studied areas of HIV-2 infection: Senegal and Ivory Coast. In Senegal, HIV-2 infections are associated with a low frequency of AIDS. Since 1985, Phyllis J. Kanki and her colleagues at Harvard University have followed a large group of HIV-2-infected prostitutes in Senegal to determine the speed at which HIV-2 progresses to AIDS. To date, only one of the women has shown symptoms of AIDS. In addition, during early attempts to isolate HIV-2 from infected Senegalese subjects, Lilly Kong, formerly at the University of Alabama, and her colleagues had little success. Apparently the virus entered cells and propagated at a reduced rate. Kong's one isolate had very mild effects.

Studies conducted in the capital of Ivory Coast, Abidjan, reveal a different trend. This city underwent social change when migrations from rural areas brought influxes of single men. The strains of HIV-2 common to Abidjan appear to be more virulent than those found in Senegal.

Although accurate information on rates of sexual interaction is exceptionally difficult to obtain, sociological studies suggest that Senegal did not experience the same kind of social disruptions found in Abidjan. Olga F. Linares of the Smithsonian Tropical Research Institute believes several cultural factors may have favored reduced partner change in Senegal: a family structure that encompasses rural and urban residences; an intact, traditional agricultural infrastructure; and an Islamic heritage that discourages premarital and extramarital sex.

Rates of sexual transmission may partly explain the evolution toward greater virulence, but the argument must be broadened to describe why HIV has become so deadly. HIV does not kill its host directly: it destroys the immune system, permitting lethal secondary infections to occur. The evolution of this characteristic apparently arose from the virus's predilection for cells critical to immunologic defenses against other pathogens, its ability to replicate quickly and its tendency to alter its genetic material at an accelerated pace.

When partner changes are infrequent and viruses are latent for long periods, a predilection for immune system cells need not destroy the host. HIV's retroviral cousin, human *T* cell leukemia/lymphoma virus type 1 (HTLV-1), for example, infects the same class of lymphocytes that HIV infects. But HTLV-1 has a suppressed level of reproduction and a less detrimental effect on the immune system in most people.

If sexual contact increases, selection may favor greater rates of replication even though such evolution may lead to an earlier death for the host and hence to the demise of any form of HIV that the person harbors. In this case, the virus makes the best of a bad situation. For a sexually transmitted virus infecting a host with frequent partner changes, immune cells are not ideal targets, because destroying the host's immune system will ensure quick death. But once a virus's biochemical machinery has become specialized to enter and reproduce in lymphocytes under conditions of low sexual contact, altering its biochemistry to enter some other less critical cell type could pose a major hurdle.

High rates of mutation, another characteristic of HIV, present additional long-term drawbacks for a virus. Frequent mutations are often costly for organisms because they tend to mangle finely tuned biochemical machinery. For HIV, increased mutation may also lead to the collapse of the host's immune system and to symptoms of AIDS.

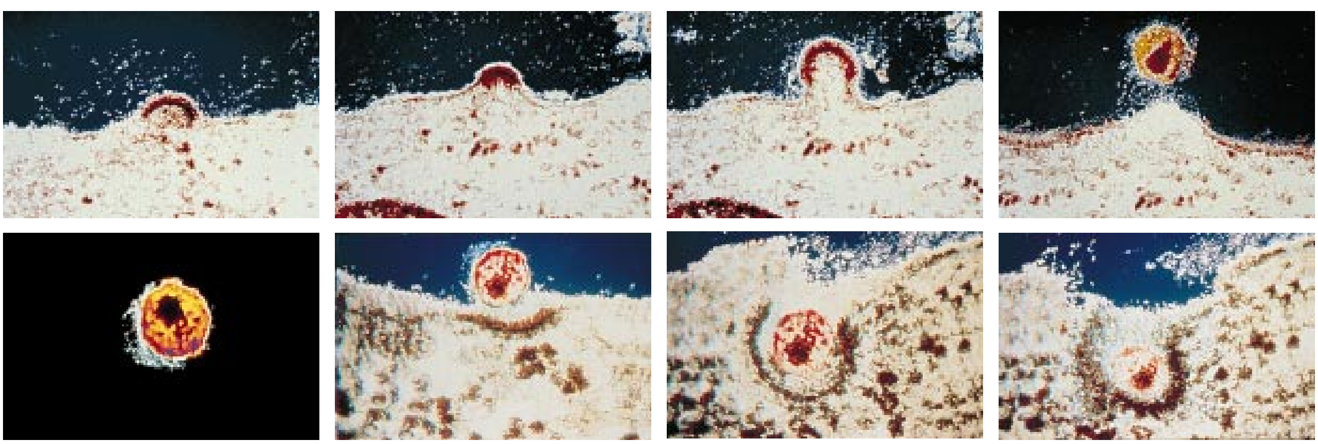
Mutations, however, may provide compensating benefits. If frequent partner changes confer an advantage on variants that have reduced latency, natural selection should favor any viruses that look different to the immune system over viruses that have already reproduced in the person. If a virus does not look different, then the immune system—primed by its previous encounter with that variant—will do it in. (This generalization breaks down during the terminal stages of HIV infection. By then, though, the point is moot because HIV is generally eliminated from evolutionary competition by the behavioral avoidance of unprotected sex with someone who is very ill.)

Being different should therefore encourage evolutionary plasticity in the proteins that protrude from the virus and are exposed to the immune system. Indeed, genes that contain the instructions for the production of these envelope proteins are more prone to mutation than is the rest of the HIV genome, and they are among the most mutation prone of any known genes in any organism. This mutability is not common to all retroviruses. Retroviruses show great variations in mutation rates in envelope proteins. HIV-1 lies at one extreme: its genes can diverge by some 5 percent in a few years. At the other end of the spectrum is HTLV-1, which may require centuries to achieve the same degree of change. Ultimately, then, a high incidence of sexual contact may explain why HIV is so dangerous. When the potential for sexual transmission increases, natural selection favors viruses with reduced latency, more rapid replication among nonlatent viruses and faster mutation rates.

The same kinds of processes should, in theory, be shaping the evolution of HTLV-1. The most detailed studies come from Japan and Jamaica, where data indicate that HTLV-1 is also variable in its virulence and that this variability is associated with rates of unprotected sexual contact with different partners. In Japan, barrier contraceptives, which inhibit retrovirus transmission, are used widely. (Birth-control pills, which do not prevent sexually transmitted disease, are restricted.) Moreover, rates of sexual partner change appear relatively low there. Accordingly, the geographic and age distributions of disease in Japan illustrate that a large proportion of infections arise from transmission from mother to child rather than from sexual contact.

In Jamaica, the analogous distributions indicate that sexual contact is the more common route of transmission. The Japanese-Jamaica comparison suggests that HTLV-1 may be more virulent in areas where rates of sexual transmission are high. Indeed, in Japan, cancers triggered by HTLV-1 occur, on average, at age 60; in Jamaica, they occur at roughly age 45. This comparison, like the geographic comparisons of HIV infection, draws attention to the need for more precise measurements within many geographic regions to determine whether retroviruses are inherently more virulent in areas with greater potential for sexual transmission.

Understanding the epidemiology of infectious diseases from an evolutionary point of view permits many assessments—including why virulence evolved in the past, what is currently maintaining it, how it may evolve in the future and, most impor-



HUMAN IMMUNODEFICIENCY VIRUS attacks the often long-lived cells of the immune system. These photographs show a newly replicated virus budding from a *T* lymphocyte, wrapping itself in the cell's membrane (*top row*). The virus travels to another *T* lymphocyte and enters when the membrane surrounding the virus fuses with that of the cell it infects (*bottom row*).

tant, how human activities can influence this evolution. If rates of infecting sexual contact decline, the virulence of HIV and HTLV-1 should drop accordingly.

The rates of unprotected sexual encounters among urban male homosexuals began to decline sharply around 1984, in response to improved knowledge about transmission of HIV. Considering that highly virulent strains of HIV generally require two to three years before they cause AIDS, any evolutionary slowing of progression to AIDS would not be detectable before the late 1980s. In addition, detection of such an effect during the late 1980s is confounded by the effects of zidovudine (AZT) treatment.

Philip S. Rosenberg and his co-workers at the National Cancer Institute have published an analysis that accounts for the life-extending effects of AZT. They have found that the growing lag between HIV infection and AIDS that was observed among male homosexuals before 1988 could be explained entirely by AZT treatment. After mid-1988, however, there was an additional lengthening of time between infection and disease among male homosexuals that could not be so explained. This unexpected delay may be the first sign of a decline in HIV virulence. It is noteworthy that infections in intravenous-drug users did not lead to significant reductions in high-risk behavior during the mid-1980s. This population also did not experience a lengthening of time before the onset of AIDS that was independent of AZT.

The next few years of data should help resolve the alternative explanations for this delay in the onset of disease, because nonevolutionary alternatives do not predict a continued lengthening. In contrast, evolutionary theory postulates a growing period between infection and AIDS in groups whose high-risk behavior has declined. More generally, the evolutionary explanation and the variations in virulence over time and geographic area suggest that an unplanned experiment is now occurring. People reducing their transmission rates fall into the experimental group. The populations in which transmission rates are not being reduced are the controls. HIV in the latter group is expected to remain more virulent than HIV in the former groups.

CONDOM USE, something these demonstrators at an AIDS awareness rally are urging, is associated with less virulent forms of HTLV-1 in Japan. Because the use of condoms reduces opportunities for transmission, the virus must be able to survive for a long time in its host. Thus, it is advantageous for the virus to be mild and long-lived so that it does not impair its host.

These trends potentially provide a basis for changing the course of virulence. By increasing the price that pathogens pay for virulence, we should be able to make pathogens evolve toward a benign state. In the case of HIV and HTLV-1, increased investment in interventions that reduce the rates of transmission by sexual contact and needle sharing should force HIV and HTLV-1 to become less damaging over time—or at least keep them from becoming increasingly virulent in each subsequent generation. Reducing attendant-borne transmission in hospitals by improving hand-washing practices, by using gloves more effectively and by increasing maternal contact with babies (since mothers tend to touch only their own newborns) should similarly inhibit evolution toward more virulent strains of hospital-acquired organisms. Further improvements in water purification should make the diarrheal pathogens in a region milder. Making houses and hospitals mosquito-proof should make the mosquito-borne diseases evolve toward a more benign state as well.

By identifying the long-term benefits of interventions, an evolutionary approach to disease control adds a novel dimension to policy making. As our knowledge in this area matures, it should provide a clearer view of which interventions are most cost-effective. We have already controlled those pathogens most readily affected by traditional interventions such as vaccination and antibiotic treatment. Pathogens like HIV and the protozoa that cause malaria remain unconquered.

In the absence of an evolutionary view, the future will repeat the past. When we impede versatile pathogens with a drug, they will evolve resistance. A pathogen's potential for change will undoubtedly lie beyond our ability to anticipate its mutability. It is for such more formidable pathogens that we desperately need a new perspective. Virulent pathogens may be transformed into mild ones not because benign coexistence is the inevitable end point of evolution but because we will have made it the most favorable outcome.

FURTHER READING

TRANSMISSION MODES AND THE EVOLUTION OF VIRULENCE: WITH SPECIAL REFERENCE TO CHOLERA, INFLUENZA, AND AIDS. P. W. Ewald in *Human Nature,* Vol. 2, No. 1, pages 1–30; 1991.

WATERBORNE TRANSMISSION AND THE EVOLUTION OF VIRULENCE AMONG GASTROINTESTINAL BACTERIA. P. W. Ewald in *Epidemiology and Infection,* Vol. 106, No. 1, pages 83–119; February 1991.

THE DAWN OF DARWINIAN MEDICINE. George C. Williams and Randolph M. Nesse in *Quarterly Review of Biology,* Vol. 66, No. 1, pages 1–22; March 1991.

EVOLUTION OF INFECTIOUS DISEASE. P. W. Ewald. Oxford University Press, 1993.

Modern Humans in the Levant

Modern Homo sapiens *preceded Neanderthals on Mount Carmel and followed a similar pattern of life for 60,000 years. Biology thus cannot explain the cultural revolution that then ensued*

by Ofer Bar-Yosef and Bernard Vandermeersch

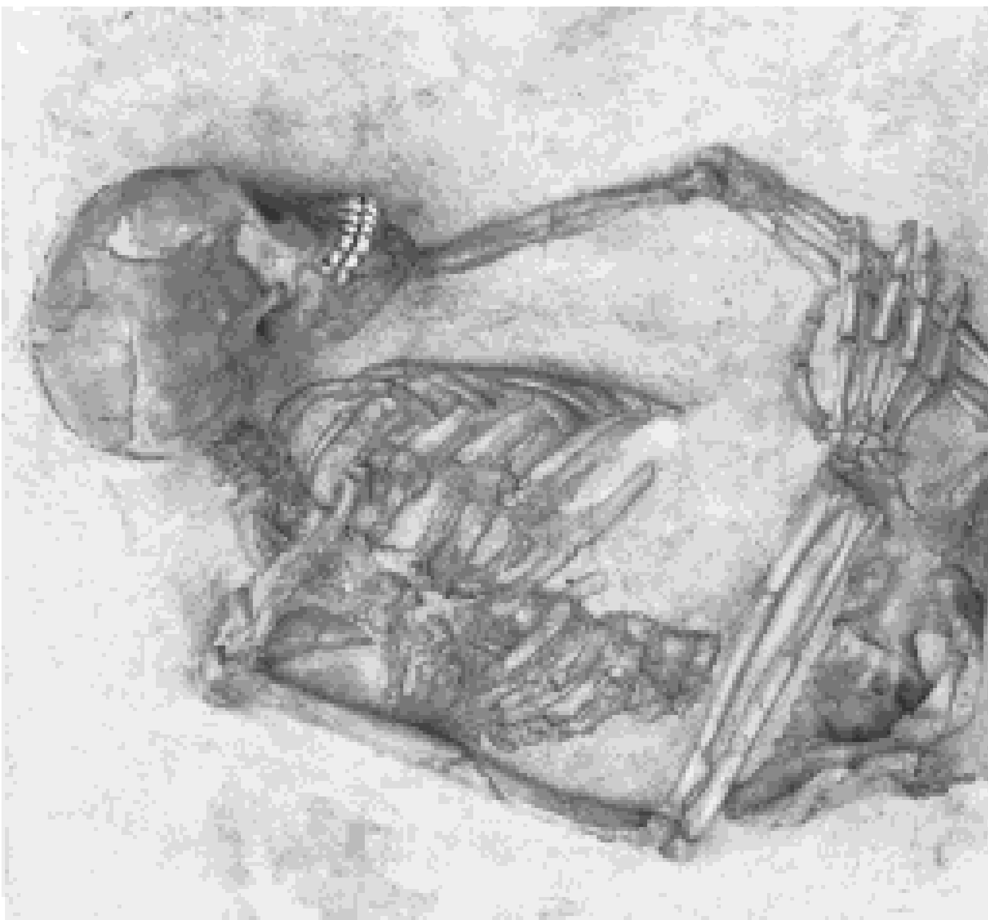
Between 40,000 and 45,000 years ago the material culture of western Eurasia changed more than it had during the previous million years. This efflorescence of technological and artistic creativity signifies the emergence of the first culture that observers today would recognize as distinctly human, marked as it was by unceasing invention and variety. During that brief period of 5,000 or so years, the stone tool kit, unchanged in its essential form for ages, suddenly began to differentiate wildly from century to century and from region to region.

In technical terms, the cultural revolution marked the transition from the Middle to the Upper Paleolithic ages. Why it happened and why it happened when it did constitute two of the greatest outstanding problems in paleoanthropology. The answers to those questions may hold the key to other mysteries as well, above all the nature and causes of the emergence of anatomically modern *Homo sapiens.*

For some years, the possibility existed that biology might furnish a straightforward explanation of the cultural revolution. There were two human types. One, the Neanderthal, was older (it was thought) and more robust than the other, called Cro-Magnon. To fashion a biological hypothesis, students of the question had only to link the archaic physical type to Middle Paleolithic culture and the gracile type to Upper Paleolithic culture. This explanation seemed to work in Europe, where paleoanthropological research had its start. There the old and stereotyped Mousterian technology was generally associated with the robust people known as Neanderthals, whereas the newer Aurignacian technology of innovative and constantly changing tools was linked to the tall and slender Cro-Magnons, who appear in the fossil record some 30,000 years ago. (The terms Mousterian and Aurignacian refer to the places in France where the artifacts were first found.)

This neat correlation of biology and culture came up against a difficulty in the Levant. Paleontologists digging at sites in the region found groups of fossils that included specimens that seemed older than their counterparts in Europe. Some specimens seemed to resemble the Neanderthals, others the Cro-Magnons. Yet both groups appeared to possess the same, primitive material culture. How did they and their artifacts fit into the pattern seen in Europe?

A definitive answer required that



OFER BAR-YOSEF and BERNARD VANDERMEERSCH are the co-organizers of a multidisciplinary project of excavations at the caves of Qafzeh, Kebara and Hayonim in Israel. Bar-Yosef is the MacCurdy Professor of Prehistoric Archaeology at the Peabody Museum, Harvard University. He was trained at the Hebrew University of Jerusalem and has conducted many excavations of Paleolithic and Neolithics sites. His interests center on the origin of modern humans and the transition to agriculture. Vandermeersch is a professor of anthropology at the University of Bordeaux I; he formerly taught for many years at the Pierre and Marie Curie University in Paris. His research focuses on human biological and cultural change during the Upper Pleistocene.

the Middle Eastern fossils be placed in chronological order. The task proved daunting. The specimens were too old to be dated by the traditional carbon 14 method, which can look back no more than 40,000 years. Uranium-thorium sequences require fossils that are clearly associated with carbonate sediments, and the potassium-argon series is better suited for dating lava flows and tuffs, which are normally found only in much older strata. In the absence of a definitive chronology, anthropologists thought it reasonable to follow the European pattern, according to which they presumed the robust Middle Eastern types to be older than and ancestral to the gracile types.

