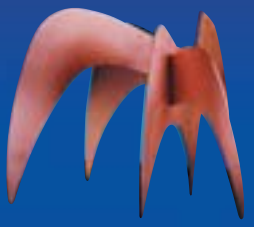
BRAINS HARMED BY HEART SURGERY • KEYS TO A LOST CITY

SCIENTIFIC AMERICAN

JULY 2003 $4.95 WWW.SCIAM.COM

NATIONAL MAGAZINE AWARD



WINNER FOR 2003

(see page 6)

SMART ANTENNAS
THEY KNOW WHERE YOU ARE

New Thinking on Cancer's Roots

How Black Holes Can Spawn Stars

Running Out of Fish

COPYRIGHT 2003 SCIENTIFIC AMERICAN, INC.

july 2003

contents

SCIENTIFIC AMERICAN Volume 289 Number 1

features

66 An enigmatic priest-king

RANDY OLSON *Aurora Photos*; DEPARTMENT OF ARCHAEOLOGY AND MUSEUMS, GOVERNMENT OF PAKISTAN

COPYRIGHT 2003 SCIENTIFIC AMERICAN, INC.

departments

columns

In May the American Society of Magazine Editors presented *Scientific American* with a National Magazine Award for editorial excellence in the Single-Topic Issue category for the September 2002 issue, "A Matter of Time." Our thanks go out to ASME and to the many contributors who made that issue a success. —*The Editors*

Cover image by Kenn Brown;
National Magazine Award "Elephant," by Alexander Calder

Scientific American (ISSN 0036-8733), published monthly by Scientific American, Inc., 415 Madison Avenue, New York, N.Y. 10017-1111. Copyright © 2003 by Scientific American, Inc. All rights reserved. No part of this issue may be reproduced by any mechanical, photographic or electronic process, or in the form of a phonographic recording, nor may it be stored in a retrieval system, transmitted or otherwise copied for public or private use without written permission of the publisher. Periodicals postage paid at New York, N.Y., and at additional mailing offices. Canada Post International Publications Mail (Canadian Distribution) Sales Agreement No. 242764. Canadian BN No. 127387652RT; QST No. Q1015332537. Subscription rates: one year $34.97, Canada $49 USD, International $55 USD. Postmaster: Send address changes to Scientific American, Box 3187, Harlan, Iowa 51537. **Reprints available:** write Reprint Department, Scientific American, Inc., 415 Madison Avenue, New York, N.Y. 10017-1111; (212) 451-8877; fax: (212) 355-0408 or send e-mail to sacust@sciam.com **Subscription inquiries: U.S. and Canada (800) 333-1199; other (515) 247-7631.** Printed in U.S.A.

COPYRIGHT 2003 SCIENTIFIC AMERICAN, INC.

Three Lessons of SARS

Surveying the worldwide panic over severe acute respiratory syndrome, contrarians have hinted that it smacks of media-fed hysteria. Compared with malaria, which annually kills a million people, isn't SARS—its death toll at about 600 as of this writing—fairly trivial? No. SARS seems to be roughly as contagious as influenza and several times as lethal as the 1918 Spanish flu that killed upward of 20 million. Known antiviral drugs do not work against it. Moreover, if even a fairly benign form of the virus becomes endemic, new strains could always mutate again to virulence. Controlling SARS would then be a chronic global burden. In view of the unknowns, the World Health Organization and local authorities have been right to err on the side of caution.

DOCTOR IN BEIJING contemplates a SARS patient.

REUTERS/CHINA PHOTO

SARS has already taught us at least three hard lessons:

New viruses can be hard to contain, but reining in damaging misinformation is harder. The disease has wrought tens of billions of dollars of damage through economic slowdowns, canceled trade and lost tourism. Some losses were inevitable consequences of the essential quarantines and travel advisories, but others were not. An Internet rumor that the government would seal Hong Kong's borders triggered a run on food and other supplies. Riots have broken out in China. Even in the U.S., where SARS cases have been few and well isolated, many people shunned Asian markets and restaurants. The WHO's short-lived advisory against travel to Toronto will be debated for years.

David Baltimore, president of the California Institute of Technology, has suggested that the media could have done more to convey that for most individuals, quarantines and other safeguards make the risk of SARS exposure virtually nonexistent. He may be right. Still, frightened people also read between the lines of whatever information they have, and official disavowals of danger are not always credible (consider the case of the British government on mad cow disease). No foolproof public information formula for preventing disease panics may exist.

Molecular understanding of a virus can be frustratingly impotent. Researchers deciphered the genetic code of the SARS coronavirus within days. Yet turning that knowledge into weapons against the disease is a much slower, harder task. Developing a SARS vaccine might take at least a year. For now, control of SARS depends largely on the blunt, Dark Ages instrument of quarantine. Biomedical science cannot create cures as fast as it gathers data. SARS is only the latest humbling reminder of that reality, and it won't be the last.

Global public health is everybody's business. Even now, few Americans probably give much thought to the health of poor Chinese farmers. Yet millions who live closely with the swine and fowl they tend represent countless opportunities for viruses to leap species and ignite new epidemics. That situation is not unique to China or even to the developing world. And it is not one we can ignore, because international trade and travel can deliver diseases anywhere, anytime.

Nothing can stop new diseases from evolving, but strong public health and hygiene systems can slow the process. They can also recognize emerging diseases and try to control them—if they have the opportunity. In the early days of the SARS outbreak, WHO officials expressed frustration that Chinese officials rebuffed their requests to investigate for themselves. Such urgent inquiries need more teeth.

THE EDITORS editors@sciam.com

COPYRIGHT 2003 SCIENTIFIC AMERICAN, INC.