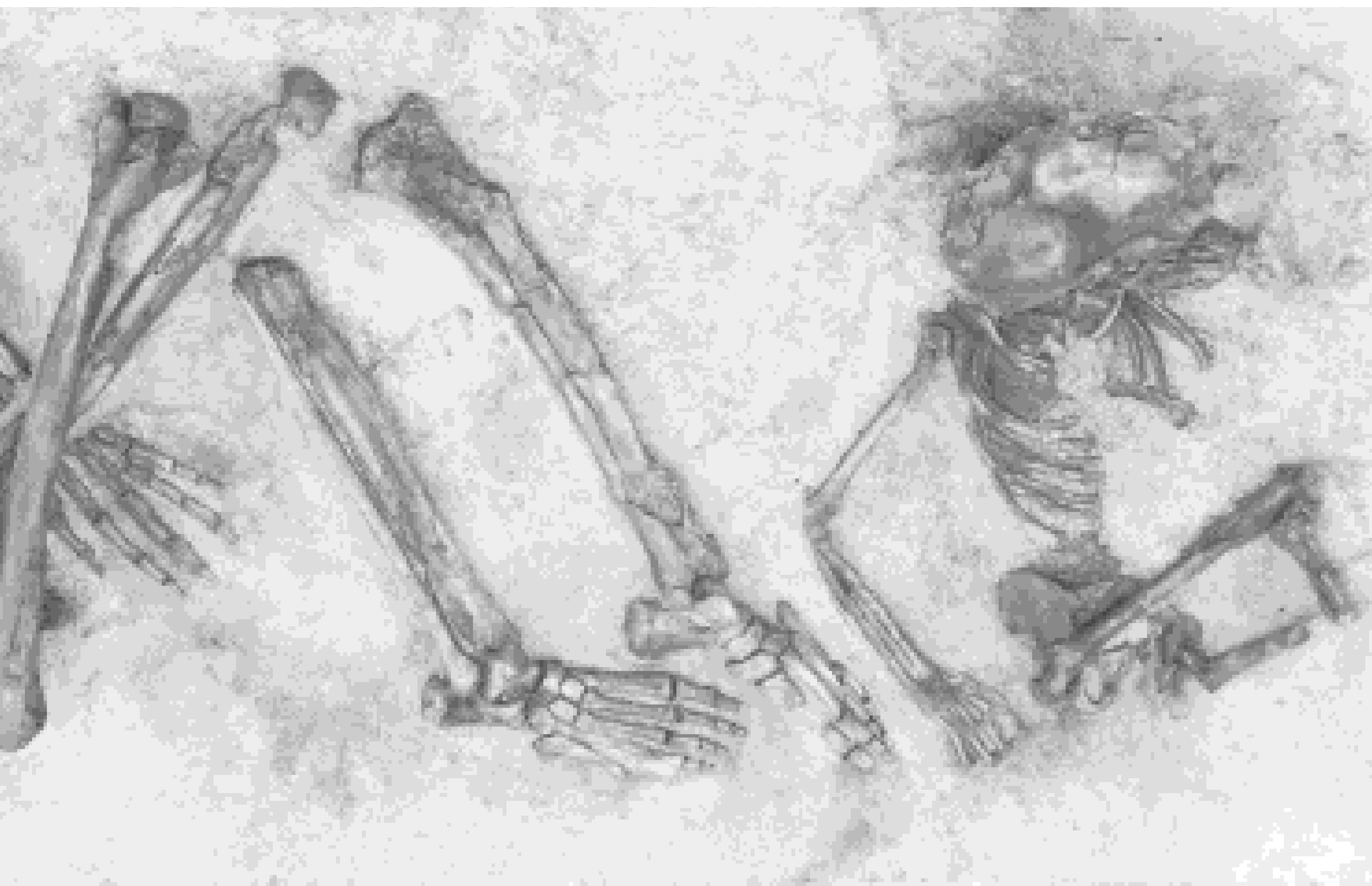
To test this hypothesis, we assembled an interdisciplinary team to construct a chronology from several kinds of evidence. The results have shaken the traditional evolutionary scenario, producing more questions than answers. Such statements may sound combative. They are not. This process of challenge is how science proceeds.

We began with the stratigraphic relations among the hominid fossils of the Middle Eastern cave sites. Then we and our colleague Eitan Tchernov of the Hebrew University of Jerusalem correlated this information with paleoclimatic and other data to estimate that modern-looking *H. sapiens* had lived in one of the caves some 80,000 to 100,000 years ago, much earlier than such people had been thought to exist anywhere. In the early 1980s our proposal was rejected, then vindicated soon after by two radiometric techniques—thermoluminescence and electron spin resonance—that had just been perfected.

The findings refuted the genealogical relation that had long been presumed to apply. No longer could the robust skeletons be identified as the ancestors of the gracile skeletons. At the same time, it became impossible to credit the emergence of modern-looking people with the cultural revolution of the Upper Paleolithic. When early modern people appear in the physical record in the Levant, they followed a way of life indistinguishable from that of archaic people. The artifacts associated with them clearly belong to the Mousterian culture.

Our findings are also relevant to another major question: Where did the human species originate? By showing that the gracile fossils were so remarkably old, we were able to tie them to the equally fascinating remains from the Klasies River Mouth caves and the Bor-

DOUBLE BURIAL from Qafzeh Cave (*right*) attests to a religious sensibility some 100,000 years ago. The infant was laid at the foot of a young woman, perhaps its mother. These people look quite modern, but their culture resembled that of the more robust Levantine Neanderthals. In Europe some 60,000 years later, modern-looking people and Neanderthals bore different cultures.



der Cave in South Africa. The South African fossils also seemed quite old and more closely related to modern than to archaic people. Could the two groups have belonged to a single, far-flung population? Some physical anthropologists say the answer is yes and conclude that modern-looking people emerged in Africa and then replaced archaic populations in other parts of the world [see "The Emergence of Modern Humans," by Christopher B. Stringer; SCIENTIFIC AMERICAN, December 1990].

Ironically, at about the same time as the new dates were being considered, molecular biologists were reaching similar conclusions about the homeland of modern people. They worked not with fossil bones but with the living fossil of DNA, which records in its many variants part of the story of evolution. By comparing a large number of samples from regional populations around the world, some of these researchers inferred that modern humans had originated in sub-Saharan Africa well over 100,000 years ago [see "The Recent African Genesis of Humans," by Allan C. Wilson and Rebecca L. Cann; SCIENTIFIC AMERICAN, April 1992].

Other investigators reject aspects of this genetic analysis, notably the posited phylogeny, or family tree. They support a rival interpretation called the multiregional model, which holds that modern-looking people emerged in many different populations at roughly the same time and that gene flow, or hybridization, maintained the unity of the species [see "The Multiregional Evolution of Humans," by Alan G. Thorne and Milford H. Wolpoff; SCIENTIFIC AMERICAN, April 1992].

THREE-DIMENSIONAL GRID at Kebara (*above*) guides archaeologists, who record the location of every bone and artifact and also examine thin sections of sediment with a microscope. Such techniques enable workers to place objects in time and distinguish periods in which the cave was continuously inhabited by humans. A cross-sectional view (*opposite*) shows the ongoing excavations by the authors and their colleagues at the mouth of the cave and in the chamber framed by its high dome. Some parts have been left untouched for future archaeologists, a practice followed by Moshe Stekelis in the 1950s but not by Francis Turville-Petre in the 1930s.

The story of the Levantine hominids begins in 1929, when Dorothy A. E. Garrod, an archaeologist from the University of Cambridge, began a five-year project to excavate Skhul, Tabun and el-Wad, three caves at the mouth of Wadi el-Mugharah (today Nakhal HaMe'arot) [*see map on preceding page*]. In the process, she uncovered the basic prehistoric sequence of the Upper Pleistocene for the entire Levant. The deepest stratigraphy, at Tabun, extended through a full 23 meters. Garrod classified the layers from top to bottom according to their typical artifacts: Upper Acheulean (layers G–F), Acheuleo-Yabrudian (layer E) and the Middle Paleolithic Mousterian (layers D, C and B and the fill in the chimney).

Garrod also uncovered the remains of an adult woman, since classified as a Neanderthal and tentatively dated to 60,000 years ago. She could not determine whether it was situated in layer B or layer C but had no doubt that the jaw she found just below it belonged to layer C. This jaw, also of a young adult, paradoxically seemed closer to the gracile morphology of the Qafzeh people than to the robust form of the Neanderthals.

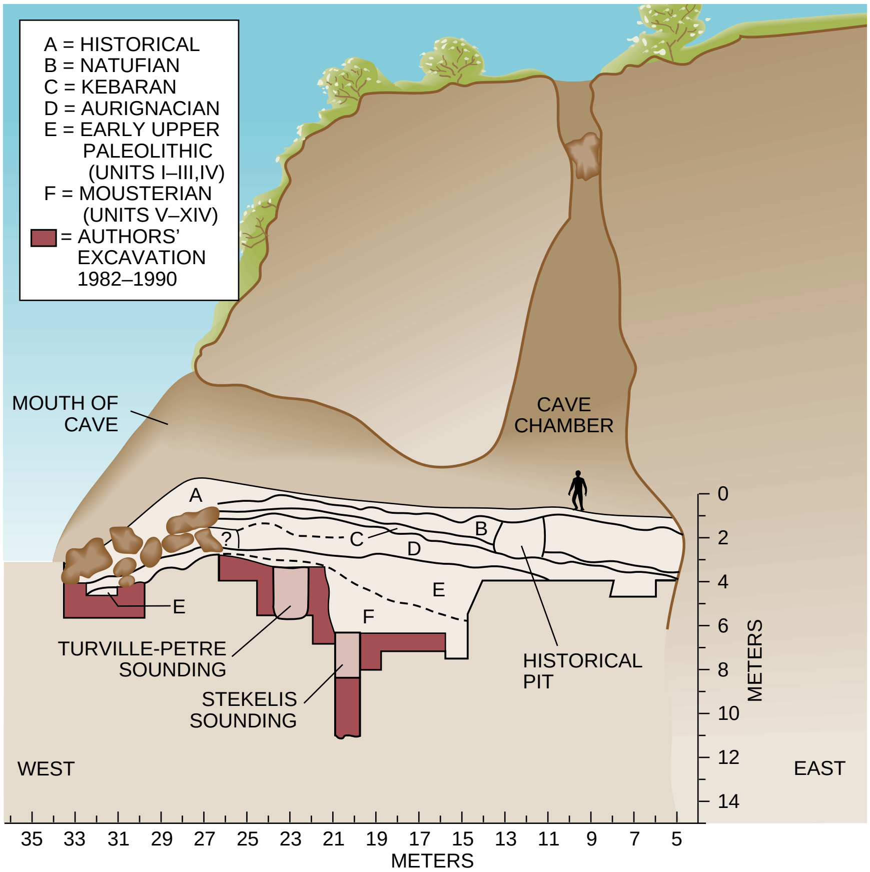
The cave of Skhul produced a most amazing collection of Middle Paleolithic skeletal materials, including several well-preserved burials. We believe the arrangement of the woman's skeleton marks it as a deliberate burial, the earliest that has been found so far. Because such sophisticated behavior is impossible without symbolic thought, the find provides indirect evidence of the existence of language. Moreover, as we shall presently show, the Neanderthals' mortuary practices—together with their stone tools—are in no way inferior to those of the gracile hominids who preceded them in the region.

The first major examination of these materials was conducted by Sir Arthur Keith of the Royal College of Surgeons in London and Theodore D. McCown, who later taught at the University of California at Berkeley. Treating the entire assemblage from Skhul and Tabun as a unit, the scholars concluded that the skeletons represented a form intermediate between Neanderthals and modern humans. Some anthropologists have considered these specimens as possible hybrids; others regard them as a locally evolved population.

In the same decade, surprising discoveries were also made in Qafzeh, a large cave near Nazareth, 35 kilometers as the crow flies from Mount Carmel. Between 1933 and 1935 René Neuville, the French consul in Jerusalem and a member of the Institute of Human Paleontology in Paris, and Moshe Stekelis of the Hebrew University jointly conducted excavations there. Outside the main chamber, in the Middle Paleolithic layers near the entrance and in the terrace in front of the cave, more than four meters of debris was excavated, exposing several human fossils.

Although the workers did not publish detailed descriptions of these fossils, Henri V. Vallois of the Institute of Human Paleontology ventured to classify them as early, or nonclassical, Neanderthals. In the 1950s F. Clark Howell of Berkeley suggested that the more modern-looking skeletons, with rounded skulls and flat faces, were probably the ancestors of the European Cro-Magnons. For this reason, many anthropologists refer to them as Proto-Cro-Magnon.

Sites elsewhere in the Middle East yielded additional Middle Paleolithic fossils in the 1950s and early 1960s. Ralph S. Solecki, then at Columbia University, unearthed two groups of skeletons in



Shanidar Cave in the foothills of the Zagros Mountains in Iraq. Both clusters consisted of Neanderthals, and both showed evidence of human, or shall we say humane, behavior. One group included the skeleton of a man who had survived the loss of an arm. Such a feat would surely have been impossible without sustained caretaking. The other set of remains showed clear signs of deliberate burial—yet more evidence of humane behavior. A similar Neanderthal burial was found in Amud Cave near the Sea of Galilee by Hisashi Suzuki of Tokyo University.

Thus, by the 1970s, the evolutionary scenario for the Near Eastern Upper Pleistocene fossils appeared relatively straightforward. Neanderthals from western Asia were represented by the skeletal remains from the Tabun, Amud and Shanidar caves. This population, it was suggested, then evolved into the local Proto-Cro-Magnons of Skhul and Qafzeh.

Two models vied to explain the next stage, which began 40,000 years ago. The first model held that the Levantine population of modern humans moved or diffused into Europe, replacing or interbreeding with local Neanderthals. The second model proposed that local Eurasian Neanderthal populations evolved, in place, into modern forms. William W. Howells of Harvard University dubbed them the "Noah's Ark" and the "Neanderthal phase" theories. They survive to this day, with modifications to take account of our new dates, as the out-of-Africa and multiregional hypotheses.

In the 1960s paleontologists and other investigators became increasingly aware that their cave stratigraphies were much more complex than they had seemed. We attacked the problem in the caves of Mount Carmel with the full weight of modern excavation techniques. Each fossil fragment and human artifact was located within a three-dimensional grid [*see illustration on opposite page*]. Geologists and micromorphological analysts examined sediments for signs that they had been disturbed by animals or humans from later stratigraphic periods. Pollen and other microscopic evidence were studied and linked to reconstructed climate conditions.

Such systematic reinvestigation began at Tabun, where the thick depositional sequence provided an excellent, long-term record of cultural changes. Excavations there were conducted by Arthur J. Jelinek of the University of Arizona between 1967 and 1972 and more recently by Avraham Ronen of the University of Haifa. The second cave, Qafzeh, was excavated by one of us (Vandermeersch) from 1965 to 1979; the other author joined in the project in the last two seasons.

The excavations in Tabun uncovered a wealth of geologic observations and well-preserved stone artifacts but few rodent fossils. Such remains derive from the deteriorated pellets of barn owls, which do not share caves with humans. Paleontologists use rodent remains to reconstruct the environment and to place the strata in chronological order.

While conducting a detailed study of the stone artifacts, Jelinek noticed that the thickness of the flake tools declined as strata progressed from layers C and D (Mousterian) through layer E (Acheuleo-Yabrudian) and layer F (late Acheulean). He showed that when this trend is plotted against the proposed time scale for Tabun Cave, the resulting curve had an angle. At that point, the process had suddenly sped up. Jelinek suggested that the acceleration reflected an evolutionary transition from Neanderthals to modern humans.

William R. Farrand of the University of Michigan calculated a timetable for Tabun by examining the site's sedimentary accumulation. He interpreted the massive sandy unit at the bottom, incorporating most of layers G, F and E, as the remains of dunes brought to the cave mouth by an encroaching sea—the result of the melting of polar ice caps during the last interglacial, about 100,000 years ago. The loess soil of layer D he explained as the mark of the onset of cold conditions around 75,000 years ago, the Mousterian period. The upper part of the sequence, layers C and B and the fill in the chimney, would have accumulated as recently as the period from 55,000 to 40,000 years ago, when the opening of the cave's chimney let rain wash in the characteristic Mediterranean *terra rossa,* or red earth.

But the reconstruction was soon shown to conflict with analysis of the faunal evidence in Tabun and Qafzeh conducted largely by Tchernov. Tchernov found that the hominid-bearing layers in Qafzeh contained several archaic rodent species not represented in layer C of Tabun. This discrepancy in the faunal markers implies that the two groups of hominids had lived at different times. Such an interpretation led us to reconsider the cave stratigraphies and the human bodies they had concealed.

At least three burials of modern-looking humans exist in Qafzeh, and their place in the sediments seemed our best clue as to their ages. One burial is that of an adult, knees partially drawn up, lying on its right side in a small, natural niche in the limestone wall. The second is a double burial, the only one known

GRACILE AND ROBUST: the gracile skull (*left*), from Qafzeh, is modern in having a high, short braincase and a definite chin. The robust skull (*right*), from Tabun, has such characteristically Neanderthal features as a projecting midfacial area and a "bun" at the back. Both populations had brains at least as large as those of people alive today.

from the Middle Paleolithic world. A skeleton of a woman between the ages of 18 and 20 years was found lying on her right side, her hands over her abdomen and her legs half folded. At her feet, perpendicular to her body, lay the skeleton of a child about six years old. Because the two skeletons were lying horizontally, cutting across the strata of the site, in a configuration resembling the boundaries of a rectangular pit, it is thought that the double burial was made in a dug-out grave.

The third burial, a boy about 13 years of age, was found in a cavity cut into the soft bedrock. The body had been laid on its back, the skull leaning on the grave's wall and the hands facing upward. A large antler of a fallow deer had been placed over the hands, across the upper part of the chest. That the antler had escaped the ravages of hyenas implies that it had been deliberately interred. In this practice, as in their tools, the early modern people cannot be said to differ fundamentally from the robust folk who came later to the region.

The great abundance of rodent remains in the lower layers indicates that humans lived in the cave only intermittently, perhaps seasonally during this period. In any case, their annual round of foraging appears to have reached as far east as the Jordan Valley and as far west as the Mediterranean shore, some 40 kilometers away. The harvest of the sea appears as perforated shells of the cockle *Glycymeris* that have been uncovered in the lower layers. What function could the shells' holes have served, except to make way for the string of a necklace or amulet? Such purposes bespeak the presence of symbolic thought.

Finally, we found a lump of red ocher, scored with clear scraping marks. A microscopic examination of the stone artifacts by John J. Shea, then at Harvard, indicates the presence of red ocher on the working edges of several tools. In later eras this material has served as a pigment and, when mixed with oil, as a waterproofing agent.

In 1979, during the final season of excavations, we tried to reconstruct the paleoclimatic sequence. First we used the strata of Qafzeh bearing rodent fossils to calibrate the climatic sequence at Tabun; then we applied the sequence to the units of Qafzeh that had yielded the human fossils. Finally, we correlated the sequence with the isotope stages from deep-sea cores. The process of triangulation is what seems to suggest that the Qafzeh hominids had lived about 80,000 to 100,000 years ago.

The heavy criticism we incurred led us to try to resolve the problem in two ways. We would apply recently developed dating methods, and we would choose a site whose state of preservation allowed a more complete reconstruction of behavior patterns. They include the use of fire, the placement of dumping areas in the cave, the kind of plant foods that had been gathered and the animals that had been hunted, trapped or scavenged. By getting such a complete picture (instead of one based only on stone tools), we believe one may test whether the Proto-Cro-Magnons and the Middle Eastern Neanderthals were in reality two different populations or species.

Kebara Cave on the western escarpment of Mount Carmel was an obvious choice because of its very full sedimentary record. Kebara is perched some 60 meters above sea level. Its arched entrance, facing north-northwest, was hardly altered through the Paleolithic period, although a later rockslide has created a small terrace in the front. We began digging in 1982.

The layers from the Upper Paleolithic had for the most part been stripped away during the 1930s, during the excavations led by Francis Turville-Petre, an independent scholar then collaborating with the British School of Archaeology in Jerusalem. Stekelis resumed excavations in 1951, and over a 14-year period his group exposed Middle Paleolithic layers and, with them, the remains of a Neanderthal baby.

We studied these layers in greater detail, particularly those associated with a series of successive periods of human habitation. The stone artifacts show a particular resemblance to layer B of Tabun and to Amud, not so much in their general form (which remained static throughout the Mousterian) but in the methods by which they had been made.

The most stunning discovery, made in 1983, was the burial of an adult male in a dug-out grave pit, later dated to about 60,000 years ago. The skull, right leg and foot were missing, and the left foot had been destroyed by water leaching through the deposits. The skull seemed to have been deliberately removed after the flesh had decayed, perhaps many months after the man had died; no other explanation can account for the undisturbed position of the bones, above all the lower jaw, which remains in place. What but a religious motive can have shaped such a mortuary practice?

The jaw constitutes the most valuable find. It is shaped like the jaw of a Neanderthal, but unlike any other such jaw it retains its hyoid, the isolated bone that controls the tongue. From the shape of the bone, Baruch Arensburg of Tel Aviv University and his associates concluded that the man's vocal tract had resembled that of modern people and that he had probably been able to make many of the vocalizations necessary for articulate speech. Another bone, the inominate, persuaded Yoel Rak of Tel Aviv University that the man's pelvis was closer to that of Neanderthals than to that of modern people.

To infer speech and other peculiarly human behaviors requires more than bones alone can provide. One must reconstruct actions from the objects that have been acted on: archaeological artifacts. Certain of these features in the Mousterian layers fill the bill quite well. Most of the bones and waste flakes, for instance, were found near the cave wall, whereas the central part of the chamber was covered by the remains of hearths. This pattern indicates a characteristically modern mode of behavior: cleanliness.

The evident presence of morphologically modern humans in the Levant some 100,000 years ago puts other western Asian fossils in a new light. Modern humans may have a rather extensive local genealogy. For example, the fragmentary skull found by Turville-Petre in Zuttiyeh Cave near Amud turns out to be between 200,000 and 250,000 years old. This individual could have been part of the population from which the gracile people of Skhul and Qafzeh were derived.

The Neanderthals, however, seem to have had no local forebears; the only record of pre-Neanderthals is in Europe, where it dates back to 150,000 years ago. In the absence of other fossil data, we must tentatively conclude that the Kebara Neanderthals stem from European stock. Either the founding population immigrated from Europe to the Levant during a single lifetime or, as seems more likely, it expanded gradually, over many generations. The latter process, called demic diffusion, might have involved some interbreeding with the people who lived along the path of diffusion.

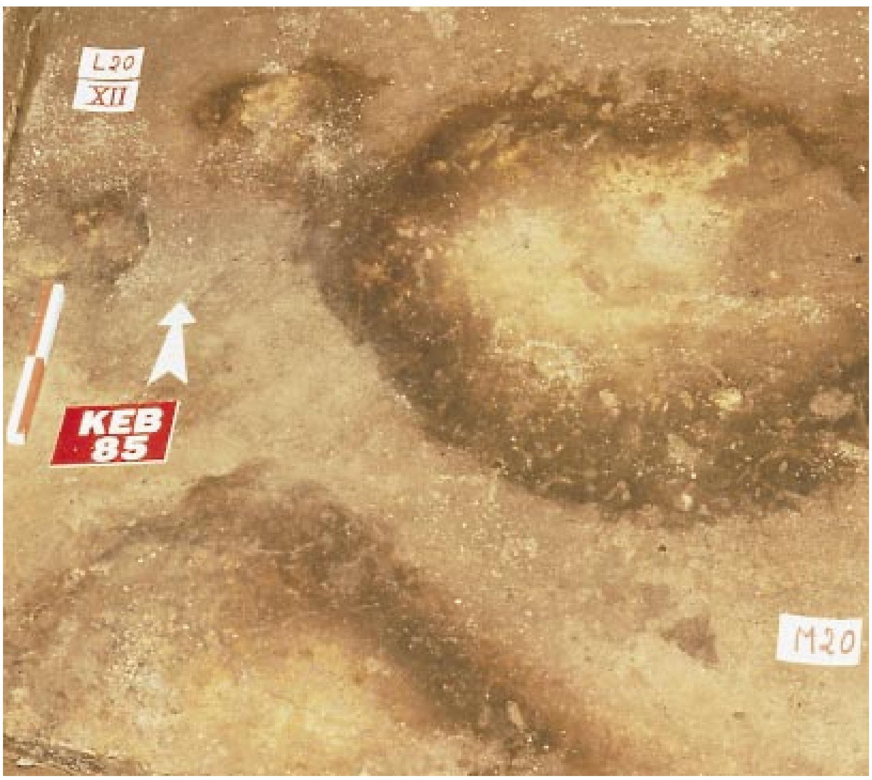
What might have induced such an expansion out of Europe? The Neanderthals were adapted to cold, as their compact bodies attest, but even they could not brave the Arctic conditions that occurred, in fairly sudden cold snaps, during the period between 115,000 and 65,000 years ago. The intense cold might have forced them southward. In support of this hypothesis, one may cite the resemblance between the Middle Eastern Neanderthals and an older skull from Saccopastore in Italy. The similarity suggests that the expansion from Europe followed the edge of the Mediterranean basin, perhaps through modern-day Turkey or the Balkans.

Whom did the Neanderthals find when they arrived in the Middle East? Several hypotheses might be tested. One may imagine that cold weather pushed the gracile people into Africa even as it pushed the robust people into the Levant, thus preventing the two groups from coming into contact. If, however, the descendants of the gracile people stayed on in the area, they may have coexisted with the Neanderthals or even bred with them. Some scientists consider such hybridization as unlikely, preferring instead to regard the populations as distinct species.

Normally, paleoanthropologists would try to work their way around the dearth of fossils by inferring behavior from the different modes of tool production, the techniques of foraging and food processing, the presence or absence of fireplaces and the evidence of artistic activity. But the relevant archaeological remains are also far from conclusive.

Chipped stone tools and the materials from which they were hewn are the most abundant remains in Paleolithic sites. The Middle Paleolithic materials were systematized by the late French archaeologist François H. Bordes, who interpreted the observable differences in tool assemblages as demonstrating the presence of several prehistoric tribes. But Lewis R. Binford of Southern Methodist University has challenged Bordes's reasoning. He argues that the Middle Paleolithic hominids probably made up a single population and that they used different sets of tools to perform different tasks.

The debate led to two decades of intensive microscopic research in which tools were made and used in one way or another; then the investigators compared the patterns of wear with those found on ancient tools. But this study failed to uncover the purposes to which the Mousterian tools had been put. In



HEARTH REMAINS were found in the middle of Kebara Cave, whereas garbage was clustered near the wall, implying a modern mode of behavior: cleanliness.

REMARKABLE NEANDERTHAL MAN from Kebara was beheaded after the flesh had decayed, leaving in place the lower jaw and the hyoid, or tongue bone. The bone's shape suggests that early humans could produce modern speech sounds.

most cases, form did not correlate with function. Although there were exceptions to this rule, it is clear that scrapers, the predominant tool, must have performed many different tasks.

One behavioral deduction that can be made from tool remains is an estimate of how far ahead humans planned. Studies have shown that humans were discriminating in their use of valuable stone and that they would go to great lengths to obtain it. Prehistoric people would not lightly discard artifacts made of first-rate materials, but they did not hesitate to throw away those made of second-rate materials, often right after having found, fashioned and used them.

Because the Levant is generally rich in flint, the Middle Paleolithic people there had to walk only a few kilometers to obtain high-quality nodules. They sometimes shaped the flint on the spot; often they carried the nodules, or cores, back to the cave. In either case, they achieved the form they desired by successively detaching flakes or points from the core. After using the resulting tool, the people sometimes resharpened it. This entire operational sequence constitutes evidence of long-term planning, a characteristically human trait.

The unmistakable signs of deliberate burials in Qafzeh, Skhul, Amud and Kebara; the perforated marine shells in Skhul and Qafzeh; and the red ocher used in those two caves and in Hayonim Cave all attest to the presence of symbolic behavior. The modern-looking hyoid bone of the Kebara Neanderthal suggests that articulate language had evolved. The Middle Paleolithic hominids possessed the elements of modernity, but the pace of innovation was unaccountably slow.

Resolution of the phylogenetic debate should eventually come from studies of DNA recovered from well-preserved human bones. But phylogeny cannot explain the cultural revolution that ushered in the Upper Paleolithic age. To solve that problem, scholars must infer changes in human behavior from the subtle traces of the physical record.

Geologists, micromorphologists and other experts must combine their efforts to distinguish the residues of human activity from those of animals and other natural agents. Archaeologists must then characterize the layouts of Middle and Upper Paleolithic campsites and seek in the differences some hint of the changes that led to the explosion of innovation. Just as no historian could explain the Industrial Revolution without reference to the Enlightenment, no prehistorian can account for the technological breakthrough some 40,000 years ago without first unraveling the mystery of Mousterian life.

FURTHER READING

PALEOANTHROPOLOGY. M. H. Wolpoff. Alfred A. Knopf, 1980.

THE SHANIDAR NEANDERTALS. E. Trinkaus. Academic Press, 1983.

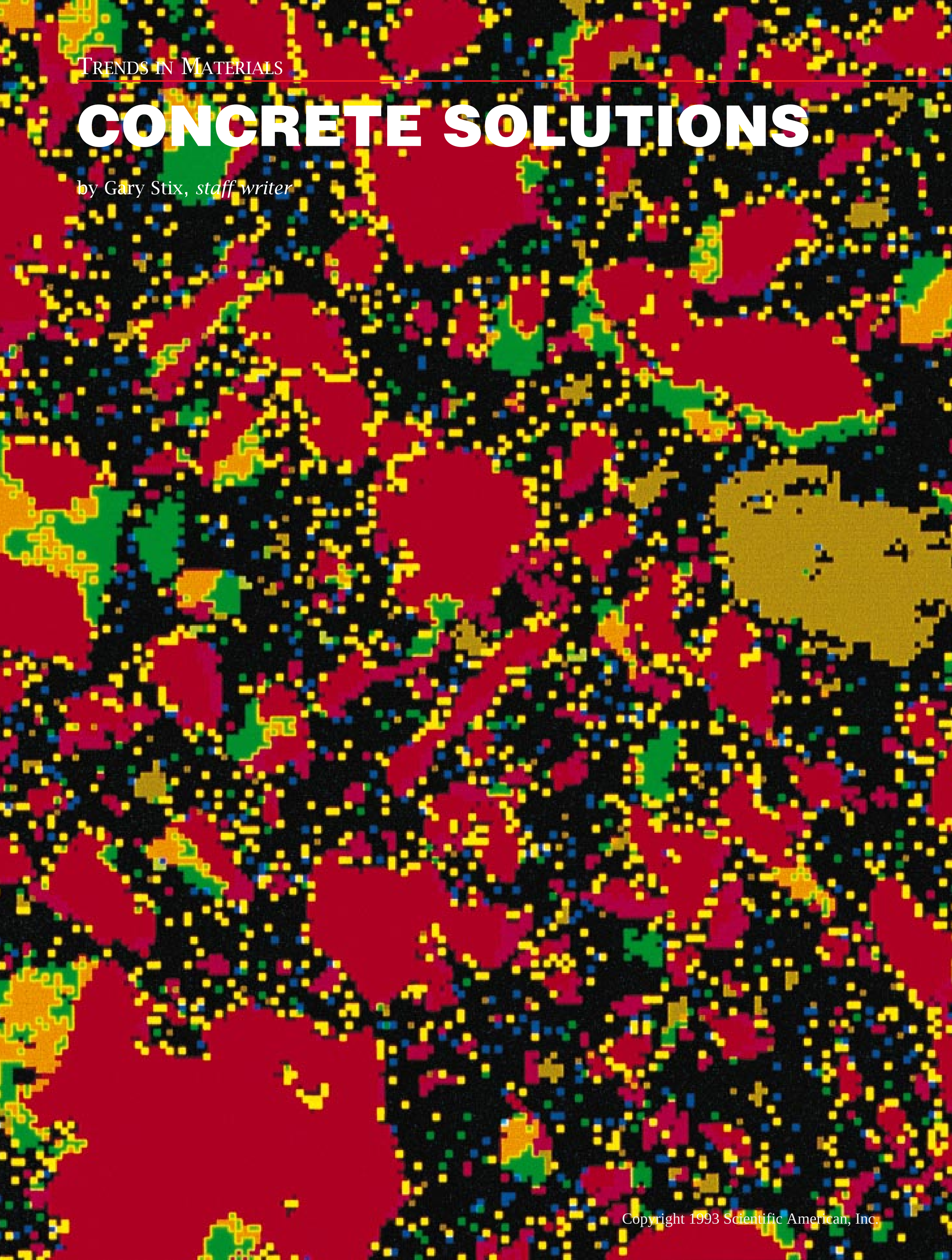
WESTERN ASIA. E. Trinkaus in *The Origins of Modern Humans: A World Survey of the Fossil Evidence*. Edited by F. H. Smith and F. Spencer. Alan R. Liss, 1984.

THERMOLUMINESCENCE DATING OF MOUSTERIAN 'PROTO-CRO-MAGNON' REMAINS FROM ISRAEL AND THE ORIGIN OF MODERN MAN. H. Valladas, J. L. Reyss, J. L. Joron, G. Valladas, O. Bar-Yosef and B. Vandermeersch in *Nature,* Vol. 331, No. 6157, pages 614–616; February 18, 1988.

ELECTRON SPIN RESONANCE DATING AND THE EVOLUTION OF MODERN HUMANS. R. Grün and C. B. Stringer in *Archaeometry,* Vol. 33, Part 2, pages 153–199; August 1991.

CONCRETE SOLUTIONS

by Gary Stix, *staff writer*



Rehabilitating the nation's aging infrastructure may depend on how well industry, government and academia tweak the properties of materials, from lowly concrete to aerospace composites.

On an engineering workstation, the way in which cement reacts with water is depicted as a wild eruption of reds, greens, yellows, blacks, magentas and oranges. As the hydration of cement particles progresses, the screen turns into a brilliant shower of yellow, punctuated with patches of blue and red. These images might have occurred to an abstract expressionist artist who enjoyed access to a high-resolution graphics display.

Such artistic achievements are the serendipitous product of serious efforts by a scientist and an engineer from the National Institute of Standards and Technology (NIST) to create two- and three-dimensional models of this ubiquitous gray paste. The work reflects a growing recognition that even the most basic of construction materials deserves renewed respectability and attention in the pantheon of materials science.

Without such efforts to improve old materials and to adapt newer ones, the battle to restore the American highway infrastructure—and make it comparable to the best roads in Europe—could well be a losing one. Indeed, a growing number of researchers, administrators and industrialists believe such a mundane use of high technology may provide benefits as great as those obtained from investigating ways to make a smaller transistor or a lighter aircraft fuselage. Certainly the task is of epic proportions. The National Research Council's Transportation Research Board has estimated that bettering the durability of roads and bridges by a single percentage point over a 20-year period would yield a savings of $10 billion to $30 billion. That money could be used to help offset an overly burdensome repair bill.

By 1995, 35 percent of the interstate pavements will have outlived their useful life, a congressional report warns. The Federal Highway Administration (FHWA) estimates that the U.S. confronts a $167-billion toll to repair bridges and roads that are

CEMENT FANTASIA, conceived at the National Institute of Standards and Technology, simulates an early stage of cement's hydration reaction. This type of computer model will help design improved versions of ordinary construction materials for more durable rehabilitation of infrastructure.

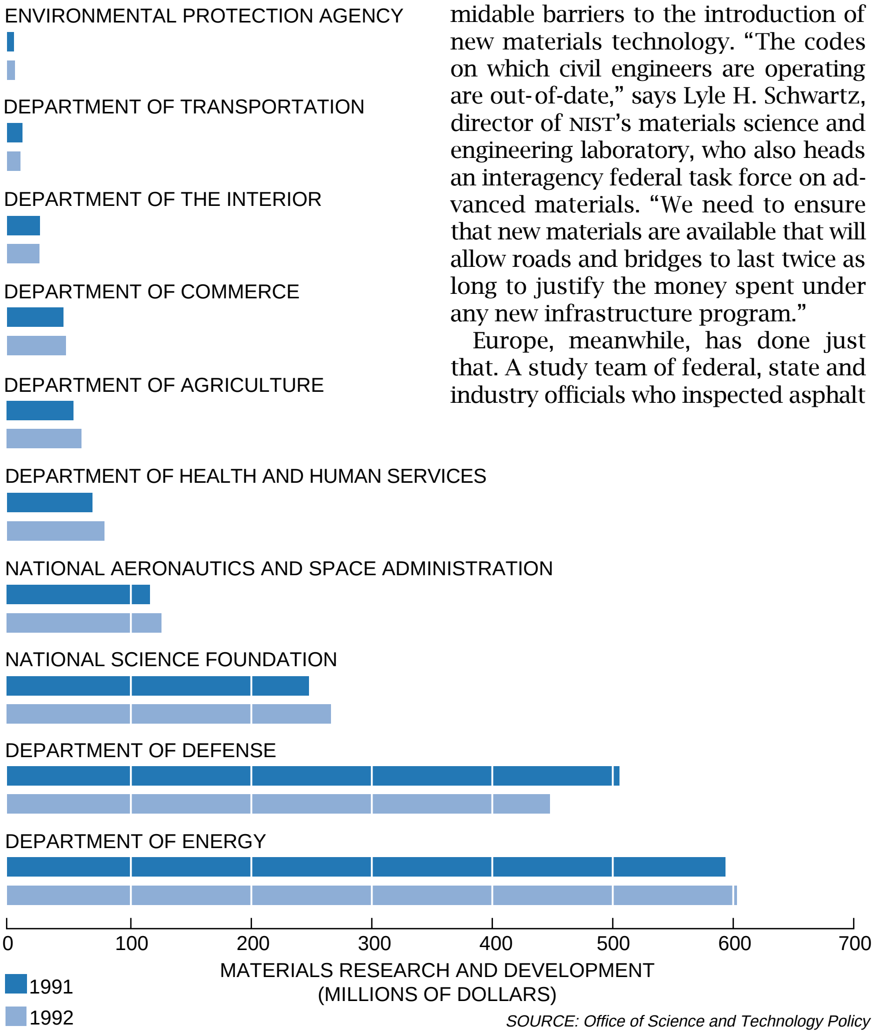
deficient or about to fall below minimal standards. To add new high-occupancy vehicle lanes and other capacity improvements needed to alleviate grinding traffic congestion would require an additional $123 billion.

Not surprisingly, the pothole, usually the focus of small-town mayoral races, has arrived in the national spotlight. The Surface Transportation Act of 1991 laid out a $151-billion blueprint for rehabilitating decaying bridges and roadways that also gave states the freedom to expand mass-transit systems. Fulfilling a 1992 election campaign pledge, President Bill Clinton is exploring an economic stimulus package to create jobs that would allocate $5 billion to $10 billion for public works projects.

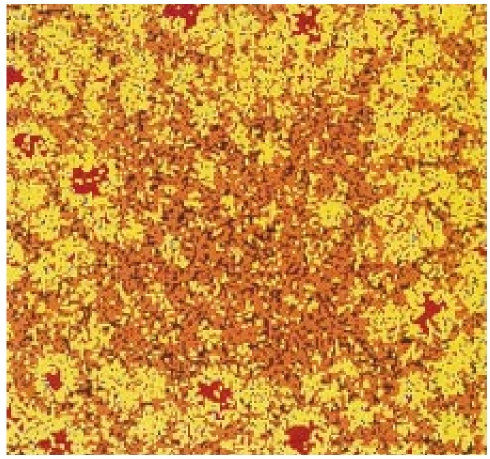
Yet these investments may produce the equivalent of temporary fill for potholes if the set of materials in use is as old as the pavements and pylons it is intended to replace. Until the late 1980s, policymakers and many research scientists had ignored the need to adopt improvements in the basic materials that constitute roads, bridges, runways, wastewater treatment plants and other parts of the nation's infrastructure, including prospective elevated guideways for high-speed trains. The Reagan administration tried to eliminate NIST's Center for Building Technology and to turn over operation of the FHWA's main research center to the private sector. Funding for materials research at the Department of Transportation and the Environmental Protection Agency, two agencies that help to supervise the public works agenda, is still minuscule compared with the amounts spent on composites for defense use or on energy-related superconductivity [*see illustration below*].

Moreover, conservative state and local construction regulations in the 50 states and 83,000 localities, which too often specify plain-vanilla varieties of concrete, steel and asphalt, remain formidable barriers to the introduction of new materials technology. "The codes on which civil engineers are operating are out-of-date," says Lyle H. Schwartz, director of NIST's materials science and engineering laboratory, who also heads an interagency federal task force on advanced materials. "We need to ensure that new materials are available that will allow roads and bridges to last twice as long to justify the money spent under any new infrastructure program."

Europe, meanwhile, has done just that. A study team of federal, state and industry officials who inspected asphalt



RESEARCH GAP on infrastructure-related materials became apparent when federal officials observed the comparatively low levels of materials research at the Department of Transportation and other agencies that shape public works policy. A $2- to $4-billion research program has been proposed to help fill this void.