How to Contact Us

EDITORIAL

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

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

SUBSCRIPTIONS

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

REPRINTS

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

PERMISSIONS

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

ADVERTISING

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

New York
Scientific American
415 Madison Ave.
New York, NY 10017-1111
212-451-8893
fax: 212-754-1138
Los Angeles
310-234-2699
fax: 310-234-2670
San Francisco
415-403-9030
fax: 415-403-9033
Midwest
Derr Media Group
847-615-1921
fax: 847-735-1457
Southeast/Southwest
MancheeMedia
972-662-2503
fax: 972-662-2577
Detroit
Karen Teegarden & Associates
248-642-1773
fax: 248-642-6138
Canada
Derr Media Group
847-615-1921
fax: 847-735-1457
U.K.
The Powers Turner Group
+44-207-592-8331
fax: +44-207-630-9922
France and Switzerland
PEM-PEMA
+33-1-46-37-2117
fax: +33-1-47-38-6329
Germany
Publicitas Germany GmbH
+49-211-862-092-0
fax: +49-211-862-092-21
Sweden
Publicitas Nordic AB
+46-8-442-7050
fax: +46-8-442-7059
Belgium
Publicitas Media S.A.
+32-(0)2-639-8420
fax: +32-(0)2-639-8430
Middle East and India
Peter Smith Media & Marketing
+44-140-484-1321
fax: +44-140-484-1320
Japan
Pacific Business, Inc.
+813-3661-6138
fax: +813-3661-6139
Korea
Biscom, Inc.
+822-739-7840
fax: +822-732-3662
Hong Kong
Hutton Media Limited
+852-2528-9135
fax: +852-2528-9281

On the Web

WWW.SCIENTIFICAMERICAN.COM

FEATURED THIS MONTH

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

Sensing Trouble

World events of the past two years have brought with them a number of new worries for the average American. The safety of the water supply, the risk of hijacking, and the threat of chemical and biological weapons being used on our shores have moved to the front of the country's collective consciousness. At the annual meeting of the American Chemical Society, presentations focusing on domestic security concerns were a noticeable addition to the program, with scientists outlining new ways to detect dangerous chemicals and describing novel applications of time-proven techniques.

Astronomers Spy Surface Ice through Titan's Haze

Imagine Los Angeles on an especially smoggy summer day: the sun's otherwise intense rays are muted, bounced back and forth off the particles in the air as if in a giant game of pinball. Light that does make its way through the dense atmosphere is unlikely to make it out again. And so it is on Saturn's moon Titan, where haze forms an atmosphere 10 times as thick as the one on Earth. This nearly opaque curtain has prevented planetary scientists from learning much about what lies beneath. Now new observations from infrared telescopes are providing the clearest picture yet of Titan's surface. The findings indicate that this moon is covered, at least in part, by frozen water.

Ask the Experts

How were the speed of light and the speed of sound determined?
Chris Oates of the National Institute of Standards and Technology enlightens.

Scientific American DIGITAL

More than just a digital magazine!
Subscribe now and get:
All current issues before they reach the newsstands.
More than 140 issues of *Scientific American* from 1993 to the present.
Exclusive online issues for FREE (a savings of $30).

Subscribe to Scientific American DIGITAL Today and Save!

COURTESY OF JET PROPULSION LABORATORY

COPYRIGHT 2003 SCIENTIFIC AMERICAN, INC.

SCIENTIFIC AMERICAN®
Established 1845

EDITOR IN CHIEF: John Rennie
EXECUTIVE EDITOR: Mariette DiChristina
MANAGING EDITOR: Ricki L. Rusting
NEWS EDITOR: Philip M. Yam
SPECIAL PROJECTS EDITOR: Gary Stix
SENIOR EDITOR: Michelle Press
SENIOR WRITER: W. Wayt Gibbs
EDITORS: Mark Alpert, Steven Ashley, Graham P. Collins, Carol Ezzell, Steve Mirsky, George Musser
CONTRIBUTING EDITORS: Mark Fischetti, Marguerite Holloway, Michael Shermer, Sarah Simpson, Paul Wallich

EDITORIAL DIRECTOR, ONLINE: Kate Wong
ASSOCIATE EDITOR, ONLINE: Sarah Graham
WEB DESIGN MANAGER: Ryan Reid

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

COPY DIRECTOR: Maria-Christina Keller
COPY CHIEF: Molly K. Frances
COPY AND RESEARCH: Daniel C. Schlenoff, Rina Bander, Emily Harrison

EDITORIAL ADMINISTRATOR: Jacob Lasky
SENIOR SECRETARY: Maya Harty

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

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

VICE PRESIDENT AND PUBLISHER: Bruce Brandfon
ASSOCIATE PUBLISHER: Gail Delott
SALES DEVELOPMENT MANAGER: David Tirpack
SALES REPRESENTATIVES: Stephen Dudley, Hunter Millington, Stan Schmidt, Debra Silver

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

GENERAL MANAGER: Michael Florek
BUSINESS MANAGER: Marie Maher
MANAGER, ADVERTISING ACCOUNTING AND COORDINATION: Constance Holmes

DIRECTOR, SPECIAL PROJECTS: Barth David Schwartz

MANAGING DIRECTOR, SCIENTIFICAMERICAN.COM: Mina C. Lux

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

CHAIRMAN EMERITUS: John J. Hanley
CHAIRMAN: Rolf Grisebach
PRESIDENT AND CHIEF EXECUTIVE OFFICER: Gretchen G. Teichgraeber
VICE PRESIDENT: Frances Newburg

Letters

EDITORS@SCIAM.COM

THE MARCH ISSUE generated varying amounts of heat. The relative wisdom of mining data from credit cards and other purchasing patterns to sniff out terrorists, in "Total Information Overload" [Perspectives], sparked some ire. But perhaps the hottest topic—literally—was "Dismantling Nuclear Reactors," and the related, contentiously debated idea of whether to store high-level radioactive waste at Yucca Mountain ["Man against a Mountain," Profile, by Steve Nadis]. The opposing sides—for and against Yucca as a permanent facility—are both wrong, argues Gregory L. Schaffer of Cupertino, Calif.: "All we really need to do is guarantee that Yucca Mountain is stable for, say, 500 or 1,000 years. If problems occur in a century or two, the technology of that era should easily solve them." Searing commentary on these and other articles in the March issue appears on the following pages.

DINOSAURS OF A FEATHER

In "Which Came First, the Feather or the Bird?" Richard O. Prum and Alan H. Brush say that fossils in China dated from 124 million to 128 million years ago help to explain how feathers developed. Earlier in the article, however, the ancient bird *Archaeopteryx* is dated from 148 million years ago.

John Stephens
via e-mail

Your article makes me wonder if porcupines are frustrated birds, given that feathers start out as tubes.

Robert W. Bishop
via e-mail

PRUM REPLIES: Regarding Stephens's question: the history of life is the shape of a tree, not a simple line. These feathered Chinese dinosaurs (dating from 110 million to 128 million years ago) are younger than the earliest bird, Archaeopteryx *(about 150 million years old). But we know from comparative analyses of their anatomy that these feathered dinosaurs lie outside of* Archaeopteryx *and other birds on the tree of life. These nonavian feathered dinosaurs represent younger samples of an earlier lineage in which feathers evolved prior to the origin of birds. Because of the shared anatomical details and a pattern of common ancestry, we can conclude that these feathers are homologous with bird feathers and evolved once in a shared common ancestor.*

*In reply to Bishop's suggestion, hollow hairs occur in a variety of mammals, including North American porcupine (*Erethizon*), North African crested porcupine (*Hystrix*) and caribou (*Rangifer*). Hairs are columnar structures of epidermal tissue with a superficial cuticle layer, a cortical layer and a central medullary layer. Hollow hairs have a simple or degenerate epidermal medullary layer at the center of the hair. This hollow space is not occupied by dermal tissue, as in a feather. These two tubular epidermal appendages evolved separately but grew to resemble each other over time.*

NOT MILK?