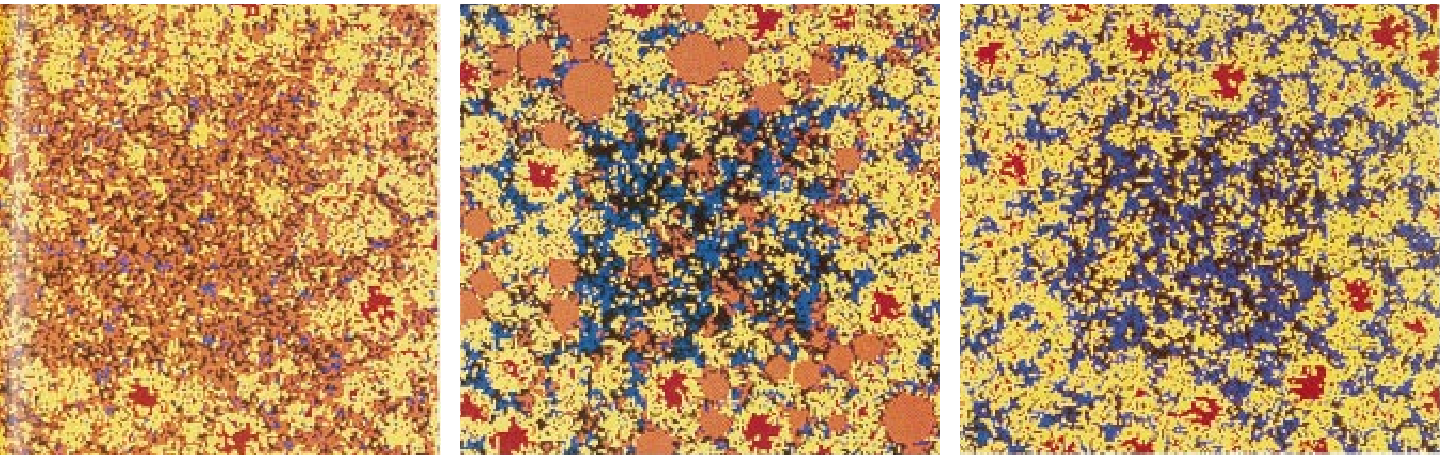


SIMULATIONS done on a Cray-2 supercomputer by the National Institute of Standards and Technology show how cement particles in concrete react with water and grow in the presence of differ-

pavements in six European countries during 1990 came home marveling at the superior condition of foreign roads. "It sort of shakes your American ego because I had always assumed that we were number one," laments Wayne Muri, chief engineer for the Missouri Highway and Transportation Department and president of the American Association of State Highway and Transportation Officials.

Lured by the Clinton administration's vow of increased spending, federal funding agencies have rolled out plans nurtured over the past few years to push infrastructure technology from the laboratory to the highway department garage or construction sites. The National Science Foundation plans to suggest more funding and better coordination among the roughly $50 million in existing infrastructure research programs. More ambitious is a proposal, involving NIST, the FHWA and the Civil Engineering Research Foundation, that calls for government and industry to devote up to $4 billion over the next 10 years to a series of projects that would employ advanced materials, including concrete, steels and composites. The program would amass data on the long-term performance of these materials in order to help ease the reluctance to specify their use in construction codes.

These projects will be welcomed by the small coterie of back-to-basics researchers who have fought hard to convince grant providers that advanced research on materials that resemble those used to build Roman aqueducts can pay off. J. Francis Young, a professor of materials science and civil engineering at the University of Illinois, remembers



ent mineral additives. The faint rectangular area at the center of each image represents the space in front of a single sand particle. By adding silica fume (*far left*), calcium silicate hydrate particles (*yellow and brown pixels*) group tightly around the sand grain. The relatively low amounts of water (*black pixels*) show that the material is less porous and thus becomes stronger. Higher porosity is evidenced by the greater number of black pixels in concrete made from cement without additives (*far right*). Adding different types of fly ash (*middle images*) produces a material with an intermediate number of pores.

the day in 1988 when he received a call from Surendra F. Shah, a professor of civil engineering at Northwestern University. Shah wanted to know whether Young would be interested in joining Northwestern, NIST and a few other institutions in submitting a proposal for one of the NSF's new Centers for Science and Technology. Its focus: cement. "I thought it would be like taking a chance in the Illinois lottery," Young says. "The chances of winning were pretty small, but what the hell."

Those at the NSF who evaluated the proposal—one of more than 350—were surprised to receive a submission on cement. Outside reviewers, though, liked the idea of bringing together physicists, geologists, chemists and civil engineers to study the chemistry of the material that binds together sand and stone to form concrete. Today the workers watch on a cathode-ray tube the reaction that occurs when calcium silicates and calcium aluminates react with water: it is an important step in making high-tech versions of a material with a 2,000-year history.

Cement Calling Cards

This new age of old materials marks the arrival of "cement lite." Young can hand a visitor a business card that identifies the NSF Center for Science and Technology of Advanced Cement-Based Materials (ACBM). The card is made from an advanced cement that can be flexed back and forth in the holder's hand or stuffed into a wallet.

Indeed, cement and concrete have undergone a renaissance during the past 20 years or so. The compressive strength of concrete used in building materials has jumped by more than a factor of four from a relatively weak 5,000 pounds per square inch or less. And the concrete recipe book has expanded: concretes have been made that set faster and are both more ductile and flexible than those made with the conventional material, called portland cement. "The most interesting area for new materials for us is in the use of concrete," says Larry T. Papay, vice president and manager of research and development for Bechtel, the San Francisco–based construction giant.

Many such materials have moved on to construction jobs. In 1989, for example, the French Ministry of Public Works built a bridge more than 350 feet long near the town of Joigny using high-strength concrete. The 5 to 10 percent price premium of the concrete was offset by the fact that 30 percent less material was needed in columns, decks and other structures.

Yet civil engineers in the $500-billion U.S. construction market remain cautious. Uncertainty about how a thinner material would perform during decades-long exposure to tractor-trailer trucks and the onslaughts of spring flooding is one cause to fret. "If a vehicle ran into a barrier of this material, the concrete might break apart, if not properly designed," worries Thomas J. Pasko, Jr., director of the FHWA's Office of Advanced Research.

Researchers at ACBM (whose members are from Northwestern, the University of Illinois, Purdue University, the University of Michigan and NIST) seek to ensure that concrete passes the test of time. Specifically, they want to understand how variations in pore structure, which gives cement its brittleness, might affect the life span of a railroad overpass or an interstate cloverleaf ramp.

Young foresees a day when a contractor will be able to stick a battery of probes into the still-hardening concrete to sample acoustical, electrical and thermal properties. These measurements could be beamed to a remote computer that will compare the data to models now being cooked up on workstations and even Cray supercomputers. The raw measurements might then be translated into relatively accurate predictions of stiffness, strength and durability.

Introducing an element of predictability into the performance of the materials, Young says, may make U.S. contractors more willing to provide assurances about the quality of their work. This practice is already widely used in Europe and is one reason for the better quality of roads there. "If you could use this technology to guarantee an extra 20 years of life, people would buy this idea," Young asserts.

Besides formulating supercomputer models, ACBM is hoping to extend the performance envelope of cement and concrete. Along with a visiting scientist from China's Tongji University, Young made small samples with silica fume that could withstand compressive stress of a record 100,000 pounds per square inch, double the amount for some mild steels and comparable to that of a structural ceramic. The experiment will not be replicated soon by a road crew. The sample had to be repeatedly compacted under high pressures while it was cured at temperatures of up to 200 degrees Celsius.

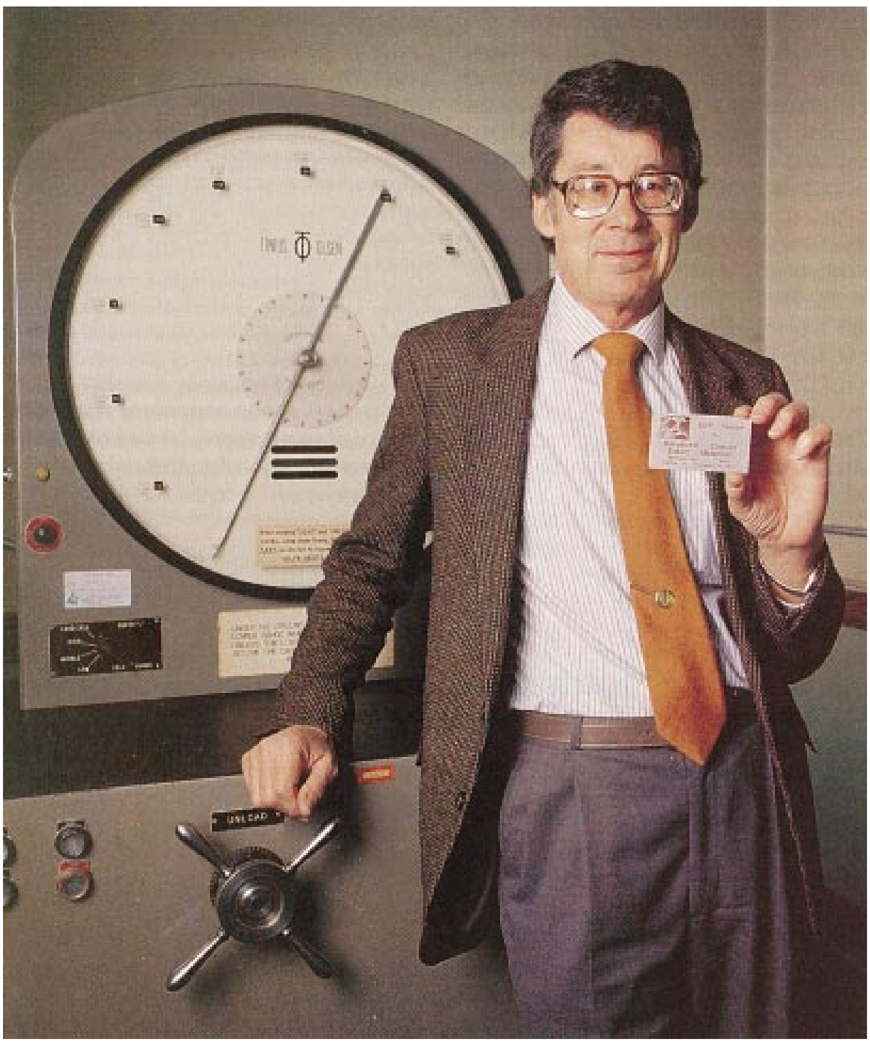
Higher-strength concrete does little to solve cement's brittleness problem. The fragility of the material is the reason that engine blocks are not made from cement. Northwestern and other universities are exploring how to make dramatic increases in its ductility. As with other types of composites, adding fibers—glass, carbon or steel—can reduce proneness to fracture. The tiny strands of fiber act to prevent cracks from progressing through the material.

Cements dosed with half a percent of fiber are available commercially. Shah of Northwestern has produced and tested a paste that is 10 to 15 percent fiber. The new material has 1,000 times the ductility and five times the tensile strength of conventional cements. This work might eventually turn up a concrete that could be laid down as a long ribbon, instead of in sections separated by joints. This would prevent the fracturing that often occurs at these interstices. In an earthquake, pylons of the material might bend instead of buckling.

Earthquakes are just one way to make the ground move. At a congressional hearing in February 1992, Representative Norman Y. Mineta of California asked what research could be marshaled to help solve a major national problem. Mineta was referring to potholes that result from temperature extremes, moisture and from axle loads of up to 20,000 pounds applied thousands of times each day to a hardened mix of stone-studded asphalt, a confection bound together with the thick, gooey dregs that remain after gasoline and jet fuel are refined out of crude petroleum.

In addition to the mail and taxes, Mineta says, the damage to vehicles from potholes is one of taxpayers' few connections to government. The Road Information Program, a Washington, D.C., advocacy organization, compiled a pothole fact sheet in the mid-1980s that estimated the existence of 55.9 million potholes in the U.S., about one for every four Americans at the time. More than just cars can fall in a pothole or its more extreme cousin, the sinkhole. In Maine, the earth swallowed a snowplow; in Kentucky, a garbage truck; in Boston, a policeman on a horse.

The pothole remains a fixture in the national psyche. In 1978 Congress deliberated over an abortive $250-million pothole bill. The ultimate example of adapting defense technology to civilian life may be contained in a thin pamphlet issued by the U.S. Army Corps of Engineers in the early 1980s. It explained what it claims is the right way to repair a pothole: dig it out and fill it with a primer coating before dumping in the asphalt. In congressional testimony last year, Robert B. Oswald, director of research and development for the U.S. Army Corps of Engineers, said this technique was adapted from methods used by the Corps for fixing bombed-out runways. "It may not be the same problem, but there are synergistic effects to that problem," Oswald told the congressmen in an earnest tone.



CEMENT CARDHOLDER is J. Francis Young, a professor of materials science and civil engineering at the University of Illinois, who has made the study of cement his life's work. Young's university is one member of a National Science Foundation program that does basic research on the material.

Rattle, Rut and Spall

The Strategic Highway Research Program (SHRP) is a recently completed five- year, $150-million effort under the aegis of the NRC. It mounted what may have been the biggest assault ever unleashed against potholes, rutting and myriad other forms of damage to pavement. The SHRP distress identification manual describes pavements with an elaborate lexicon. A diseased roadway, according to this 74-page manual, may exhibit blow-ups, popouts, map cracking, spalling and rattling.

SHRP enlisted more than 200 scientists and engineers and put at their disposal a third of the program's funding, which included access to instrumentation, ranging from nuclear magnetic resonance imaging to ion exchange chromatography. Their mission: to perform the most detailed study ever of the chemical and physical properties of asphalt. Asphalt, a cheaper but less durable material than concrete, covers more than 90 percent of paved U.S. roads.

Asphalt pavements recently celebrated their centennial in the U.S. They have done so by exuberantly rutting and rattling. The problem can be traced to the Arab oil embargo. As refiners began to get crude oil from anywhere they could find it, the quality of asphalt declined. "The asphalt you get from Venezuelan crude is different from the asphalt you get from Alaskan crude," says Damian J. Kulash, SHRP's executive director.

Researchers at Pennsylvania State Uni-

"HOLEBOT," a robotic pothole-repair truck, was built at Northwestern University by James R. Blaha (*right*), Ralph Salle and others with a $1.2-million award under the National Research Council's Strategic Highway Research Program. The truck, which can fill a pothole automatically in less than 10 minutes, may sell commercially for $300,000 to $400,000.

versity employed a series of physical models to characterize how the stiffness of asphalt changes in different climates. SHRP has used these studies to work up specifications for asphalt refiners and for state highway departments to test how the material will resist cracking and other damage in environments as far afield as Florida and Alaska.

This effort also focused on simplifying procedures for using instrumentation that measures the stiffness of the asphalt binding material and its flow properties, or rheology—implements that had previously been confined to food-processing and chemical laboratories. Currently technicians use inexact devices that gauge asphalt texture by inserting a needle into the material to tell how well the asphalt will be compacted during construction. "It's about as primitive as sticking a fork into a cake," Kulash says. The new specification and testing technology will let a technician better analyze how well a grade of asphalt will last under different weather and traffic conditions, an impossible task with simple penetration tools. "Right now we run a pass-fail system—there's no incentive to make the honor roll," Kulash says. "Someone who has a material that resists moisture and lasts five times as long can get paid the same as someone who barely meets the standard."

During its five-year tenure, SHRP also initiated the first leg of 20-year pavement tests in all 50 states. It also let a contract to Northwestern University for a next-century answer to the pick and shovel, a prototype for a robotic pothole-repair vehicle. The machine was designed by James R. Blaha and others at Northwestern's Basic Industrial Research Laboratory. It can trim the jagged edges of a pothole, vacuum up loose material, dry and heat the surface, then blast patching material into the hole with enough force so that the material does not need to be compacted. Yet the "holebot" may not gain immediate acceptance. "If you replace two guys at the back of the truck earning $20,000 with one guy at the front earning $50,000, you don't save anything," says the FHWA's Pasko. "But if we don't try these innovative techniques, we'll be repairing things the same way 20 years from now."

Still, the scope of SHRP was not broad enough to satisfy some critics. When the program was first proposed in the mid-1980s, Fred Moavenzadeh, a professor of civil and environmental engineering at the Massachusetts Institute of Technology, asserted that the results of asphalt research could be no more than small improvements in existing technology. As an editor of a construction industry journal, Moavenzadeh has advocated looking to aerospace and defense research as a source of ideas for civil construction materials.

The combination of aerospace and civil engineering may strike some observers as a juxtaposition of opposites. Laboratory samples of metal-matrix composites can command thousands of dollars a pound. Concrete, at a few cents a pound, is cheaper by weight than chewing gum. But Moavenzadeh thinks that argument is specious. "The cost of the materials can be dropped because the performance requirements for infrastructure are not as stringent as those in aerospace, and the materials expenses are a small portion of what is paid for the structure over its lifetime," he insists.

If the cost issue can be overcome, composites would offer a number of advantages in rehabilitating or replacing the roughly 200,000 bridges that are reported to have some deficiency [see "Why America's Bridges Are Crumbling," by Kenneth F. Dunker and Basile G. Rabbat; SCIENTIFIC AMERICAN, March]. Composites would permit civil engineers who entered the profession because they enjoyed playing with Lego sets as children to fulfill a lifelong dream. A reinforced plastic bridge could snap together and be bonded with glue. A lane segment, weighing about one fifth that of a steel and concrete structure, could be carted on a flatbed truck and eased into place overnight.

Several industry-government collaborations will be jockeying for defense dollars that the Clinton administration has promised to divert to civilian projects. McDonnell Douglas Aerospace, the division of McDonnell Douglas that makes the AV-8B, an attack plane whose fuselage and wings are made entirely from composites, has joined a consortium of companies and research groups that hope to build bridges with a variant of the organic polymers used in military aircraft. Additional participants are Washington University and several other St. Louis–area companies and institutions. "There's a whole body of knowledge from aerospace that could enable the U.S. to do this much better than other countries," says D. James Dorr, McDonnell Douglas Aerospace's director of research and development for advanced materials and structures.

Also in the works is a composite and concrete bridge column. The composite,

BRIDGE OVER THE RIVER TAY was built last year in Aberfeldy, Scotland, from glass fiber–reinforced composites. Components of the footbridge's deck were glued together (*bottom left*), and then the deck was winched over the two A-frame towers by a student. The project demonstrated an advanced composite civil structure.

carbon fibers in an epoxy resin, would form a shell into which concrete could be poured. This design would eliminate the need for reinforcing the concrete with steel bars, which are subject to corrosion. By varying the amount of fibers and the direction in which they are placed, the material properties of the composite shell could be tailored to resist an earthquake's shear forces.

Zoltek, a St. Louis supplier of carbon fibers to McDonnell Douglas and other aerospace companies, is working on a process to manufacture up to one million carbon fibers at a time, which could reduce the cost from $20 per pound to less than $5. Although nominally an attempt at defense conversion, the consortium will require at least $20 million in federal money to get started. The initial funding may come from the Defense Department's Defense Advanced Research Projects Agency.

Another project that could come out of a collaboration with the aerospace industry is the most ambitious composite bridge to date. The congressional delegation from California succeeded in snaring $1.6 million in funds from the surface transportation bill for the University of California at San Diego to plan the building of a composite bridge that would connect two sections of the campus on opposite sides of Interstate 5. The bridge, supported by cable stays, would measure 450 feet in length and would carry full-size trucks.

Gilbert A. Hegemier, a U.C.S.D. professor of applied mechanics, said the approximately $50 million needed to build the bridge would also test concepts that could be exploited in the repair of existing structures and in the development of manufacturing techniques to make Lego-like construction a reality. "The hand-assembly methods for making composites for Stealth aircraft need to be automated," Hegemier says. Negotiations to form a consortium to construct the bridge with a few still unspecified aerospace firms have snagged on intellectual property rights and issues of legal liability. "So far there has not been any traffic-bearing bridge that uses advanced composites," worries Frieder Seible, a U.C.S.D. professor of structural engineering. "What happens if the bridge collapses? Who is liable when you develop new materials and systems?" The FHWA and the California Department of Transportation are also involved with U.C.S.D.

Whereas these projects are still on the drawing board, a British civil engineering firm last year supervised the construction of a footbridge to demonstrate more systematic design and manufacture of composite structures. Maunsell Structural Plastics in Beckenham, England, claims the cost of a bridge it built over the River Tay at a golf course in Aberfeldy, Scotland, is competitive with that of concrete and steel structures. "Our design system tries to use the minimum amount of material with the least labor input," says Peter R. Head, a Maunsell technical director.

GEC Reinforced Plastics translated Maunsell's design into 500 square yards' worth of interlocking hollow panels that were bonded together with epoxy cement to form boxlike beams. The plastic panels were made by pulling glass fibers through a die into which a catalyzed polyester resin was injected. Over a period of two weeks, a student crew from the University of Dundee locked and glued into place the 200-foot bridge deck that sits atop two upright A-frame towers.

Head makes the controversial claim that the streamlined construction system could bring costs into line with those of concrete and steel structures. The Aberfeldy project, at more than $400,000, is about 10 percent cheaper than a conventional bridge, he contends. Maunsell hopes this year to find partners to apply its technology to larger bridge structures.

Not everyone believes Maunsell's assertions. Organic composites may be 50 to 1,000 times more expensive than concrete. Also lacking is an industrial base to supply prospective builders of plastic bridges. "I'm convinced that cheaper composite structures can be done about 10 years from now," says Hota V. S. GangaRao, a professor of civil engineering at West Virginia University, whose research is geared toward using composite reinforcing bars to replace steel rods.

Although organic composites do not rust like steel, their durability remains to be tested in the field: constant changes in temperature, humidity, vehicle loading and stress in the connections between composite panels may shorten a structure's lifetime. "Look at Roebling's Brooklyn Bridge," says Joseph M. Plecnik, a professor of civil engineering at California State University at Long Beach who is doing research on composite bridge decks and cabling. "It's lasted 100 years. It's hard to say composites will last that long."

Other research from the aerospace industry may relieve the uncertainties of composites by policing flaws with advanced sensors. These sensors can be devised on the basis of experience gained with technology originally developed for aircraft. The ability to probe strains on bridge decks and columns could provide warning of damage or failure. The University of Toronto's Institute for Aerospace Studies wants to disperse optical strain gauges connected by a fiber network throughout a composite column or truss. A force applied to the sensors would cause a change in wavelength that could be noted by optoelectronic detectors.

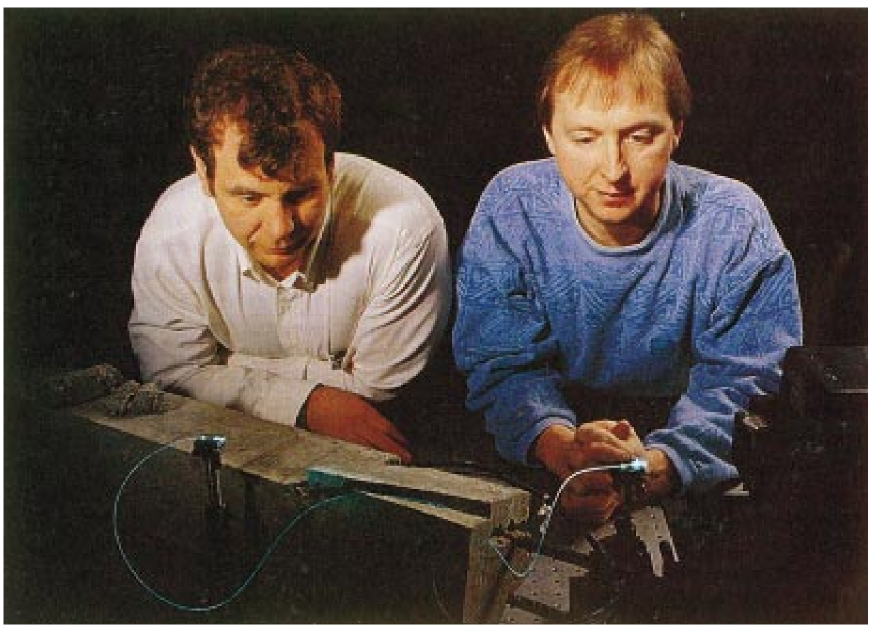
An Inside View

Networks of sensors connected by nervelike optical fibers presage "self-aware" buildings and bridges. The University of Vermont has just finished installing 2.5 miles of optical cabling into the concrete walls, ceilings and support columns of a new medical research center. Besides detecting tiny movements, the network is equipped with optical sensors to determine whether cracks are forming. The building is also outfitted with conventional electrically operated strain gauges and acceleration sensors. With a touch of hyperbole, an Associated Press headline proclaimed: "Stafford Hall is world's smartest."

The more than $11-million addition to the Burlington, Vt., campus will serve as an experiment to probe how a building should best contemplate itself. Development of a hairline crack should not cause the building to be evacuated. But Peter L. Fuhr and his colleague Dryver R. Huston are just beginning to wrestle with how to interpret the building's vital signs. "Most of the data are rubbish, so we're trying to figure out what is important," says Fuhr, a former National Aeronautics and Space Administration researcher who worked on a laser satellite communications system until his arrival at the university in 1985.

Fuhr thinks the best place for sensors may be within a dam wall or in seismically active areas, which experience large variations in stress. Last year Fuhr and mechanical engineering professor Huston helped to install 4.5 miles of fiber-optic cable on a dam on the Winooski River to monitor the water pressure on the dam, the flow of water through the spillway and vibrations from the generators. The computerized monitoring system automatically inspects the structure several times daily and is capable of placing an emergency call.

Bridges or roads might also respond to the world around them. Sending an electrical pulse through a coaxial cable set into a bridge column—or even sending a current through the steel reinforcing bars—could detect internal damage by measuring the time it takes the signal to return to a receiver, an idea being explored at Virginia Tech's Center for Intelligent Material Systems and Structures. Intelligent pavement has been investigated by the Westinghouse Science and Technology Center



STRUCTURAL NERVOUS SYSTEM has been fashioned from a network of optical fibers by University of Vermont professors Dryver R. Huston (*left*) and Peter L. Fuhr. The fibers can detect the presence of cracks and flaws in concrete.

in Pittsburgh. Researchers have applied a small magnetic field to asphalt seeded with ferromagnetic particles to measure thickness or detect voids in the hardened surface.

A bridge that heals itself was suggested by work at the University of Illinois's Architectural Materials Laboratory. There Carolyn Dry, a professor of architecture, is embedding treated glass or polymer fibers in fresh concrete. If a bridge column cracked or was ruptured by an earthquake or other stresses, the hollow fiber would break, releasing an adhesive that would halt further fracture. In another approach, a coating on polypropylene fibers would be eaten away by deicing salts, which would release an anticorrosion chemical to protect metal reinforcing rods.

Many bridge inspectors still drag along a chain until a drumlike sound testifies to the location of a hollow area underneath. The lessons of how to get technology from the university laboratories to sagging support columns or pothole-ridden roads might be better learned by crossing the Atlantic.

Before traveling to Europe in 1990 with other industry and highway officials, Wayne Muri, the Missouri engineer, had dismissed out of hand reports of the superiority of European roads. On the first stop, Stockholm, the trip from the airport to the downtown area convinced him otherwise. "The absolute smoothness and lack of any rutting in obviously very heavy traffic immediately sent up an alarm saying, 'Something different is going on here,'" Muri remarks. Because of more durable designs and pavement materials, European roads are built to last up to 40 years, twice as long as those in the U.S. They also carry a greater volume of trucks with axle loads that are 25 percent heavier than in the U.S., which causes them to absorb twice the wear. But the Europeans also pay double the cost or more for a better-quality, thicker roadbed.

What most impressed Muri was a type of asphalt developed in Germany to resist damage from studded tires. It is highly resistant to the rutting that occurs when heavy trucks bear down on asphalt pavement, a major cause of damage to roads in the U.S. The material, stone-mastic asphalt, which is now being tested by several state highway departments, consists of a network of crushed stones that are packed together until they touch. The stones are stuck together by a hard asphalt binding material that contains fibers or polymer additives.

Muri is, of course, an advocate for spending to build new highways. But the states that he represents are, in fact, one of the barriers to incorporating new materials into public works projects. Indeed, both laws and the attitudes of government officials and contractors are literally set in concrete.

Local officials can choose from about 16,000 road contractors nationwide for paving and rehabilitation. Many contractors are family-run outfits that survive through their willingness to abide rigidly by construction codes that have specified the same set of materials for decades. In comparison, France has nine large contracting firms that pave 80 percent of the country's major roads. Unlike most U.S. contractors, many of them retain a staff of in-house researchers, a prerequisite for adapting new materials technologies. "Most companies here don't think in terms of research and development," concedes Harvey M. Bernstein, president of the Civil Engineering Research Foundation (CERF). "They don't have the staff to understand it, and they are largely dealing with day-to-day problems."

Institutional Roadblocks

Those who could bring materials innovation to this market are often discouraged from even trying. Novophalt America, a U.S. subsidiary of an Austrian firm, has peddled a high-durability asphalt that contains a polymer additive. In a 1991 report the congressional Office of Technology Assessment documented how Novophalt met with resistance from public works officials, despite having amassed data on the pavement's performance from demonstration projects on highways and airport runways throughout North America. "The U.S. highway industry is very bureaucratic and extremely conservative in the way it operates," says Walter Tappeiner, Novophalt America's president. "It's hurting itself and the taxpayer."

Frustration does not stem simply from the buy-American strategy. Corning developed a quick-setting, high-strength cement from a glass, stratlingite-hydrogarnet. The material could set in only two to four hours and had other favorable qualities, such as low porosity and chemical reactivity. Frederic Quan, Corning's manager of research contracts, wrote a case history for CERF describing what he perceived as the obstacles Corning faced in penetrating the market for highway materials. Quan cited the industry's aversion to risk, the high liability costs, the low profitability and the lack of familiarity with innovation. "There doesn't seem to be a mechanism to fund leapfrogging technology with high payoffs," he wrote.

The difficulty of introducing technology into public works prompted the FHWA and CERF to establish a testing facility for new products. The Highway Innovative Technology Evaluation Center is being set up this year to provide a comprehensive performance evaluation, which could be used by the various states for product certification. "The purpose of the center is to provide a one-stop entry into the marketplace for innovators," says L. Gary Byrd, a consulting engineer who has done a number of studies for the NRC's Transportation Research Board and the FHWA.

But plastic bridge decks and stone-mastic asphalt highways will require that both government and industry devote even more spending to basic materials. Although the 1991 Surface Transportation Act provides more funding for highway research, the total amount is still only 0.25 percent of that spent on highways, little more than a twentieth the comparable percentage for the aerospace industry.

Calculations of the compressive strength in concrete may appear to have little bearing on a nation's competitive standing. But one premise of Clinton's election platform was that the nation that builds the best transportation and communications networks will prevail. Just how much economic benefit in increased productivity comes from investing in infrastructure is a subject of heated debate. But even the most hardened cementheads do not doubt that building more durable roads may help, for instance, a company that attempts to set up a streamlined production system in which parts are delivered to a factory only as needed.

Muri recalls that three or so years ago the governor's office received a not totally facetious bill for more than $50,000 from General Motors's huge plant in Wentzville, Mo. The bill was for losses suffered when suppliers' trucks were unable to reach the plant because traffic on Interstate 70 had slowed to a near halt for emergency road repairs. The delay had stalled the plant's just-in-time production system.

Durable, safe roadways can be built only if materials scientists learn to control the properties of a dull-gray paste that looks like a pointillist masterpiece on a Cray computer. "When somebody says we want to do research on concrete, somebody else usually points to the roadbed and says we already do it," says NIST's Schwartz. "This ignores the fact that potential for improvement in this area is absolutely enormous. A small investment over the next five years could make an impact on all investment in infrastructure over the next 25 years."

High Technology

Will a new venture finally launch industry into space?

For so long that it has almost become boring, the National Aeronautics and Space Administration has been trying to sell corporations on adding low-earth orbit to their roster of research facilities. But neither hard nor soft sell has convinced companies to ante up more than token dollar sums, and commercial space research to date remains pretty much pie in the sky.

Now Westinghouse and two partners believe they can get private space research off the ground—with more than a little financial help from NASA. Sometime this spring the first Commercial Experiment Transporter (COMET) mission is scheduled for takeoff on an unmanned spacecraft from a launch pad at NASA's flight center on Wallops Island, Va. COMET will carry 11 experiments 300 nautical miles into low-earth orbit. Thirty days later a recoverable capsule will deposit some of them in the Utah desert; others will orbit for more than two years before being destroyed by reentry of the remaining part of the spacecraft, the service module.

Westinghouse hopes this relatively low cost, hire-a-spacecraft approach to marketing will spur industry demand. Many of the other 32 U.S. commercial launches since 1982 carried communications satellites aloft. These missions required researchers to provide their own ground station and occupy themselves with myriad planning details. Even before COMET flies, Westinghouse has established an independent launch program, called Westinghouse Space Transportation and Recovery Services, or WESTAR, and begun to offer, for $35 million, a full package of launch services to any takers. "This is one-stop shopping," says Thomas E. Haley, WESTAR program manager. "You bring it to one organization, and we do all of the launch services as well as orbit and recovery."

Despite the private-sector talk, NASA is kicking in $85 million of the estimated $100 million to $200 million in start-up costs for three COMET launches. The space agency is also supplying the funding to get the experiments on board the spacecraft. Indeed, neither Westinghouse nor its partners, EER Systems and

ROBERT PROCHNOW

PACKAGE TOURS to low-earth orbit are being marketed to industry, government and academia by Thomas E. Haley, manager of Westinghouse's WESTAR program.

Space Industries, would have undertaken COMET without NASA support. "NASA can take credit for jump-starting this industry," Haley says.

The 450 pounds in payload that make up the experiments on the first COMET flight are commissioned with the Centers for the Commercial Development of Space. These institutions, most of them affiliated with universities, were set up, starting in 1985, by NASA's Office of Commercial Programs. They still rely on the agency as their biggest source of funding. (A similar scheme is being used by Spacehab, a privately built experimental module that will be flown on the space shuttle. Eight flights have been scheduled at six-month intervals; the first is set for April 28.)

Among the experiments COMET will carry are one concerning protein crystal growth—which involves Eli Lilly, Du Pont, Merck and others—and demonstrations of telecommunications by Motorola. Motorola is working through the space commercialization center at Florida Atlantic University to perform experiments that will help define technical requirements for the low-earth orbit element of a worldwide cellular telephone network.

COMET will provide longer exposure to a low-gravity environment than has the space shuttle or any other U.S. commercial launch service. The recoverable capsule that will descend to the military testing grounds in Utah's Bonneville Salt Flats, the first for a U.S. commercial launch, will remain in orbit for a month. Most shuttle journeys last for a week or so; previous U.S. commercial space flights on expendable rockets have supplied only a few minutes of exposure to microgravity.

Space pioneers Max Faget and C. C. Johnson of Space Industries designed the recoverable capsule using proved 30-year-old technology; the spacecraft resembles the manned capsules of the Mercury and Gemini programs.

Westinghouse built the expendable service module; it will contain four of the 11 experiments and will stay in orbit for two years. Both capsules will be lofted into space by a Conestoga 1620 booster rocket supplied by EER, a launch services company that built the pad on Wallops Island. EER also procures engines and components and assembles the Conestoga rockets.

Although slightly more than a month remained before the tentative liftoff date of March 31 (which is about six months behind schedule), WESTAR had yet to find a buyer for an extra 100 pounds of payload for experiments that would be funded entirely by a company, university or government agency. An independent WESTAR flight might be scheduled for the fall of 1994, Haley says, although many, if not all, of the experiments would probably be done for the Department of Defense or another federal agency. "There's interest in industry and academia, but they aren't ready yet," Haley remarks.

Indeed, few researchers are now convinced that there is a big payoff in microgravity investigations to grow crystals of proteins for drug research or semiconductor materials for electronics firms. After a decade of experiments, little more than one of five protein crystals produced in space was superior to those made in terrestrial laboratories, according to a recent article in *Nature.*

Moreover, a National Research Council report last summer noted that NASA's "zeal" to encourage the commercial use of space may lead to lower-quality industry experiments that are conducted without the peer review to which university experiments are subjected. Robert F. Sekerka, a professor of physics and mathematics at Carnegie Mellon University, characterized industry involvement in this program as opportunistic: "If government weren't putting up the money and industry were paying full freight, how much would they put in? The answer is, probably nothing."

WESTAR, however, hopes that if the price is right, industry and academia will eventually get on board. It plans to charge $1 million for a 10-pound payload that will be shoehorned in with other experiments. This is still enough to carry out a small protein growth experiment. The ability to repeat experiments at predictable intervals may enable researchers to discover the best uses for microgravity, Haley says.

If the effort is successful, Haley predicts that WESTAR may fly as many as 10 missions a year by the turn of the century, as costs steadily drop. "A lot of scientists want to fly something in January and then bring it back and tweak the experiment and then send it up in August to see the effect of the change in parameters," Haley says. "By doing that, you may be able to short-circuit many years of ground-based experimentation." —*Gary Stix*

Advertising Space

If Michael F. Lawson gets his way, a familiar billboard pitch, "This Space Available," will soon take on new meaning. As the chief promoter for the Commercial Experiment Transporter (COMET), Lawson is peddling a wealth of marketing angles.

For $500,000, almost three quarters of the price for 30 seconds of television advertising during the Super Bowl, Columbia Pictures was nearing an agreement on the right to advertise an upcoming movie on the side of the spacecraft for the COMET mission, scheduled for liftoff into low-earth orbit this spring. "This mission will start the ball rolling to get corporate America involved: this is a true commercial event," says Lawson, whose five-year-old Atlanta-based firm, Space Marketing, has become COMET's official ad sales representative.

With the outside of the rocket reserved, Lawson was still trying to peddle a berth inside the capsule that is to return to the earth after a month. For a million dollars or more, a corporate sponsor could procure the right to have commemorative gold coins or a bag of corn chips accompany the commercial cargo, which contains pharmaceutical, telecommunications and other experiments.

Lawson's firm has coordinated such events as sending full-scale models of spacecraft to industrial trade shows and state fairs. Now it has become involved in trying to convince soft-drink, fast-food and insurance companies that the surface of the mission's Conestoga rocket is the best idea in outdoor advertising since skywriting. Putting a camera in the capsule on this or a future mission might lead to the first suborbital commercial. "You could have a bag of Fritos floating around in space," Lawson suggests.

The rocket will be decorated with a cylindrical billboard that will be visible for only a few fleeting moments. But the promotion should encourage hordes of space buffs to dial 1-900-9-ROCKET to tape a brief message, for a mere $3.50, that will be beamed both out into space and back to the earth from the orbiting capsule. "A message for the future of mankind," as Lawson puts it. A lucky child may even get to push the button to launch the rocket.

Of course, the marketing prospects are as vast as space itself. The Sea of Tranquillity could become the extraterrestrial equivalent of the Strip in Las Vegas—or at least the next Epcot Center. Where will this all end? Lawson answers with a question. "Will there be a McDonald's on the moon in 15 or 20 years? I think that's a little farfetched," he concedes. "But astronauts may be there consuming McDonald's products to gain sponsorship that reduces the cost of space missions." —*G. S.*

TOMMY THOMPSON *Black Star*

A LITTLE PIECE OF SKY is what Michael F. Lawson is promoting as chief ad salesperson for COMET.