Clifford J. Rosen's research into the mechanism of osteoporosis ["Restoring Aging Bones"] is fascinating. But I absolutely could not believe my eyes when I turned and saw a picture of a glass of milk! Research 20 years ago debunked milk as a good source of calcium, because proteins in the milk can cause bones to lose the mineral. "Rebuilding the Food Pyramid," by Walter C. Willett and Meir J. Stampfer, in the SCIENTIFIC AMERICAN January issue, also expressed concern about dairy in our diet, noting that countries with the highest rates of consumption suffer the most fractures.

Rosen's work is invaluable, but without first correcting a poor diet in a patient, this could be a case of using a high-tech solution to fix a low-tech problem.

Andrew Benton
Flemington, N.J.

COPYRIGHT 2003 SCIENTIFIC AMERICAN, INC.

Letters

SHOCKING OVERLOAD

I found "Total Information Overload" [Perspectives] doubly shocking. If the Pentagon and the Transportation Security Administration do what you suggest, analyzing individuals' transactions for clues about terrorist activities, that would threaten our privacy and waste public resources for minimal prospects of enhanced security.

On the other hand, I fear that your unscientific depiction demeans data mining unfairly, lessening prospects for real security gains. Data mining need not be mindless pattern matching. Instead of your caricature, suppose a Computer Assisted Passenger Prescreening System exploited information such as terrorist watch lists, passport activity and criminal records? This could augment passenger safety and lead to justified arrests—benefits lacking in security systems focusing only on the physical weapons detection that you advocate.

Alan Porter
Professor Emeritus, Industrial and Systems Engineering and Public Policy
Georgia Institute of Technology

NUCLEAR WISDOM

As a geologist and member of Maine's Advisory Commission on Radioactive Waste and Decommissioning, I applaud Matthew L. Wald's balanced reporting in "Dismantling Nuclear Reactors." I'd like to mention an issue that has not received adequate scientific and social scrutiny: that of interim spent-fuel storage. His article touches on the Independent Spent Fuel Storage Installation at Maine Yankee and ones like it at other operating and closed power plants. Interim spent-fuel storage has become a necessity as our nation continues to struggle with the politics of long-term housing of these materials. But a comparison of the most simplistic criteria for siting waste facilities and power plants exposes the folly of our current approach. Waste facilities should be isolated from the hydrosphere and placed far from population centers as an extra measure of safety. But power plants are sited near water bodies for cooling and are generally near population centers to reduce transmission losses.

Furthermore, discussion is needed of the security aspects of our present system, which will spawn perhaps 100 or more storage facilities with varying degrees of protection, versus a central interim facility isolated from water and humans. Stacking all the accumulated waste in one secure desert location might be better than our current unplanned system.

Robert G. Marvinney
Maine Geological Survey

HIGH-LEVEL **nuclear waste produced by power-generation plants finds a temporary home in on-site storage systems.**

Decommissioning of nuclear plants starts from the wrong premise—namely, that the safest option is to restore the site to its original condition. There is no justification for such an arbitrary requirement. It would be far simpler to remove the nuclear fuel (for use elsewhere), lock the doors and the gates, paint the outer walls green and wait 1,000 years for everything to cool off. Any other action is fraught with quite unnecessary danger.

The obligation to restore a nuclear power site does not apply to any other structures, such as conventional energy stations, grain silos or cathedrals. The cost of its application to nuclear power plants is enough to price green nuclear energy out of the market, possibly another dastardly ploy by the oil giants.

Graham Hills
via e-mail

GUMMY PRINTS

In Working Knowledge, Mark Fischetti writes, "Fingerprint readers offer greater security, because it is almost impossible to fake a human digit." Unfortunately, you don't have to completely fake a human digit to fool readers. Simply breathing on the device can cause it to reactivate and recognize the latent fingerprint of the previous user (search the Web for "capacitive latent fingerprint"). Also, Tsutomu Matsumoto of Yokohama National University in Japan demonstrated in January 2002 that most fingerprint readers can be fooled by a "gummy" finger, easily created with gelatin and a fingerprint lifted off of a smooth object. More information is available online at www.cryptome.org/gummy.htm

Jeff Martin
Seattle

POLITICAL SCIENTISTS

Bringing more scientists into the decision-making processes at the U.S. Department of State is a very important goal—and was a major recommendation of the National Research Council study on science that I chaired. "From Lab to Embassy," by Sally Lehrman [News Scan], reaffirms this goal. A sidebar states that fellowships have been increased recently, in response to the NRC report, but incorrectly implies that the American Association for the Advancement of Science fellowship program is new. AAAS has had such a program for 23 years.

Robert A. Frosch
Belfer Center for Science and International Affairs
John F. Kennedy School of Government
Harvard University

ERRATUM In "Connect the Pings," by Wendy M. Grossman [News Scan], one of the companies should have been referred to as BAE Systems, rather than BAe Systems.

DAVID MURRAY, JR.

COPYRIGHT 2003 SCIENTIFIC AMERICAN, INC.

50, 100 & 150 Years Ago

FROM SCIENTIFIC AMERICAN

Alien Reality ▪ Mechanical Food ▪ Riot Bones

JULY 1953

GENESIS BY LIGHTNING—"University of Chicago chemist Harold Urey has championed one theory as to how life began on earth. It suggests that a billion years ago or so the earth's atmosphere consisted of methane, ammonia, hydrogen and water vapor. Under the action of lightning discharges or of ultraviolet radiation, these compounds were split into free radicals, which recombined in chance ways to form more complex molecules. A few months ago Urey had one of his students, Stanley L. Miller, assemble a mixture of methane, ammonia and hydrogen over boiling water in an air-tight glass system and circulated the vapor continuously past an electric spark. By the end of the day the mixture turned pink; after a week it was a deep, muddy red, and it contained amino acids—the building blocks of proteins."