Virtuous Viruses

Agents of disease are being turned against cancer

According to current estimates, up to a fifth of all cancers are caused at least in part by chronic viral infection. The culprits include human *T* cell leukemia/lymphoma virus, human papilloma, Epstein-Barr (EB) and hepatitis B. HIV, the AIDS virus, can probably be added to the list. Although using antigens from those viruses to mobilize the immune system to fight cancer has long been considered a possibility, only recently have researchers begun to report positive results.

A British biotechnology start-up, Cantab Pharmaceuticals, has developed a potential immunotherapy for cervical cancer based on a viral protein. And several companies have expressed interest in bringing to market a vaccine against EB virus, thought to be a cause of both nasopharyngeal cancer and Burkitt's lymphoma, a cancer of the lymph glands common among children in Africa. Other companies may follow suit.

MICHAEL MANNI

IMMUNOTHERAPY for cervical cancer based on viral proteins was developed by Cornelia S. McLean of Cantab Pharmaceuticals. Human trials will begin soon.

Cantab's product may be the first to be based on a demonstration in animals that viral proteins are able to invoke an immune response that can prevent and induce regression of tumors. The company, which is devoting its efforts to developing therapies, was formed in 1989 by Alan J. Munro, the former head of immunology at the University of Cambridge, and Abingworth, a venture capital group. A formal agreement with the university grants Cantab exclusive license to commercialize immunology research at the university.

The work that paved the way for Cantab's cervical cancer product was done in cattle by M. Saveria Campo and her colleagues at the Beatson Institute for Cancer Research in Glasgow and at Glasgow University. When the researchers injected cattle with a protein from bovine papillomavirus, known as E7, it strengthened the animals' ability to fight off tumors in the mouth, which develop when cattle are injected with the virus itself. The E7 treatment stimulated the cellular immune response, a major arm of the immune system, more strongly than did the actual virus. Although its precise role is not known, it is probably no coincidence that E7 interacts with an important cellular protein that controls growth.

Campo's results, which will be published shortly in the *Journal of General Virology,* made a powerful case for investigating whether there is a similar effect in humans. Bovine papillomavirus is genetically quite similar to human papillomavirus type 16, which causes lesions of the cervix that can become malignant. Cancer of the cervix is estimated to kill more than 300,000 women worldwide every year—almost as many as breast cancer.

Cornelia S. McLean, then in the department of pathology at Cambridge, proved that proteins from the human papillomavirus can also induce a strong immune response. McLean, now at Cantab, used as a vehicle vaccinia virus, which has widely been used for smallpox vaccination. She injected mice with vaccinia that had been modified to produce the E7 protein from human papillomavirus type 16 and then monitored the reaction that occurred when she grafted E7-producing cells onto the mice's skin. Pretreatment with the modified vaccinia caused the mice to reject the grafts faster than they normally would, indicating that the animals had been sensitized to the protein. (McLean's research was published in the February issue of the *Journal of General Virology.*)

Cantab has teamed up with Leszek K. Borysiewicz, an immunologist at the University of Wales College of Medicine in Cardiff, to see whether the E7-producing vaccinia generates cellular immune responses in humans. Initially about 10 patients with advanced cervical cancer will be injected with the modified vaccinia, which also produces another viral protein called E6. The trials will start "in the near future," according to Borysiewicz. Although there is a risk associated with using live vaccinia, Munro explains, experiments suggest that a live virus can produce a far stronger immune response than can a purified protein, which is not "seen" by the immune system in the same way. "We're using the sophistication of the immune system," he says.

Campo's work on bovine papillomavirus was supported by the London-based Cancer Research Campaign, which is also hoping to develop a vaccine that might protect against papillomavirus infection. But in January the organization captured headlines by announcing the development of a different preventive vaccine, one aimed at EB virus.

EB virus is believed to cause several types of cancers, probably in combination with other factors. It infects about 90 percent of the world's population, but in developed countries it usually causes either no illness at all or a transient malaise called glandular fever. The virus is, however, often found in malignant cells in Hodgkin's disease. In Africa, it is found in Burkitt's lymphoma cells, and in China, it is thought to be responsible for about 50,000 deaths a year from nasopharyngeal cancer.

The new EB vaccine, unlike Cantab's cancer therapy, is a synthetic version of a protein normally found in the virus. The culmination of more than two decades of effort, the product is based on research by Andrew J. Morgan of the University of Bristol and John R. Arrand and Michael Mackett of the University of Manchester, along with collaborators at the University of Birmingham.

The vaccine, which is made in mouse cells, has been shown to prevent a lymphoma in one type of monkey. If it works in humans, it could become the first vaccine designed specifically to prevent cancer, although hepatitis B vaccines, which have been used for more than 10 years, almost certainly prevent liver cancer.

John Green of the Program for Appropriate Technologies in Health in Seattle, which is collaborating with an international task force on EB virus, says several companies have expressed interest in perfecting the EB vaccine. Yet the effort faces many difficulties. Perhaps the chief obstacle is that the countries most in need of such a vaccine are poor and have other pressing health priorities. Hepatitis B vaccine, for example, has not been widely used in developing countries.

Morgan of the University of Bristol believes the work on EB virus will eventually lead to a therapeutic vaccine for nasopharyngeal cancer or lymphoma as well as a preventive vaccine. It may, however, be based on a different protein from the one used in the prophylactic product. In people infected with EB virus, the immune system shows increased activity shortly before nasopharyngeal cancer becomes clinically apparent, Morgan notes. A vaccine that stimulated the same activity might, he suggests, be a candidate for a therapy.

Until recently, Morgan says, “people in the vaccine world didn’t appreciate the complexity of the immune response and the fact that it can be selectively altered.” Now it seems viruses have given Cantab and others a raft of promising strategies. *—Tim Beardsley*

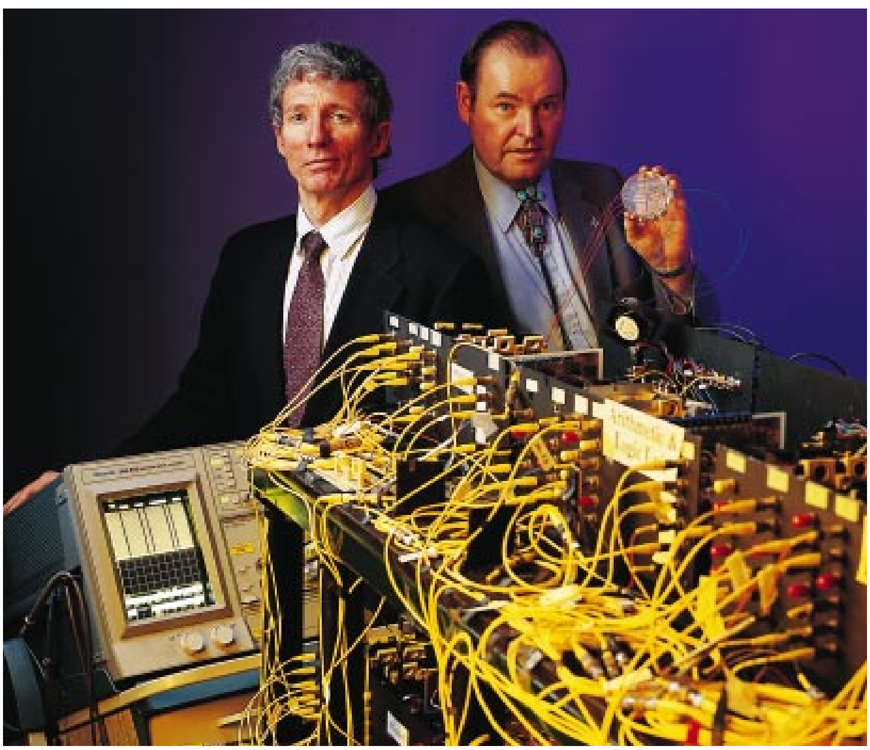
Light Motif

An optical computer stores its program in space-time

Digital computing took a giant stride in 1948, when workers at the University of Manchester completed Mark I, the first all-electronic computer to hold its own program in memory. The infant technology of optical computers took a similarly bold if shaky step early this year when Harry F. Jordan and Vincent P. Heuring of the University of Colorado’s Optoelectronic Computing Systems Center demonstrated an (almost) all-optical computer that stores and manipulates its instructions and data as pulses of light.

Like the Mark I, the bit-serial optical computer (BSOC) is big and dumb, built to prove a principle in architecture rather than to test new devices or perform real work. The computer can do little more than basic switching, counting and arithmetic, and its memory holds a mere 128 bytes—fewer than there are characters in this sentence. But it performs these tasks in a way that differs radically from any computer that has ever been built before.

Virtually all electronic digital progeny of Mark I have used flip-flops—devices in which relays, vacuum tubes or semiconductor circuits take and hold a charge—to store data. In such machines, information remains physically confined. But optical analogs of the electronic flip-flop are still very slow. So the BSOC uses dynamic memory instead: the infrared pulses that make up its bits zoom endlessly around a circuit of optical fibers at the speed of light. “The information is always on the move; in a sense, it is being stored wherever it happens to be in space-time,” Heuring explains. “The trick is to arrange matters so that you get interaction of information in the same space at the same time.”



GEOFFREY WHEELER *Black Star*

FIRST OPTICAL COMPUTER to store its own program does so dynamically in long coils of optical fiber. Harry F. Jordan and Vincent P. Heuring of the University of Colorado used lithium niobate switches to control 12-foot “bits” of laser light.

The BSOC accomplishes this feat of synchronization by using a clock running at 50 megahertz and a path for the light that spans carefully measured distances between each switch. Just before each tick of the clock, control pulses throw switches that route arriving bits from path to path, in and out of delay lines made from meter-long coils of optical fiber. Whereas an electronic microprocessor would fetch information it needs from a memory address, the BSOC’s optical processor waits for the data to come to it. Counters keep track of what is where.

It is the switches that make the system nearly all-optical. AT&T, which supplied the Colorado group with the devices, constructed them by diffusing titanium through the surface of lithium niobate, a transparent solid, to form two channels for the light. When a voltage is applied, the pathways are parallel; when it is dropped, the channels cross. To throw the switch, a control pulse must be converted from light to electricity and then amplified. Jordan uses a conventional desktop computer connected to the switch amplifiers to program the BSOC and to display its results, although in theory a fiber-optic interface would work as well.

So the BSOC does need electricity to run. But by relegating electrons to a minor role and doing away with static storage altogether, the researchers have simplified the structure considerably. "This may be the first time you can actually see an entire computer—memory registers, input/output, arithmetic and instruction units, everything—on a single overhead projector slide," Heuring boasts. "Normally that's impossible: there are so darn many flip-flops and latches and gates and buses and so on."

Dynamic memory has other benefits. It makes the machine speed-scalable, for example. Because the design is based on the time it takes for light pulses to travel from one processor to the next, the same blueprint will work at half the scale and twice the clock rate. This is fortunate, since as it stands the BSOC sprawls across an entire desktop. Jordan and Heuring are now integrating the BSOC on a disk of lithium niobate attached to a disk of silicon or gallium arsenide that will contain the switch electronics. The palm-size computer should be about one four-hundredths the size and run 400 times faster, at 20 gigahertz.

Simply increasing the speed without changing the size also produces interesting results, Jordan discovered. "If your switches are fast enough, you can double the clock rate and end up with two independent machines running on the same hardware 'in parallel'—actually interleaved in time." At 100 megahertz, in other words, the BSOC develops a split personality.

Dynamic memory does have its price, however: capacity and efficiency. Storing information in space-time requires a lot of one or the other. And the longer the circuit, the longer a processor may have to wait for its data to swing by. Heuring admits this price limits the complexity of potential applications. "A BSOC would be ideal for an AT&T or cable TV switching network that routes and rearranges serial information coming in from five cities and going out to five different cities," he says. "But you would not do word processing on it."

In fact, the first practical optical machines will almost certainly be hybrids, with optoelectronic switches and some kind of static storage. Recognizing this, Jordan and Heuring have already moved on to a parallel-processor design that trades coils of fibers for empty space and serial switches for square arrays called optoelectronic integrated circuits (OEICs). The OEICs have miniature photodetectors on one face and microlasers on the other. They would stand on edge like so many dominoes, drawing electric power from their base and receiving control pulses from above. Heuring's new design would use computer-generated holograms to rearrange the bits from one OEIC to the next.

The Colorado researchers see a big payoff if they can build a computer that uses OEIC chips containing a million or more switches. "Normally, video has to be scanned in serially. That's the big bottleneck with virtual reality and a lot of computer graphics and animation," Heuring points out. "With OEICs, we could in principle work on an entire frame of video—or an entire computer-generated polygon—all at once."

Neither researcher expects to have a finished product for five to seven years, even assuming they find funding. But one year's work should prove or disprove the principles involved, "and we could have some preliminary results within three years," Heuring predicts. If the history of electronic computers has a lesson, it is that such predictions tend to be far too pessimistic.—*W. Wayt Gibbs*

Putting the Brakes on Tankers

The wreck of the *Braer* off the Shetland Islands in January is the latest in a series of costly calamities caused by oil tankers losing power on the high seas. The vessel went aground without dropping its anchors, but they would probably have made little difference. Anchors are useless for stopping a laden vessel in a strong current because the brakes on the windlasses overheat; the strongest anchor chains will snap as they are suddenly drawn taut.

A British marine engineer, Keith Ridgway, believes he has invented a system that can prevent similar tragedies in the future. In the late 1980s Shell International sponsored him to design a windlass with a special hydraulic motor that enables it to keep a constant tension on the anchor chain, playing it out when forces are strongest and hauling it in when they abate. A ship adrift tends to move sideways. When an anchor is dropped and sets, the vessel will swing bow-on to the current. Ridgway's computer simulations indicate that his windlass can absorb the energy of the ship as it swings about and so bring it to a halt without breaking the extra-long chain. The chain can then be locked.

Shell patented Ridgway's idea but never completed the detailed design work because it sold many of its large tankers, and the company has relinquished the patent. According to Ridgway, who is now at the University of Sheffield, tanker owners have been unwilling to spend the $400,000 he estimates it would cost to retrofit a supertanker, even though technical experts have hailed the idea. "Nobody has said this wouldn't work," he says.

Ridgway says the *Braer* disaster prompted him to look again at how to get his system fitted onto a large civilian ship (some military assault ships have incorporated the principle, he notes). Suitable hydraulic motors are already in production for other purposes. —*Tim Beardsley*

MATTHIAU POLAK *Sygma*

WOULD an advanced anchor system have kept the* Braer *off the rocks?

Global View

A map of the earth that provides the real picture

To Tom Van Sant, founder and chief executive officer of GeoSphere Project in Santa Monica, Calif., seeing is believing when it comes to understanding the complex global environment. "We want to be able to show people the relationships and dynamics between different earth systems without the need for translation," he says.

Van Sant is delivering his message with the help of remote satellite imagery, computer graphics and a three-dimensional model of the earth. The "reality globe," as he refers to it, is a seven-foot sphere bearing the image of the earth as it would appear from space without any clouds. The globe serves as a three-dimensional screen for projections depicting the earth's changing environment, from patterns of ocean circulation and continental drift to tropical deforestation and global climatic change. The company is also outfitting the globe with internal fiber-optic lights to enable viewers to watch such scenes as the world's cities at night and global seismic activity.

The reality globe is the centerpiece of what GeoSphere Project leaders call the Earth Situation Room, or ESR. Because an individual can see only half of the globe from any angle, the ESR also contains wall screens on which the whole earth image and the overlays can be projected. The monitors allow users to zoom in and view any region of the earth in detail. Ultimately, Van Sant envisions an international network of ESRs serving as visual libraries of earth systems for regional television networks, schools and research organizations. "This is the most advanced scientific visualization project in the world," Van Sant says.

It is also an idea that is catching on. The four-year-old company installed its first ESR at the World's Fair in Seville, Spain, last May and has contracted to build seven more this year. But GeoSphere is not restricting itself to selling ESRs, which can carry a price tag of up to several million dollars. It is also adapting the digitized data bases and cloudless image of the earth for desktop computers and shelf-top globes. Last June company officials demonstrated their first interactive computer laserdisk globe at the Earth Summit in Rio de Janeiro. They also plan to issue similar interactive software on CD-ROM. Both computerized versions will sell for less than $100, making them affordable for schools and individuals.

SURMANI *Van Sant Studio*

GLOBAL MAP based on satellite images is the world's "most advanced scientific visualization project," says GeoSphere Project's CEO and founder Tom Van Sant.

Central to GeoSphere Project's emerging success is the cloudless image of the earth from space. Designed in collaboration with researchers at the National Aeronautics and Space Administration's Jet Propulsion Laboratory in Pasadena, Calif., the image is a seamless mosaic of hundreds of pictures of the earth taken by the National Oceanic and Atmospheric Administration's *TIROS-N* satellites.

Color-coding information for the image's nearly 600 million pixels requires 2,400 megabytes of hard-disk storage capacity, and a bank of roughly two football fields of televisions would be needed to display it at its full resolution of four kilometers. "It's more science than the artist would want and more art than the scientist prefers," says Lloyd Van Warren, a software engineer at the Jet Propulsion Laboratory.

That laboratory's team is now working to update the current image to a sharper, one-kilometer resolution. At four-kilometer resolution, for example, zooming in on the Aral Sea, which straddles Soviet successor states Kazakhstan and Uzbekistan, clearly reveals that it has been severely depleted by irrigation. At one-kilometer resolution, farms that have prospered at the sea's expense come to life.

On the low-tech front, GeoSphere recently reached a licensing agreement with two globe makers to manufacture more traditional globes with a new twist. The minigeospheres will of course feature GeoSphere's cloud-free view of the earth. But interchangeable clear acrylic overlays will depict such features as political boundaries, animal migration routes, energy resources and population distribution.

GeoSphere is also working to acquire data bases and convert the information into visual overlays. So far data have been provided by more than 50 international organizations, U.S. federal agencies and environmental organizations. Among them are the United Nations Environmental Program, NASA, NOAA and the Worldwatch Institute.

In many cases, the collaborative organizations are given copies of the finished visuals in exchange for their hard-won data. In others, getting the message out is reward enough. "GeoSphere is helping to fulfill our mission of making the scientific information we collect comprehensible to the public," says Douglas R. Newman, who is NOAA's chief of constituent affairs. It also might just inspire a few children to become scientists. *—Robert Service*

Frequent Flaps in the Deregulated Skies

A little less than 15 years ago Alfred E. Kahn and his colleagues on the now vanished Civil Aeronautics Board (CAB) began dismantling the regulations that dictated what cities airlines could serve and what prices they could charge. The idea, of course, was to unfetter the airlines for the benefit of both shareholders and customers. "I always said that this comes as close as we can get in the social sciences to an experiment," remarks Kahn, who is now a professor emeritus at Cornell University.

The current results are troubling. Consumers have profited at the expense of the airlines, and calls for reregulation are being heard in Washington. Average ticket prices are estimated to be 10 to 20 percent lower than they would have been under regulation. At the same time, the industry has lost about $8 billion in the past three years.

There is a long list of reasons why the "experiment" hasn't panned out quite the way free marketers hoped it would. Key among them are the ways in which economists simply failed to capture the complexity of the real world. "It's not entirely accurate to say that economists had a clear model and then tested it," comments Severin Borenstein, a former CAB colleague of Kahn who is now at the University of California at Davis.

The resemblance of the airline industry to a free market, Borenstein says, did not dictate that the rules of perfect competition held sway. "Economists take a model of perfect competition and slap it on a market that looks pretty competitive—but really isn't," he argues. "Hard scientists realize that tiny aberrations from the model cause huge distortions in the results."

Consider, for example, the "contestable markets" hypothesis: economists postulated that even if only one airline served a particular route, it would still charge low prices because to do otherwise would be to invite competition. The hypothesis was based on the notion that aircraft are mobile resources. Unfortunately, theory ignored the need for airport gates and takeoff and landing slots, all of which may be scarce on desirable routes—but which incumbent airlines "own." As a result, prices on single-carrier routes are on average 10 percent higher than prices on routes served by two carriers, which in turn are on average 10 percent higher than those on routes served by three or more.

Economists also thought significant economies of scale did not exist in the airline industry. Consequently, they failed to foresee the overwhelming marketing advantages of hub-and-spoke systems. Passengers tend to choose the carrier offering the most flights to a given destination, and so an airline can keep others out of its markets by dominating a hub. For instance, Atlanta is virtually the fiefdom of Delta Airlines.

Other economies of scale include frequent-flier programs, which instill brand-name loyalty in passengers, and "commission overrides," which give travel agents similar incentives for steering business to a particular carrier. Computerized reservation systems, which can influence ticketing decisions by skewing the presentation of flights, also work to the advantage of the large airlines that own them.

> *"The self-destructiveness of airline management belongs in a Monty Python movie."*

Misunderstanding the structure of the air-transport market directed economists toward simplistic conclusions. But even more unrealistic was the assumption that market participants would work rationally to maximize their profits. Airlines in pursuit of market dominance (and the resulting monopoly profits) have repeatedly slashed fares below the levels required to recoup their operating costs. "The self-destructiveness of airline management belongs in a Monty Python movie rather than on the business pages," observes Richard C. Leone, chairman of the Port Authority of New York and New Jersey, the organization that oversees JFK, La Guardia and Newark airports.

The government has also thrown grit in the market mechanism. The Reagan administration chose not to enforce antitrust laws, allowing dubious competitive practices to proliferate. And now that winners and losers appear to have emerged, the government has been reluctant to let major carriers die. Bankruptcy laws permit an insolvent airline to continue flying while negotiating with its creditors. The resulting cost advantage, charges American Airlines chairman Robert L. Crandall, undercuts fares for the entire industry.

Economists and policymakers are convinced that the current situation is untenable and that the U.S. needs at least four major airlines to ensure some degree of competition. But they differ bitterly about what to do next. One idea in vogue, especially with some airlines, is, not surprisingly, reregulation. Its most vocal proponent is Paul S. Dempsey of the University of Denver. He advocates government guarantees to help carriers borrow money to buy new planes, "modest" price controls and partial rationing of airport gates and landing slots.

Others argue that significant government interference with prices may in fact turn the tables by aiding airlines and hurting consumers. They propose only a minimum of intervention to smooth out the kinks in the market. One bill before Congress aims to eliminate biases caused by computerized reservation systems. Other policymakers are rethinking how the government apportions gates and landing rights. Vernon L. Smith of the University of Arizona suggested early in the course of deregulation that air carriers be forced to bid for slots rather than receiving them by fiat. No simple solution is yet in sight.

Some moves favored by economists are politically unpalatable, such as the banning of frequent-flier programs, which often reward business travelers for spending their employers' money. Equally vexing is the notion of boosting competition by opening the U.S. domestic market to foreign carriers.

Meanwhile most airlines show little sign of abandoning their pursuit of market share at the expense of solvency. But a beacon glimmers in Dallas, where Southwest Airlines operates at a consistent profit by offering only point-to-point flights (which are cheaper to operate than hub-and-spoke routes), eschewing fancy in-flight service, avoiding rapid expansion and keeping fares consistently low. Their rule, according to Leone? "We won't send planes if we can't make a profit on the flight." Now, there's an idea that might fly.

—Elizabeth Corcoran and Paul Wallich

The Rise and Fall of the Lunar M-pire

While rummaging through a storeroom in the Louvre last fall, curator Jean-Jacques LeMaire discovered a carton containing some unpublished papers of Jules Verne. Unfortunately, a mouse had nibbled through the papers many years before. LeMaire and his colleagues are desperately trying to piece the material back together. So far they have restored a previously unknown chapter from Verne's 1865 classic Round the Moon. *LeMaire has kindly agreed to publish the material in* Scientific American.

Major Elphinstone paced back and forth inside the dome of the Baltimore Observatory as his associate J. T. Maston peered through the eyepiece of the observatory's main telescope. "I don't see them yet," Maston said nervously.

"May I look?" Elphinstone asked.

"Of course," Maston replied, picking up a large notebook of mathematical calculations. "If the quantity of guncotton had been misjudged by a mere 1 percent," he said, "then there could be a delay of several hours. I believe we should not yet conclude that some terrible fate has befallen Barbicane, Nicholl and Ardan."

"Quite right," Elphinstone sighed. "One must think positively."

The Major shivered. "It is devilishly cold out here on Long's Peak! I suggest we retire to the main building for an hour and then return to the observatory to renew our search."

"I am a little concerned lest any news come through on the ticker-tape machine." They stared at the tiny machine that had been installed in a corner of the dome.

"Any news will be recorded on the tape, Maston. We can easily check it when we return."

Maston reluctantly agreed, and the two gentlemen walked quickly to the warm lounge of the building. A servant was summoned to bring them coffee and sandwiches.

The Major noticed a map of the world on the wall—German territory in orange, French in green, American in purple and the British Empire in pink.

"Soon we shall redraw that map," Elphinstone boasted.

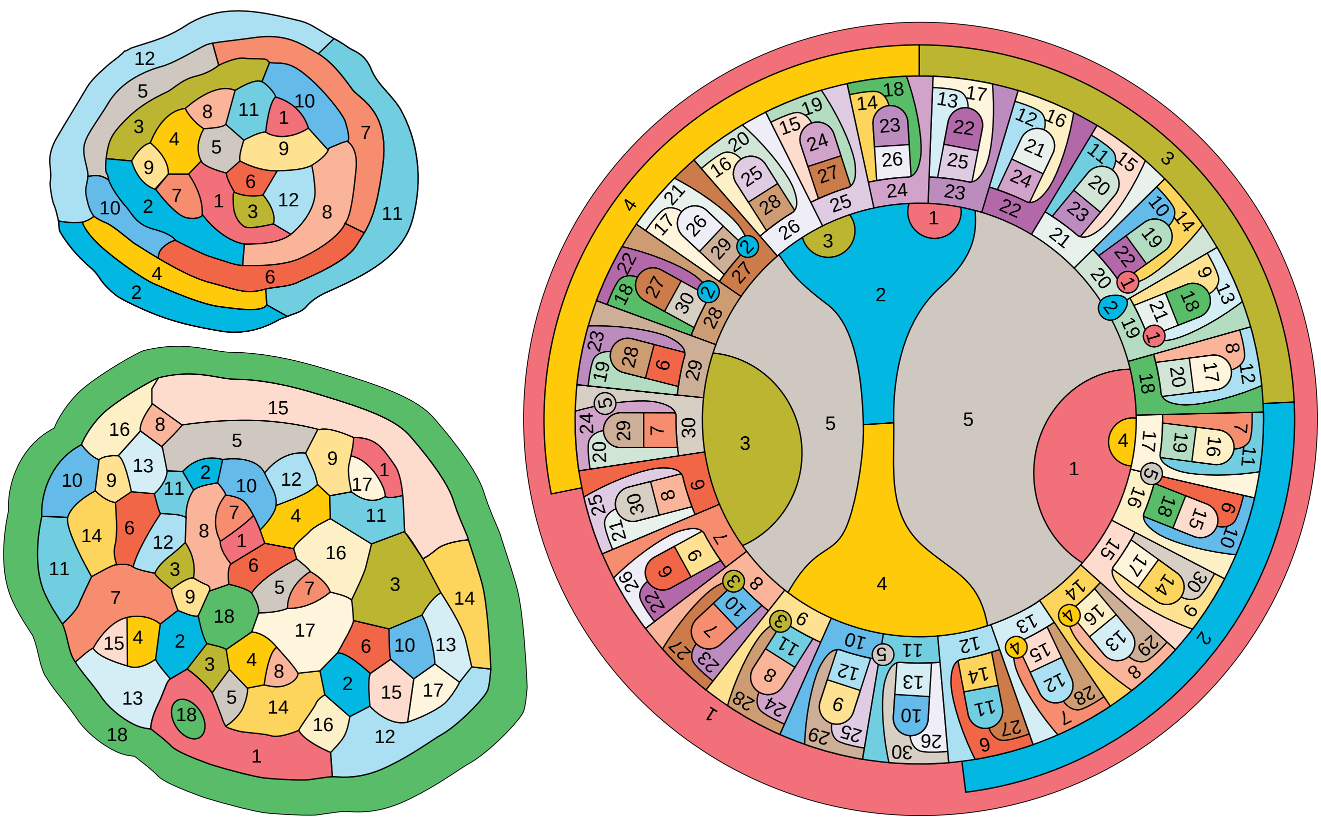
"Pardon?"

"We will adjoin a map of the moon, and it, too, will be colored purple. An American colony on the moon. The sun may never set on the British Empire, but the American Lunar Colony will rise above every nation on the earth."

"Ah," said Maston, who had little interest in building an empire. "There might have to be some green, too."

"Excuse me?"

"On the map of the moon," he said



CARTOGRAPHER'S NIGHTMARE:** If one were to make a map of several empires, each of which consists of two countries, what would be the maximum number of colors needed to guarantee that no two adjacent countries would be the same color and that countries in the same empire would be the same color? The map at the top left requires 12 colors, and in fact, no map of empires of two countries ever needs more than 12. For empires consisting of three countries, the maximum number of colors needed is 18* **(bottom left), and for empires of five countries, the total is 30* **(right)**.

apologetically. “Don’t forget that Ardan is French.” Seeking to change the subject, he wandered over to the map. “I wonder why cartographers use so many colors? There must be at least a dozen.”

“So?”

“I was told not long ago by a relative, a mathematician at Harvard, that someone named Percival Heawood has proved that any map on the surface of the globe can be colored with no more than five colors. The problem, if I recall correctly, was first posed by the Englishman Francis Guthrie in 1852.”

(Editor’s note: In 1879 Arthur Kempe, a barrister and a member of the London Mathematical Society, claimed a proof that four colors would always suffice; however, 11 years later Heawood found a subtle error. For almost a century, nobody knew whether every map can actually be four-colored. Finally, in 1976, with substantial computer assistance, Kenneth Appel and Wolfgang Haken of the University of Illinois proved that five colors are never needed.)

Elphinstone thought for a moment. “But surely one color will suffice?”

“Hmm? Oh. I failed to mention that adjacent countries must receive different colors.”

“I see,” the Major said. “What about 100 countries adjacent only at a common point, like slices of a pie?”

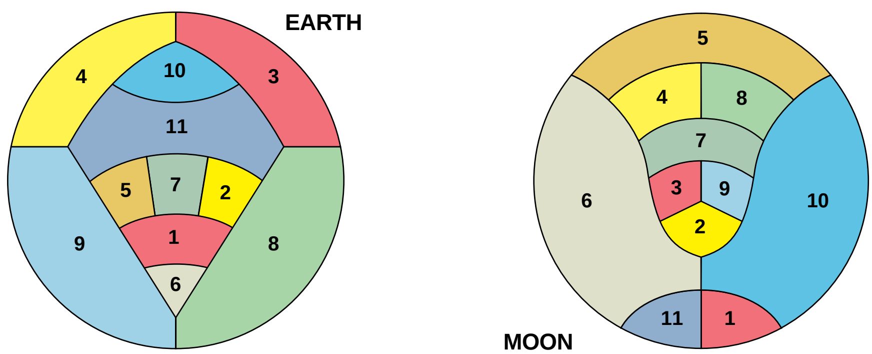
“I should clarify my terms,” Maston answered. “By ‘adjacent’ I mean that they share some common boundary of nonzero length. Now, I notice this map appears to use many more colors than is necessary. But I suppose criteria other than mere adjacency were used.”

Elphinstone finished his coffee and called for a brandy. Suddenly Maston leaped to his feet in excitement. “I have just remembered an extension of the map-coloring problem,” he explained. “Suppose that instead of countries, we consider empires. Countries belonging to the same empire must receive the same color, of course. Different empires may also receive the same color, however—subject to the same rule as before: if two countries are adjacent, then they must receive different colors.”

“Logical.”

“Indeed. The implication, of course, is that if any two empires possess adjacent territories anywhere on the globe, then those empires must receive different colors. Let me say that an empire that contains exactly two countries—each country being a single connected region—is a 2-pire; one that contains three countries is a 3-pire, and so on. Then an empire with m countries would be called an m-pire. I must apologize for the feeble joke.”

“Yes,” the Major muttered. “You must.”



MAPMAKER’S INCUBUS: What is the maximum number of colors needed for a map in which every empire consists of two countries, one on the earth and the other on the moon? Nine colors are sometimes needed, as shown above, but no one knows whether there are some such maps that require 10, 11 or even 12 colors.

“In my defense, I should point out that it is not mine, but Heawood’s.”

“I wonder if he realized that an empire with x countries would be called an x-pire?”

“Good point,” Maston conceded. “Anyway, in 1890 Heawood proved that one needs at most 12 hues to color any given map consisting of 2-pires.

“Heawood discovered a map with 12 mutually adjacent 2-pires [*see illustration on opposite page*], which perforce required distinct colors, thus proving that 12 is the smallest possible number of colors in general. Heawood went further and proved that every map consisting of m-pires requires no more than $6m$ colors.”

“And did he show that this number is also the best possible?”

“No, he did not. But he conjectured such a result. In fact, he suggested that there always exists a map with precisely $6m$ mutually adjacent m-pires.”

(Editor’s note: In 1984 Brad Jackson of San Jose State University and Gerhard Ringel of the University of California at Santa Cruz proved Heawood’s conjecture. A map of 18 mutually adjacent 3-pires, discovered by Herbert Taylor of the University of Southern California, is shown on the opposite page as well as a map with 30 mutually adjacent 5-pires.)

The Major ordered a second brandy. “Would it make any difference if some of the countries in an m-pire were on the moon?” he interjected.

Maston thought for a moment. “Probably,” he said. “After all, we are now considering maps on a pair of spheres instead of just one. I suppose the simplest case would be when each country on the earth is the possessor of a 2-pire, of which the second country is a lunar colony. I believe methods akin to Heawood’s will show that the maximum number of colors required lies somewhere between eight and 12.”

(Editor’s note: Rolf Sulanke of Humboldt University in Berlin showed that some such maps require nine colors, but it is still not known whether the correct answer is 9, 10, 11 or 12. The reader might like to consider 3-pires, each having one territory on the earth, the moon and Mars. For three planets, the optimal number of colors is either 16, 17 or 19; for m *greater than or equal to four, it is either 6*m *– 2, 6*m *– 1 or 6*m*.)*

Elphinstone and Maston made their way back to the observatory.

“Still no word,” Elphinstone reported, checking the ticker tape. Maston turned the great handle that moved the telescope on its mountings. He bent down and placed his eye against the instrument.

“Any luck?” Elphinstone asked.

“No, not yet.” Maston fiddled with the settings. “Ah! There she is!”

“May I look?” The Major could barely make out a tiny dot against the lunar landscape. “So they made it. Soon there will be a new map with a purple moon!”

“And a dash of green,” Maston added.

“Yes, of course,” Elphinstone said, taking another peek through the telescope. “My word! I think I see several other spacecraft.”

The ticker-tape machine suddenly began to chatter. Maston rushed over to read the message: “INTERCONTINENTAL NEWS AGENCY REPORTS THAT MANNED MISSILES WERE LAUNCHED TO THE MOON TODAY BY ARGENTINA, BELGIUM, BRAZIL, BRITAIN, CHINA, GERMANY, HOLLAND, JAPAN, PORTUGAL, SPAIN, RUSSIA AND THE UNITED STATES.”

The Major stared at Maston. “I believe you were saying purple, green and between seven and 10 additional colors?”

FURTHER READING

PEARLS IN GRAPH THEORY: A COMPREHENSIVE INTRODUCTION. Nora Hartsfield and Gerhard Ringel. Academic Press, 1990.

BOOK REVIEWS by Philip Morrison

Molecules at Work

The Machinery of Life, by David S. Goodsell. Springer-Verlag, 1992 ($29).

In this impressive and most original book, the author, both a research biochemist and a painstaking master of computer graphics, offers a guided visual trail into the essential mechanics of living cells. The reader should not expect a swift, effortless ride but rather an enjoyable stroll to pass and ponder scores of intricate drawings of molecular machinery at work.

These are no casual sketches; all the molecules are strictly dimensioned, their structure documented. For those mechanisms shown in context, the drawings offer "a rough idea of the gross shape and size of the molecule" by a smooth, simple outline that suggests three dimensions. A special gallery of a dozen or so isolated but important molecules is presented in color. All drawings show only the outside surface form of the atoms that define the private space within which no other atom can penetrate.

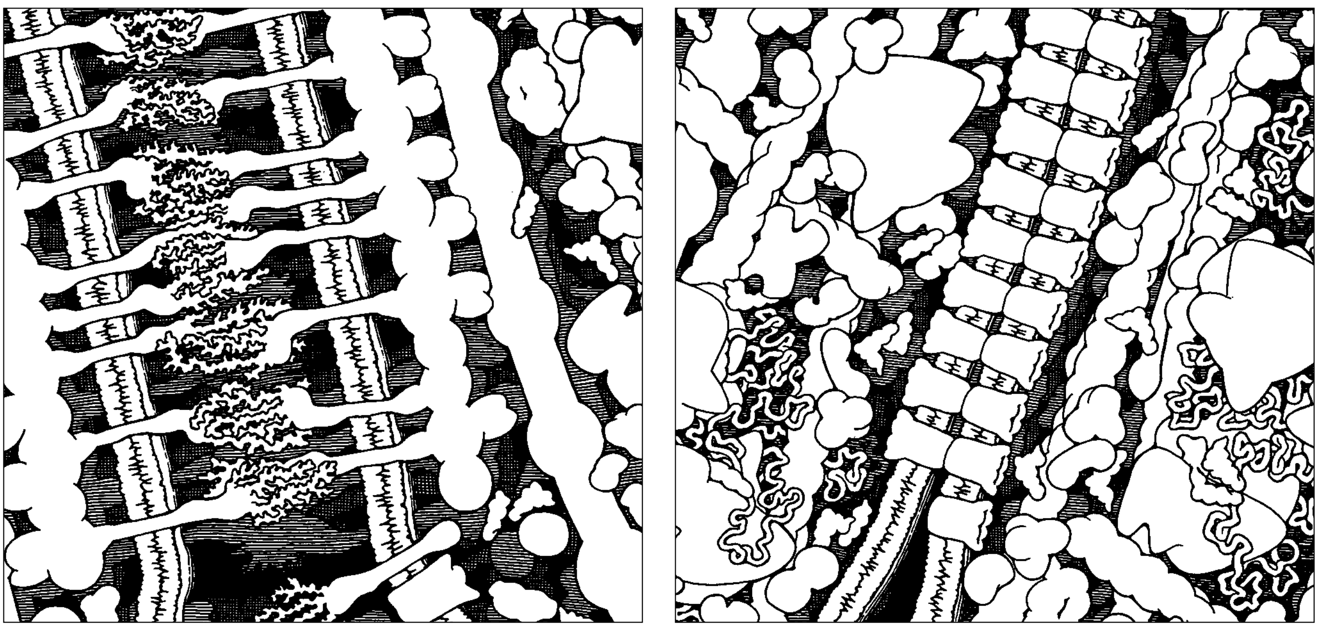
What is here are illustrations of vital machines, their frames, levers and links, springs and detents. The visual outcome is clear and informing, although this world, at once intimate and alien, is seldom beautiful. Many key portions of cells are mapped mostly on a scale of one million to one. That scale rules events; ubiquitous gravity, for instance, is here without real effect.