PAGING AGENT MULDER—"Is Man alone in space? As for the possible duplication of man on other planets, no animal is likely to be forced by the process of evolution to imitate, even superficially, a creature upon which it has never set eyes and with which it is in no form of competition. Nor could an animal, however gifted in mimicry, ape a man if it came among men. The individual sitting next to you in the theater could not conceivably be an insect masquerading as a man. Even if the body duplication (down to clothes) was perfect, the creature's instinct-controlled brain, its cold, clock-like reaction, in contrast to our warm mammalian metabolism, would make the masquerade hopeless."

JULY 1903

THE CONVENIENCE CENTURY—"To the American, who is now so accustomed to mechanical contrivances that he no longer is astonished by them, the automatic restaurant is but the logical development of the vending machine. This establishment, in New York City, is fitted up elaborately. Its electric lights, its dazzling mirrors, and its resplendent marble outshine everything on Broadway [*see illustration*]. On the upper floor the patrons purchase what they desire; in the basement the food is cooked, and lifted to the floor above by means of small elevators."

AN AUTOMATIC RESTAURANT, New York City, 1903

ART OF THE LETTER—"The letter of a century ago has still a certain literary value. Nowadays we only 'correspond' or we 'beg to state.' It still remains for our children to discard the forms of polite address which have come down to us. The letter of the future will be a colorless communication of telegraphic brevity."

JULY 1853

IGNORANCE—"A terrible riot occurred Wednesday night at the residence of Dr. George A. Wheeler, New York, caused by the finding of some human bones on the premises. A mob of 3,000 collected, armed with clubs, axes, and stones. The premises were completely gutted by these savage ignoramuses. Nobody was killed, though some police officers were injured. Not one of the mob who had his arm or leg broken, but would get carried to a doctor to get it set, and how could the doctor do this unless he was acquainted with the anatomy of the human body?"

AGE OF THE EARTH—"Wonderful geological calculations were contained in a paper read by Sir Charles Lyell before the Royal Society in London, on the coal fields of Nova Scotia. He believes that the carboniferous formation of that country was once a delta like that of the Mississippi. If we include the coal fields of New Brunswick, there are 54,000 cubic miles of solid matter. It would take more than two millions of years for the Mississippi to convey to the Gulf of Mexico an equal amount of solid matter at a flow of 450,000 cubic feet per second. This is a subject for deep reflection and examination by all Biblical geologists especially."

COPYRIGHT 2003 SCIENTIFIC AMERICAN, INC.

OUTBREAKS

Disease Dustup

DUST CLOUDS MAY CARRY INFECTIOUS ORGANISMS ACROSS OCEANS **BY OTTO POHL**

SANDSTORM blows particulates out from the Sahara Desert in Africa (*landmass at right*) over the Canary Islands in the Atlantic Ocean. The storm occurred in February 2001.

ORBITAL IMAGING CORPORATION *Photo Researchers, Inc.*

On February 11, 2001, an enormous cloud of dust whipped out of the Sahara Desert and moved north across the Atlantic, reaching the U.K. two days later. A few days afterward, counties across the island began reporting simultaneous outbreaks of foot-and-mouth disease, a viral sickness of livestock (sometimes confused with mad cow disease). For Eugene Shinn, a geologist at the U.S. Geological Survey in St. Petersburg, Fla., that coincidence suggested an obvious link.

The idea that large-scale disease outbreaks could be caused by dust clouds from other continents has been floating around for years. But it seemed far-fetched. In the U.S. government, "no one wanted to listen to me," Shinn remembers about his proposal that something as amorphous and uncontrollable as a dust cloud could bring the disease to America.

But the theory is now gaining acceptance as scientists find that it may explain many previously mysterious disease outbreaks. Although the world's dry areas have always shed dust into the atmosphere—wind blows more than a billion tons of dust around the planet every year—the globe's dust girdle has become larger in recent years. Some of the changes are part of nature's cycles, such as the 30-year drought in northern Africa. Others, including the draining of the Aral Sea in Central Asia and the overdependence on Lake

COPYRIGHT 2003 SCIENTIFIC AMERICAN, INC.

LEAVING DDT IN THE DUST

Dust carries more than just disease. Ginger Garrison of the U.S. Geological Survey suspects that DDE, a breakdown product of DDT and a dangerous endocrine disruptor, is blowing over from Africa to the Caribbean. She is currently analyzing dust samples from Mali, the Caribbean and the ocean areas in between. She has also visited Mali to track the source of these toxic dust–borne chemicals. "There has been a definite change in what goes into the air in West Africa," she says. "In the past 12 to 15 years, there has been an incredible increase in the use of pesticides and plastics incineration."

Chad in Africa, are the result of shortsighted resource management. Poor farming practices also hasten desertification, creating dust beds polluted with pesticides and laced with diseases from human and animal waste.

For Shinn and his co-workers, it was a strange disease outbreak in the Caribbean in the early 1980s that first brought to mind the connection between dust and disease. A soil fungus began to attack and kill seafan coral. The researchers doubted that local human activity was the culprit, because the disease was found even in uninhabited places and islands devoid of soil. In addition, Garriet W. Smith of the University of South Carolina demonstrated that because the soil fungus could not multiply in seawater, it required a constant fresh supply to continue spreading.

Smith analyzed the African dust blowing across the Caribbean and was able to isolate and cultivate the soil fungus *Aspergillus sydowii,* with which he infected healthy seafans. USGS investigators then showed how the *Aspergillus* fungus and other organisms could survive the long trip from Africa protected by dense clouds of dust.

Researchers are now finding evidence that supports the link between sickness and dust. Ginger Garrison of the USGS believes that there is a direct link between bacteria-caused coral diseases such as white plague and black-band disease and African dust storm activity. In addition, outbreaks of foot-and-mouth disease in South Korea last year followed large dust storms blowing in from Mongolia and China.

Other organizations are now joining the USGS in tracking dust. NASA has satellites that are carefully monitoring dust storms, which can cover an area as large as Spain. The National Oceanic and Atmospheric Administration has just opened a station in California to track Asian dust as it passes over the U.S. (Although the SARS virus could theoretically cross oceans in a dust storm, the epidemiology so far indicates that person-to-person contact is the only way SARS has spread.)

The findings on international dust storms have also attracted the attention of those who are concerned about bioterrorism. "Anthrax will certainly make the trip" in dust from Africa to the U.S., remarks Shinn, who recently completed a terrorism risk assessment for the U.S. Dust clouds could be considered, in effect, a very dirty bomb.

Otto Pohl is based in Berlin.

ASTROPHYSICS

Frozen Stars

BLACK HOLES MAY NOT BE BOTTOMLESS PITS AFTER ALL **BY GEORGE MUSSER**

Demolishing stars, powering blasts of high-energy radiation, rending the fabric of spacetime: it is not hard to see the allure of black holes. They light up the same parts of the brain as monster trucks and battlebots do. They explain violent celestial phenomena that no other body can. They are so extreme, in fact, that no one really knows what they are.

Most researchers think of them as microscopic pinpricks, the remnants of stars that have collapsed under their own weight. But over the past couple of years, a number of mavericks have proposed that black holes are actually extended bodies, made up of an exotic state of matter that congeals, like a liquid turning to ice, during the collapse. The idea offers a provocative way of thinking about quantum gravity, which would unify Einstein's general theory of relativity with quantum mechanics.

In the textbook picture, the pinprick (or singularity) is surrounded by an event horizon. The horizon is not a physical surface, merely a conceptual one, and although it marks the point of no return for material plummeting toward the singularity, relativity says that nothing special happens there; the laws of physics are the same everywhere. For quantum mechanics, though, the event horizon is deeply paradoxical. It allows information to be lost from our world, an act that

COPYRIGHT 2003 SCIENTIFIC AMERICAN, INC.

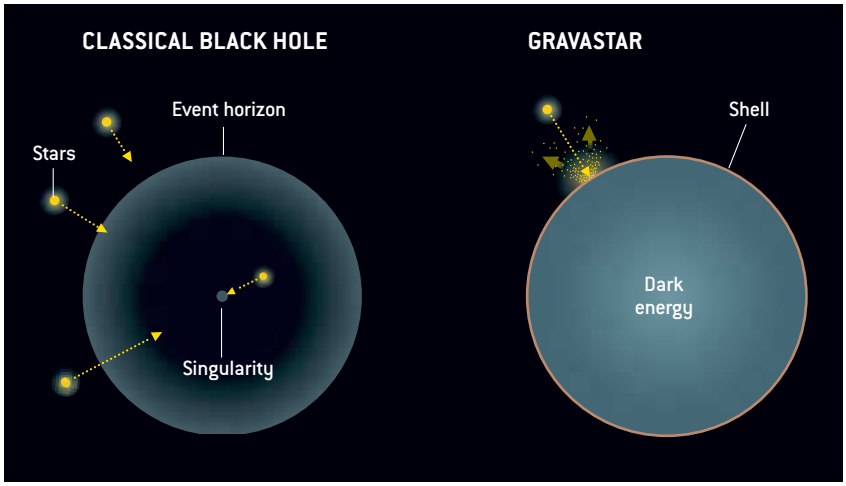
quantum theory forbids. "What you have been taught in school is almost certainly wrong, because classical black hole spacetimes are inconsistent with quantum mechanics," says physicist George Chapline of Lawrence Livermore National Laboratory.

The new conceptions of black holes eliminate the event horizon altogether. The basic idea is that there does, in fact, exist a force that could halt the collapse of a star when all else fails. That force is gravity itself. In matter with certain properties, gravity switches from being an attractive force to a repulsive force. Such a material, going by the name "dark energy," is thought to be driving the acceleration of cosmic expansion.

Last year physicists Pawel O. Mazur of the University of South Carolina and Emil Mottola of Los Alamos National Laboratory reasoned that a pocket of the stuff might freeze out, like ice crystals, during the collapse of a star. The result, which they call a gravastar, would look like fried ice cream: a crust of dense but otherwise ordinary matter stabilized by a curious interior. The crust replaces what would have been the event horizon.

Another proposal goes further. It conjectures not only that dark energy would freeze out but that relativity would break down altogether. The idea comes from a dark-horse contender for quantum gravity, the proponents of which are struck by the resemblance between the basic laws of physics and the behavior of fluids and solids (also known as condensed matter). In many ways, the equations of sound propagation through a moving fluid are a dead ringer for general relativity; sound waves can get trapped in the fluid much as light gets trapped in a black hole. Maybe spacetime is literally a kind of fluid.

What makes this approach so interesting is that the behavior of condensed matter is collective. The details of individual molecules hardly matter; the system's properties emerge from the act of aggregation. When water freezes, the molecules do not change, but the collective behavior does, and the laws that apply to liquids no longer do. Under the right conditions, a fluid can turn into a superfluid, governed by quantum mechanics even on macroscopic scales. Chapline, along with physicists Evan Hohlfeld, Robert B. Laughlin and David I. Santiago of Stanford University, has proposed that a similar process happens at event horizons. The equations of relativity fail, and new laws emerge. "If one thinks of spacetime as a superfluid, then it is very natural that in fact something physical does happen at the event horizon—that is, the classical event horizon is replaced by a quantum phase transition," Chapline says.



CLASSICAL VIEW portrays a black hole as an infinitely dense point (a singularity), which draws in matter such as stars, and an event horizon, which marks the point of no return. But in a black hole regarded as a ball of dark energy (a "gravastar"), infalling matter disintegrates at the dense shell.

For now, these ideas are barely more than scribbles on the back of an envelope, and critics have myriad complaints about their plausibility. For example, how exactly would matter or spacetime change state during the collapse of a star? Physicist Scott A. Hughes of the Massachusetts Institute of Technology says, "I don't see how something like a massive star—an object made out of normal fluid, with fairly simple density and pressure relations—can make a transition into something with as bizarre a structure as a gravastar." Mainstream theories of quantum gravity are far better developed. String theory, for one, appears to explain away the paradoxes of black holes without abandoning either event horizons or relativity.

Observationally, the new conceptions of black holes could be hard to distinguish from the classical picture—but not impossible. Gravitational waves should reveal the shape of spacetime around putative black holes. A classical hole, being a simple object without a true surface, has only a couple of possible shapes. If one of the gravitational-wave observatories now going into operation finds a different shape, then the current theories of physics would be yet another thing in the universe to get torn to shreds by a black hole.

UNEVENTFUL HORIZON

What would happen if you fell into a black hole? That depends on the theory. According to general relativity, you would feel weightless throughout your journey, even when you crossed the event horizon and entered the hole. Everything immediately around you would be falling in, too, so you would have no reason to suspect anything strange. The tidal forces that make a hole so deadly would not necessarily kick in until later. Only those of us watching from Earth would realize what was happening. "It would be impossible, in the framework of general relativity, to build a little self-contained sensor with an alarm that would go off and say, 'Warning: you have just fallen into a black hole—prepare to die,'" says physicist Scott A. Hughes of the Massachusetts Institute of Technology.