The drawings of macroscopic machines always label materials. Human engineering rests on metal, wood, plastic and ceramic. Cell materials are fourfold, too: clumped and folded proteins and nucleic acids, thready lipids (fats and oils), and branching polysaccharides. (A sample of each opens the color gallery of molecules.) All of them operate in a bath of salty water, 70 percent of the cell by weight, which envelops the cell contents. Each material interacts with that watery environment; lipid molecules, for instance, have two distinct ends. Their heads chemically seek water; their tails shun it. Cell-surface membranes use double molecular layers of such lipids, stacked tail to tail; only the heads make surface contact with watery matter both inside and out; within the fluid sheet made by the joined, oily tails, specific mobile proteins drift.

Life is not only structure but active process. The machinery has to operate. Thermal motion rules; molecules rapidly turn, twist, jostle, even drift. The molecular machines advance a step or two whenever during their ceaseless dance the right atomic partners chance to touch, then to part, exchange or stick. The fifth constituent of the living cell is not to be pictured: it is chemical energy, subtly dependent on molecular surroundings, shape, internal quantum states and ionic charges.

A set of illustrations presents the 10 programmed steps to "sugar breaking" that degrade commonplace glucose. Many steps enable transfer of fuel energy with little loss in heat to build transportable molecules of ATP, welcome energy donor everywhere within cells. The complex jigs and dies of protein, those big, chunky enzymes that in sequence attach or remove a variety of small molecules, are all shown with active surfaces shaded. But their catalytic functions, able to distort reactant molecules to speed chemical change without fiery temperatures, are too far from external form to become clear to a nonbiochemical reader. The stepwise molecular figures are of interest, though not as illuminating as the text itself. Glycolysis, "the bane of biochemistry students," remains mainly obscure.

Elaborate and fascinating drawings present crowded sections of living cells, all in still life, of course. The intestinal bacterial form called *Escherichia coli* is



NEIGHBORING CELLS in the human body are bound together by spot desmosomes, shown in cross section in the illustration on the left. Proteins extend across the opposed membranes, binding together in the space between. Cells communicate and share resources through tiny pores in gap junctions like those shown on the right (magnification 1,000,000 times).

as numerous as any organism on the earth and biochemically better known than any other. One entire capsule-shaped cell of *E. coli* decorates the jacket; within the book five key regions are shown magnified another power of 10. Their contents are mapped biomolecule by molecule; the densest pattern, the small water molecules of the pervasive matrix, is suppressed, except in one exemplary corner. The composite cell wall is sectioned to disclose five or six functional layers, from the gluey outermost fringe of brushes of protective polysaccharides to the innermost membrane of them all, studded with protein ion pumps for transport and energy release. Deep within the cell lies the DNA, in this minute organism not safeguarded within a walled nucleus but as a single long, rambling loop of message tape, wound for storage around many little reels of a disk-shaped protein.

A yeast cell represents the larger eukaryotic cell, our own kind of cell, with the evolved advantages of internal compartments and an intricate framework. There is too much wonderful mechanism to summarize; the most striking visually is the three-armed "triskelions" that pull out little bubbles of membrane that carry digestive protein molecules away to work on the debris inside the cell vacuole.

Then we view in molecular detail a small sample of specialized human cells and tissues. Blood serum with its antibody proteins and a strange hexapod protein (a member of the "complement cascade," part of the immune system) are dazzling, although simple compared with the neuron synapse, in which neurotransmitters are seen in voyage across the cleft. We see the forms of vitamins, poisons, drugs and viruses, all molecules that are biologically potent even in numbers very small compared with the multitudes of molecules of air, water and food we steadily consume. The slender book ends on a pleased note of fruitfulness that should continue as long as our molecular insight grows.

A Mayan in St. Petersburg

BREAKING THE MAYA CODE, by Michael D. Coe. Thames and Hudson, 1992 ($24.95).

Michael Coe is a Yale anthropologist of the pre-Columbian, a remarkably candid insider, who chronicles the origins and accomplishments of recent Maya scholarship. Here he relates "one of the greatest intellectual achievements of our century." This true story is as bizarre, contentious, readable and tragic as the grand novels of contemporary Neotropical fiction.

The volume opens by setting the stage with the rise of epigraphy and a vivid account of the growth of Maya archaeology. About halfway into the book we come to the crux—how in the 1950s and against all odds a young Russian began to decipher the Maya glyphs in far-off St. Petersburg. Full entry of the Maya into history is a matter of the past decade or so, or of tomorrow, for the texts are articulate now but not yet fluent.

It seems worthwhile to open with the one bit of strictness that undergirds all recognition of lost scripts. Count the number of symbols. From Etruscan to Sanskrit, all alphabetic writing systems display sign lists that include 20 to 40 distinct symbols; the syllabic systems from Persian to Cherokee or Japanese count some 40 to 85 signs; beyond that, "logographic" systems enter, with symbols for words—not at all picture writing—where the count runs from 500 for Hittite up to above 5,000 for Chinese.

The Maya count is some 800 glyphs in what is no great body of text. Such counts are not trivial to assess. Even an alphabet is multiple: inscriptions cross time, signs do change, uniqueness is not to be found in the complex carvings. The Maya signs in use at any one epoch number only 200 to 300. The redoubtable code breakers of wartime have done and can do little for ancient scripts. Their methods are made for long, precise text strings, often five-letter groups in a known tongue, ingeniously concealed. The glyphs are not like that at all. (It should be said that for 100 years we have read with growing precision the calendrical and astronomical texts of the Maya. Those have a helpful deep structure that transcends any language.)

Old scripts always record a spoken tongue. In all logographic scripts, Sumerian, Egyptian or Chinese, principles are mixed. Phonetic elements and a variety of auxiliary forms are there to relate signs to the spoken words, via some homely rebus or syllabic marking. The great Champollion, who cracked the essentially phonetic nature of Egyptian hieroglyphics from the Rosetta Stone, went to Egypt only late in his short life. But he was a linguist of parts; he already knew Coptic, the tongue descended from ancient Egyptian, when he was 17.

Yuri Valentinovich Knorosov, now 70, is the Maya approximation to Jean-François Champollion. His tale begins with a near impossibility, amid the flames of fallen Berlin. An artillery spotter, he fought his way into that city in May 1945 and chanced to watch the National Library burn. He could save just one book. It was a one-volume 1933 edition of the three known Maya codices. He brought it back along with four battle medals. The young veteran went to Moscow University to study Egyptology, but he found his way to an interest in other old tongues. His professor challenged his brilliant student with Mayan; on graduation, Yuri went to the Institute of Ethnology in St. Petersburg, where he is still, and took up his work.

By 1952 he had found his way. His Rosetta Stone was really only a much-studied and long-rejected fragment, some pages and an inconsistent Maya "alphabet" written out in the 1550s by the Franciscan Diego de Landa, with the aid of learned, even princely, informants in Yucatán itself. (We know it only through a chain of transcriptions and editings over time.)

In the same work Bishop de Landa first listed the Maya day signs in an introduction to their calendrical system, long known as fruitful. Now the Russian's new look at the material gave clear evidence that Mayan, too, was as pervasively syllabic and phonetic as Egyptian. Yucatec and Cholan would be as relevant as Coptic had been in Egypt, and the scholarly dictionaries of Maya tongues offered him workable readings that explained de Landa's few dozen badly misunderstood syllables. The glyphs do echo the spoken word. Of the 800 glyphs, about 150 are now known to be syllabic or phonetic. There are plenty of gaps still.

In Britain a potent skeptic dwelt. He was John Eric Sidney Thompson, a man of wide if selective erudition and utter self-assurance, whose magisterial writing dominated modern Maya studies from the 1920s until his death in 1975. His voluminous work held real value; even his list of the glyphs of the Maya is still standard. But he brooked no deep critic, although he lived manifestly in an absurd dream of high-mindedness. "Beyond doubt," the glyphs have spiritual and mystical meaning, he wrote, on "the threshold of the…Maya soul." He mocked poor Knorosov, a Marxist whose papers often appeared flanked by the Soviet journal's obligatory cold war propaganda, a linguist at that, everything that Thompson derided. Nor did Knorosov get it all right; often he went astray. A few pioneers dared Thompson's wrath, but on the whole the phonetic line of work went underground.

In the end the new path drew new people. A young artist, Linda Schele, found her way by chance to Palenque and its inscriptions in 1970. Now she is an incisive scholar and a charismatic leader; around her annual workshops (often in Austin, where she is professor) the decipherments flow. The contributors are largely—not all—young; undergraduates, graduates, graduate students, even many amateurs outside the academy bring in new results and queries. You can read one stela in full here, in Schele's rendering keyed carefully to a drawing of the glyphs. It is a little dull, all about the assumption of the throne of Piedras Negras by their Lady Queen, but state monuments read like that in all climates.

Tragedy is pervasive in any account of the Maya. Most of the spade archaeologists—the populists, few of whom speak any Maya tongue—keep their distance from these translations; what can these elite texts tell us of the people? Sour grapes, Coe says. When Knorosov came to Guatemala in 1990 to accept the president's gold medal and see the glyphs for the first time, the death squads warned him against visiting and so honoring the old sites. Texts are a heritage; hope for a new Maya polity is still alive among millions in the villages, but anathema to right-wing murderers in anxious capitals. Gabriel García Márquez will understand.

Spooky Action Up Close

TAMING THE ATOM: THE EMERGENCE OF THE VISIBLE MICROWORLD, by Hans Christian von Baeyer. Random House, 1992 ($23).

Cocksure Samuel Johnson once refuted the idealism of the philosophers simply by kicking a large stone "with mighty force." Atomic reality, long a target of critical philosophers, is demonstrated nowadays with like concreteness. At IBM, some physicists can deftly tease half a dozen atoms into chosen sites on a crystalline surface. They nudge the atoms with the minute tip of the scanning tunneling microscope as they watch a scanned color-coded image built up out of the varying tip current. All deflections of the probe as it feels along the atoms are fed back to the robotic manipulator, so that the operator's fingers on the controls sense proportionate resisting forces, as if the atoms were hen's eggs being nested within their carton. Visual and tactile magnification must alike be a few millionfold; then can we see our atoms and touch them, too.

Hans von Baeyer, a physicist at the College of William and Mary, writes as vividly of sweeping ideas as of yesterday's conversation. His text is as personal as it is charmingly allusive, aptly citing philosophers old and new, yet never distancing the reader. He sums up recent findings in several labs, some of which he has visited, to report how fully those once conjectural atoms have been caught in our perceptual reach.

In Seattle he entered Nobelist Hans Dehmelt's wonderfully realized dream of "a single atomic particle floating forever at rest in free space." There von Baeyer met little Astrid, an atom so familiar as to have earned a name of her own. Astrid was carefully tended and fed by cunning laser beams while ever so gently enclosed in a spacious and subtle trap. Like the fox in Antoine de Saint-Exupéry's *Little Prince,* she might have said, "To you I am nothing more than an atom like the other atoms. But if you tame me...I shall be unique in all the world." An isolated atom afloat in space, all but stationary in the dark field of an ordinary microscope, she shone as a dot in pale blue, a hue more distinctively her own clan's than any tartan can be, no flash or track or video scan but a lasting source of real light in real time, ready and waiting for any eye to view.

In Rochester, Carlos R. Stroud, Jr., and his team probed instead deep within their atoms, passing from enjoyment of a lovely tamed creature to the intensity of repeated surgical intervention. Using a sequence of precisely planned disturbances—first with delicate microwaves, then by fierce laser pulse, then after waiting for an atomically long time, sudden dissection by an electric field—they could map the electron within the atom.

All these predicted marvels are not enough. Von Baeyer reports the present doubts of able physicists that the inherent duality of quantum theory between point and cloud can be the last word. A little garden of dazzlingly ingenious experiments lately have enlivened the old question. For light or sound waves, the phenomenon is familiar: well-prepared waves that can arrive via alternative paths interfere, both to reinforce and to cancel. One open slit casts a blurry image unlike the neat fringes made by two.

Now go to the level of distinct atomic events. Map the arrival pattern of many identical atoms sent one at a time toward a pair of open slits. Interference of such whole atoms was reported by four experimental groups during 1991, no small feat, since the two precise slits apt for the optical lab at submillimeter scale must here still be neat, yet 1,000 times finer. If each tiny atom passes through either one slit or the other, how can the slit through which an atom did *not* go yet affect its final destination?

The clever doubters found more trouble. (A fine summary of recent experiments, "Quantum Philosophy," by John Horgan, was in the July 1992 issue of this magazine.) These experiments all imply, as does the theory, that any sure information on which slit any particular particle arrived ends its participation in interference. This remains true even if the choice of slit is found by a marker process that does not directly disturb the atom's flight and even (an experiment still in preparation) if knowledge of the slit choice is left open to an optional erasure well after the particles have passed. Erase, and you see interference; retain, and the two slit images merely add. Notice that in this case all the atomic arrivals together will record a blurred pattern. The observation of interference is itself only a retroactive one, found long after all particles have reported by deleting the arrival record of just those atoms whose path was reliably noted.

The explanation given by quantum mechanics is not new, but it is firm. Atoms, like photons, are described by a quantum field, something different from its two limiting cases, classical wave and particle. Physically alternative paths divide a coherent wave front between them, yet any pellet must pass through either one slit or the other. The quantum theory describes all the results we see by clear, unambiguous rules that show just how the case "both A and B" differs intrinsically from "either A or B." Photons and atoms alike obey; what is hidden from us is the inherent differences in flight times.

Einstein first laid down a strong challenge to quantum mechanics along these lines; he would not accept "all that spooky action at a distance." "No doubt the day will come," he said, "when we will see whose instinctive attitude was the correct one." At present, most physicists hold, firmly though not dogmatically, that the day of decision has passed. The atomic world seems to be the curiously dicey place Bohr foresaw, much more like everyday life than like the fully analyzable trajectories of classical physics. Our author rather inclines to favor the doubters; he hopes that the atoms now so open to detailed study will lead us to a new understanding.

Perhaps more novelty lies in what we call empty space. The quantum vacuum is no simple void but a plenum, crammed by fluctuating fields and transient particles. Yet a macroscopic volume of "empty" space, simply confined by metal walls, exerts the very pressure predicted. A similar confinement, as von Baeyer reports from Daniel Kleppner at M.I.T., measurably slowed the "spontaneous" decay of an enclosed atom. Is it around the wilderness in the vacuum that little crystals of new quantum physics will first form?

ESSAY by Anne Eisenberg

Scientists in the Movies

President Bill Clinton said recently that he would admit to being a wonk—but not a nerd. No wonder. He probably learned the distinction at the movies, watching that apotheosis of nerds Jerry Lewis, playing a chemist in *The Nutty Professor,* careening into class in his white lab coat and tripping on the way to the lecture platform. Or maybe he picked up the difference watching *Revenge of the Nerds*—the nerds are computer-obsessed undergraduates who pin pictures of Einstein on the wall and don't get into fraternities.

"Nerd," it turns out, is movie shorthand for scientists, engineers and assorted technical types who play chess, perhaps, or the violin (but never the saxophone). At one time, "nerd" meant "socially inept person," but nowadays the meaning has narrowed, no doubt because of movies like *Revenge of the Nerds,* to "socially inept technical person," making it the first generic derogatory noun for scientists and engineers.

Nerd. What a linguistic tribute. And it joins a string of adjectives, each more dismal than the next, that are closely associated in the popular imagination with the noun "scientist," among them "power-driven," "unbalanced," "insensitive" and, most widespread of all, "mad," as in "mad scientist."

Blame the spread of these uncomplimentary terms on the movies, the place where words and images combine to produce today's notion of the typical scientist: a buffoon, an obsessive or a maniac.

This was not always so. True, seekers of knowledge have often had to contend with stiff penalties for daring to tamper in God's domain. But until the late 1950s, there usually was a noble Dr. Ehrlich for every deranged Dr. Frankenstein. In fact, in the 1930s, 1940s and even mid-1950s, Hollywood released a series of reverent film biographies, including *The Story of Louis Pasteur, Dr. Ehrlich's Magic Bullet* and *Madame Curie,* which earned scientists adjectives like "visionary," "inspired" or, at the least, "helpful." But now this genre of film—as well as parallel movies about inventors such as Edison and Bell, which dramatized the thrills and benefits of technology—is a thing of the past.

Today most depictions of scientists are fictional portrayals in which the scientists do harm either to themselves or to others. The drama of discovery has been replaced by the drama of destruction. In dozens of Faustian horror films, such as *The Fly* or its remake, the damage is confined to the scientist; in countless other films, scientists bring trouble on us all, deliberately (in their quest for power) or unwittingly (when their inventions escape their control). Sarah, the heroine of the hugely popular film *Terminator 2: Judgment Day,* expresses the common view when she denounces all scientists, in the person of one Dr. Dyson, whose invention has led to global war. "You [scientists] think you're so creative," she says. "All you know how to create is death and destruction."

In his book *Taming the Tiger: The Struggle to Control Technology,* Witold Rybczynski dates the change in the image of scientists to the second half of the 20th century, when, he says, people began to worry deeply about the influence of technology. Certainly, in the movies, the idea that scientists are productive people who by and large advance the human condition has evaporated, replaced in camp versions by Vincent Price in a fright wig up on his mountain playing the organ or in more serious offerings by Dr. Strangelove rolling his wheelchair across the War Room to calculate the odds of surviving a doomsday machine.

Even in the lightest Hollywood fare, the movie scientist brings threats, not benefits. Granted, there is the rare sympathetic throwback, usually in sports comedies in which the scientist unveils an invention that causes baseballs to repel bats (*It Happens Every Spring*) or basketball players to rise in the air (*The Absent-Minded Professor*). But in the main, the audience is treated to a parade of scientists who shrink their kids, hinder sensitive robots in their search for self-fulfillment (*Short Circuit*) or chase adorable mermaids (*Splash*) and extraterrestrials (*E.T.*) in hopes of dissecting them. Scientists cringe as they watch their screen counterparts; indeed, I know some who have become so defensive that they routinely misread the AIDS-activist slogan "Silence = Death" as "Science = Death."

Only in the Saturday serials and their successors has the image of the good scientist remained unscathed; there scientists are still the galaxy's best problem solvers. Where would Captain Kirk be without his unflappable, analytical science officer Mr. Spock? Flash Gordon without Dr. Zarkov? The vampire hunters without Professor Abraham Van Helsing? Of course, the scientists do not get top billing—usually they are the sidekicks, not the stars—and most testaments to their scientific abilities are indirect: we are unlikely to hear adjectives such as "good" or "useful" applied to them, although occasionally they merit "logical." Still, they have avoided "mad" and "power-driven," and clearly they are handy people to have on the team when the time comes to drive off aliens or sterilize vampire soil.

It's time for a change in image. There's one small sign that movie audiences may be ready for the shift. Recently Errol Morris made a documentary based loosely on Stephen W. Hawking's best-seller *A Brief History of Time.* The film is that rarity, a nonfiction account of a scientist's life. In it, Morris interviews Hawking's friends, colleagues, students and family relations who think well of him; they talk sympathetically of Hawking's boyhood, his intellect, his development and, most of all, his extraordinary resilience. Nowhere is there the slightest suggestion of the scientist as the primary source of disorder; instead we see a flattering, inspirational story of someone who succeeds against the stiffest of odds.

Hawking's colleagues, too, come off quite well, particularly Princeton physicist John A. Wheeler, explaining, among other things, the origin of the term "black hole." "After you get around to saying 'gravitationally completely collapsed star' about 10 times," he says, "you look desperately for something better." Wheeler's lucid explanatory style delights viewers, who can for a change not only admire a scientist but actually understand what one of them is saying—even in the movies. Who knows, the rehabilitation of movie scientists may continue—is James Gleick ready with his screen version of *Genius: The Life and Science of Richard Feynman*?

ANNE EISENBERG, a professor at Polytechnic University in Brooklyn, is the author of four books on scientific writing.