In the new models, however, there would be no doubt when you had reached the horizon: you would slam into a shell of hyperdense material, or the particles in your body would disintegrate into gamma rays.

SAMUEL VELASCO

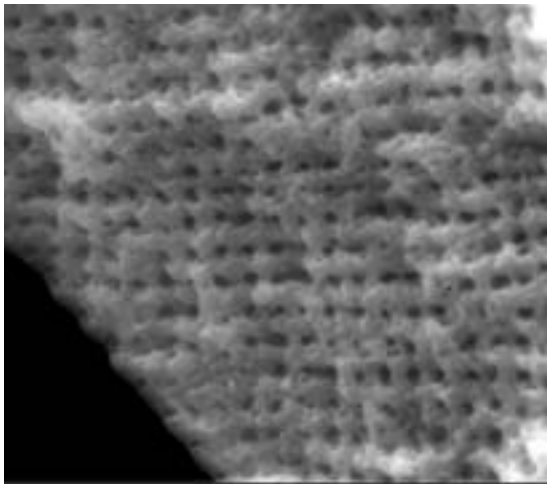
COPYRIGHT 2003 SCIENTIFIC AMERICAN, INC.

PHYSICS

Holographic Control

LIQUID-CRYSTAL HOLOGRAMS FORM PHOTONIC CRYSTALS **BY GRAHAM P. COLLINS**

Photonic crystals influence light in the way that semiconductor materials affect electric currents. Typically made out of a regular array of cavities in some refractive medium, a photonic crystal reflects or transmits light depending on the light's wavelength and the interplay of all the tiny wavelets scattered by the holes. In one respect, photonic crystals lag far behind their silicon-based cousins: it is difficult to modulate a photonic crystal's properties—for example, switching one from reflecting to transmitting. Recently research groups have demonstrated a versatile way to make a class of materials consisting of a polymer interspersed with liquid-crystal "droplets" whose optical response can be controlled by applying a voltage.



THREE-DIMENSIONAL LATTICE of liquid-crystal droplets (etched away for this micrograph) forms a photonic crystal whose properties can be controlled by applying a voltage. The holes are each separated by about 0.25 micron.

The fabrication begins with a soup of monomer molecules and liquid-crystal molecules, all sandwiched between two sheets of a substrate, such as glass plated with a thin layer of conducting material. The solution is irradiated with two or more laser beams, which are aligned and polarized to generate a specific interference pattern—the alternating dark and light areas that occur when laser beams overlap. (This is the hologram of the technique.) At the bright points in the pattern, the monomers link up and form a complex network of polymer. As this reaction proceeds, fresh monomers diffuse from the dark regions to the bright regions, causing the liquid crystal to accumulate in the dark regions. The end result is a solid polymer with droplets of liquid crystal embedded in a pattern corresponding to the dark regions of the holographic interference pattern.

The material functions as a photonic crystal, because the liquid-crystal droplets, whose optical axes are randomly oriented, scatter light. Active control of the photonic crystal is achieved by applying a voltage, which causes the optical axis of each droplet to line up. These aligned droplets present light with the same refractive index as the surrounding polymer matrix—the material becomes transparent, like a uniform piece of clear plastic.

When only two laser beams are used, the droplets are arranged in planes through the polymer, forming what are called diffraction gratings. Devices of this kind, which are technically not photonic crystals because they are structured in only one direction, were made as long ago as the late 1980s.

The idea of incorporating liquid-crystal material in true photonic crystals was proposed in 1999 by Kurt Busch of the University of Karlsruhe and Sajeev John of the University of Toronto. (John was also one of the originators of the basic photonic-crystal concept in 1987.) A typical early effort at realizing this idea involved a crystal made of close-packed spheres of silica with the intervening spaces filled with liquid crystal. This infiltration approach is limited, however, in the range of structures that can be made, and the construction requires several steps. Holographic fabrication, in contrast, can generate an arbitrary, regular lattice structure in a single step. Four or more laser beams are required to generate a fully three-dimensional array of droplets. The three-dimensional pattern is determined by the wavelengths and directions of the beams; the shapes and sizes of the individual droplets are determined by the relative polarizations and intensities.

Two groups have recently produced such photonic crystals. Timothy J. Bunning of the U.S. Air Force Research Laboratory at Wright-Patterson Air Force Base in Ohio, along with his co-workers, demonstrated a three-dimensional photonic crystal whose diffraction could be completely extinguished. At Brown University, Gregory P. Crawford, Michael J. Escuti and Jun Qi showed that three-dimensional crystals can switch from one optical state to another over a narrower voltage range than simpler one-dimensional gratings do. They also showed that a stop band—wavelengths blocked by a crystal—could be varied across a small range of wavelengths,

MICHAEL J. ESCUTI *Brown University*

A SHOCKING TRANSFORMATION

Shock waves could offer a new way to control a photonic crystal's properties. Using computer simulations, John D. Joannopoulos and his co-workers at the Massachusetts Institute of Technology predict that shock waves in a photonic crystal can have three dramatic and potentially useful effects on light. First, light can be trapped at a shock front for a controllable length of time. Second, the light can be "upconverted" to a higher frequency, even if the beam is weak (previously, the process occurred only in certain optical materials for very high intensity light). Finally, the bandwidth of the light can be narrowed by an order of magnitude, a feat achieved by no other nonquantum process.

COPYRIGHT 2003 SCIENTIFIC AMERICAN, INC.

opening the possibility of constructing a tunable filter.

The liquid-crystal photonic crystals made so far have comparatively weak optical properties because the liquid crystal bends light only slightly more than the polymer substrate does. A goal of current research is to devise crystals that have a bigger refractive index contrast. With their stronger optical effects, such crystals would be of greater use for applications such as switches, filters and reflective displays.

BIOTECH

Sugar Added

CHEAPER, BETTER PROTEIN DRUGS THROUGH SWEETENING

BY W. WAYT GIBBS

Proteins are the workhorses of biochemistry. They catalyze and metabolize; they regulate and signal; in the form of antibodies, they seek and destroy. Much of biotechnology entails searching for proteins that could serve as medicines and then making them in sufficient purity, potency and quantity.

The latter job may be the hardest one. The standard way to manufacture protein medicines—typically in huge fermenting vats filled with genetically engineered hamster ovary cells—is labor-intensive. Mammalian cells are complicated; it takes skill and attention to keep them fed and healthy.

That is one reason why, gram for gram, pharmaceutical-quality human proteins are dearer than gold. Omalizumab (trade name Xolair), a genetically engineered antibody recommended for approval in May for the treatment of asthma, is expected to cost about $10,000 a year, 10 times the price of existing asthma drugs. Despite the high cost of therapeutic proteins, demand for them is soaring. With nearly a dozen on the market and about 500 more in clinical trials, manufacturing capacity is becoming a major bottleneck.

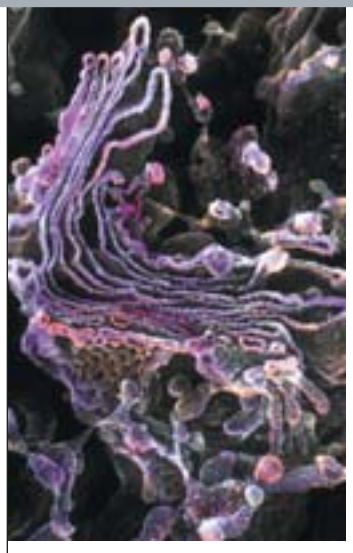
There are cheaper and more scalable ways to make medicinal proteins. Biotech firms have rejiggered the genes of all kinds of species to produce human proteins in the eggs of chickens, in the leaves and seeds of plants and in the bodies of insects. And for decades, companies have used vats of microorganisms to secrete some simple proteins, such as insulin.

But a serious limitation to these approaches is that they yield proteins that are incomplete. Only mammal cells naturally attach the right sequence of sugar molecules onto proteins that enables them to fold into the correct shape and achieve full potency within a human body. This process, called glycosylation, works differently in lower organisms than it does in humans and other mammals.

Recently several biotech firms and academic researchers have demonstrated new methods to manufacture glycosylated human proteins with the right sugary tails, no mammal cells required. GlycoFi, a startup in Lebanon, N.H., does it in yeast. At the March meeting of the American Chemical Society, GlycoFi chief scientist Tillman U. Gerngross reported that the company had engineered strains of *Pichia pastoris,* a yeast found in tree bark, that can make a desired human protein and attach nine of the dozen or so sugar molecules normally tacked onto it.

To do this, GlycoFi scientists drew random combinations from a library of more than 1,300 genes for enzymes. They managed to insert the genes into a yeast cell so that they are active in the endoplasmic reticulum—the part of the cell, along with the Golgi apparatus, where freshly minted proteins get glycosylated. "We're working toward a system that can rapidly generate every form of glycosylation on any given human protein and then automatically test to see which one has the best activity," Gerngross says.

In February, Gerd G. Kochendoerfer and his co-workers at Gryphon Therapeutics in San Francisco reported success with an even more straightforward approach: create the proteins from scratch using the same synthesis steps that are routine in chemicals factories. Proteins are long chains of linked amino acids. Gryphon scientists have inserted marker amino acids at specific points in the chain, then used them as handles to guide the reactants from one step to the next.



GOLGI APPARATUS contains the biochemical machinery in a cell that adds sugar molecules to proteins, increasing their potency.

P. MOTTA AND T. NAGURO/SCIENCE PHOTO LIBRARY

GOING FOR A CELL BREAK

Making protein drugs might be much cheaper and faster if cells were not involved. James R. Swartz of Stanford University has created cell-free liquids that synthesize human protein. An early version of the technique was licensed to Roche Pharmaceuticals, which now uses it in commercial production. The new version of the medium, which contains the innards of *Escherichia coli* bacteria, produces human proteins at less than half the cost of standard hamster-cell fermenters. It cannot yet add sugars to the proteins, but because the method is so much speedier and less expensive, Swartz says, "we are now thinking about producing patient-specific vaccines for people with lymphoma."

COPYRIGHT 2003 SCIENTIFIC AMERICAN, INC.

"We have total control of the protein structure, down to the last hydrogen atom," Kochendoerfer claims. The company was able to produce a synthetic version of human erythropoietin (which boosts red blood cell production and commands a $5-billion market) that is more than twice as potent as the hormone brewed in hamster cells. Protein yields are still low—1.5 to 20 percent, depending on the complexity of the product. Even so, the manufacturing cost is already competitive at large scales with standard fermentation techniques. And, he adds, "we can make proteins with multiple sugars attached to specific locations even more homogeneously than the human body [can]."

MATERIALS SCIENCE

Alloy by Design

COMPUTATIONS LEAD TO AN UNUSUALLY FLEXIBLE METAL **BY STEVEN ASHLEY**

Metallurgists have long sought to sit down at a computer, key in the elemental formulation for a new alloy, see how it works on the screen and then go into the lab to mix up a batch. Ideally, this digital development method would replace the tedious trial-and-error process that dates back to before medieval alchemists first tried to turn base metals into gold.

Word recently arrived from Japan that such progress may soon be in the offing. Researchers at the Toyota Central Research and Development Laboratories in Nagakute report in the April 18 *Science* that advanced computational models and tools led directly to the invention of a new class of titanium-based alloys.

Toyota's so-called gum metal alloys are strong, tough and heat-stable yet exhibit a remarkable degree of elasticity and plasticity over a temperature range extending from –200 to 300 degrees Celsius, according to Takashi Saito and his colleagues. The material can stretch several percent and return to its original length again and again, even after repeated elongations. In contrast, the best-known shape-memory alloy, nickel-titanium, which also demonstrates this superelasticity, soon gets hard and brittle with frequent deformations.

Though originally meant for automotive springs, gaskets and the like, the patented metal compositions are too costly for anything other than specialized premium applications such as microscrews, shape-recovery eyeglass frames (in production), medical implants and catheters, heat-tolerant springs for spacecraft, even long-hitting golf clubs.

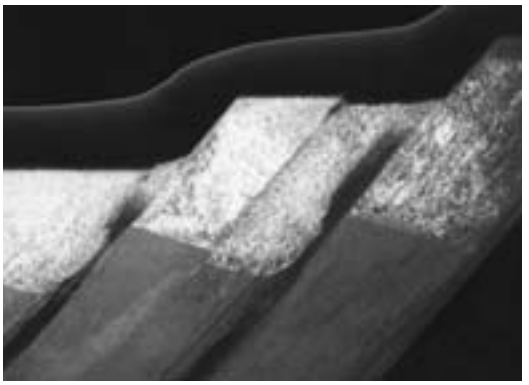
Gum metal alloys—composed of titanium, tantalum, niobium, zirconium and sometimes vanadium, together with a minimal leavening of oxygen—are consolidated into ingots under high heat and pressure but with no melting. The impressive physical qualities, Saito says, appear after vigorous cold-working, in which the metal is forced through a die at room temperature.

In designing the alloy, the Toyota team optimized three molecular properties. One was the number of bonds each metal atom forms with its neighbors. The others relate to the bond strength among atoms and the amount of attraction between outer electrons and nearby atoms. A combination of calculations, digital modeling and computer-directed experiments led to the final elemental recipes.

Several titanium experts in the U.S. are intrigued but as yet seem unconvinced by the inventors' explanation for the alloy's behavior and the theory they present to account for it. "Saito's group has a sterling reputation in the field," says Daniel Eylon, a materials engineer at the University of Dayton, but "we've all seen similar claims propounding new theories of plastic metal deformation that later turned out to be wrong. Metallurgists will need to see more data before we'll know if their work is truly significant. If it is, we'll all be learning some new physics."

DEFORMED THINKING

Takashi Saito and his Toyota co-workers have boldly proposed a new theory of plastic deformation to explain gum metal's surprising properties. In most metals, bending or other physical manipulation causes atoms in an ordered crystal structure to shift or dislocate. Continued deformation sends these "wrinkles" propagating through the crystalline lattice. Further repeated force eventually yields a tangled microscopic grain structure that is hard and brittle on the macroscopic scale. The Toyota researchers claim that their materials display no such atomic rearrangements. Instead their alloys respond by forming planar cracks, or "giant faults" (*below*), between crystalline sheets. These sheets then slide across one another like geologic plates. This energy-absorbing phenomenon results in gum metal's stretchy behavior, they believe.



MICROTECTONIC PLATES several hundred microns thick are thought to slide across one another, accounting for gum metal's enduring flexibility.

COURTESY OF TOYOTA CENTRAL RESEARCH AND DEVELOPMENT LABORATORIES

COPYRIGHT 2003 SCIENTIFIC AMERICAN, INC.

The Price of Pills

DOES IT REALLY TAKE $897 MILLION FOR A NEW THERAPY? BY CAROL EZZELL

Forty F16 jet fighters, or $802 million. That's how much it takes to develop a new drug, according to the first academic analysis of the process published in 12 years. That number reaches $897 million if postmarketing studies—additional clinical research that the U.S. Food and Drug Administration sometimes requires as a condition for approving a new drug—are taken into account, the report's authors announced in May.

These sky-high prices (in 2000 dollars) have prompted disbelief and consternation among some critics, who allege that the pharmaceutical industry is inflating the true cost of drug development to justify the escalating price tags of many therapies. The naysayers also accuse big pharma of seeking to justify its tax credits for research and development and to dissuade Congress from rolling back those benefits.

Drug companies often counter that clinical research—testing new therapies in patients—has gotten more difficult and therefore more expensive in recent years. Clinical trials for treatments for chronic diseases, such as arthritis, often require thousands of patients who must be followed for years, they say. Moreover, the companies cite statistics that only 21.5 percent of drugs that enter human tests ever make it to market, so they must recoup their costs on the therapies that do.

Who is right? It's hard to tell. The new analysis was led by economist Joseph A. DiMasi of Tufts University's Center for the Study of Drug Development, which receives roughly 65 percent of its funding from the pharmaceutical industry. (The funds are unrestricted—Tufts says companies cannot direct how they are spent.) But that connection worries some skeptics. James Love of the Washington, D.C.–based Consumer Project on Technology, one of the pharmaceutical industry's staunchest fault finders, comments that he considers the Tufts center "a think tank on behalf of industry."

Love and others note that the study relied on data confidentially provided by the companies. Ten pharmaceutical firms turned over cost information on a total of 68 randomly selected new drugs to DiMasi and his collaborators. The researchers' analysis placing the cost for developing a new drug at $802 million appears in the March *Journal of Health Economics* (they extended the figure to $897 million in May).

DiMasi bridles at the suggestion that the data were tainted or not representative of all new drugs under development. "The methodology was sound," he maintains. "I was satisfied that the people [from the drug companies who provided the data] were being honest with me." The clinical trials collectively involved 5,303 patients, with a price of roughly $23,000 per patient.

"The problem with these studies is they just don't jibe with publicly available data on the cost of clinical trials," argues Love, who says that a more realistic number is $10,000 to $12,000 per patient. "That would cover pretty much everything you'd want to do" to a given patient in any type of trial, he suggests. He also points out that DiMasi's cost-per-patient figure isn't the whole story, because it adds up to only $122 million. DiMasi counters that the other $680 million reflects preclinical research and the cost of failures.

In an editorial accompanying the Tufts article, Richard G. Frank of Harvard Medical School contends that the analysis considered only those drugs that were "new chemical entities" with little known about them, whereas many drugs in clinical testing are chemical relatives of existing drugs whose actions and side effects can be anticipated to a degree. Frank asserts that a significant proportion of drug development costs typically reflects business decisions to drop other drugs because of competition or market size. Still, he warns against tampering with the drug development process too much: "Regardless of the exact cost figure estimated," he writes, "if we are not cognizant of the complex, risky and costly process of drug development, public policy can damage an industry that has over the past generation bestowed enormous benefits on society by improving the effectiveness of health care."

VICTOR DE SCHWANBERG *Photo Researchers, Inc./SPL*

news SCAN

A BALLOONING NUMBER

Analyzing the cost of drug research and development is tough because the industry protects such information for competitive reasons. In 1987 economist Joseph A. DiMasi of Tufts University led a research group that put the price at $231 million, or $318 million in 2000 dollars. But by the mid-1990s pharmaceutical firms were routinely citing a $500-million-per-new-drug figure, which apparently stemmed from an attempt by the Boston Consulting Group to extrapolate upward for the increasing size of clinical trials. That number was bandied about until last year, when a drug company executive announced the new $802-million number at a conference, months before the full analysis appeared in the scientific literature.

COPYRIGHT 2003 SCIENTIFIC AMERICAN, INC.

DATA POINTS: SPACE WORMS

Tiny worms survived the disastrous February 1 reentry of the space shuttle *Columbia*. The species, *Caenorhabditis elegans*—sent into space to test a synthetic nutrient source—was found in canisters among the shuttle's debris. These soil worms are popular model organisms in biology, providing crucial information about such concepts as programmed cell death and longevity. In 1998 *C. elegans* became the first multicellular organism to have its genome sequenced.

Length of adult: 1 millimeter

Number of cells: 959

Life span: 2 to 3 weeks

Number of generations that the recovered worms were from the original space worms: 4 or 5

Number of base pairs in *C. elegans* genome: 97 million

Number in human genome: 3 billion

Number of protein-coding genes in *C. elegans*: 19,099

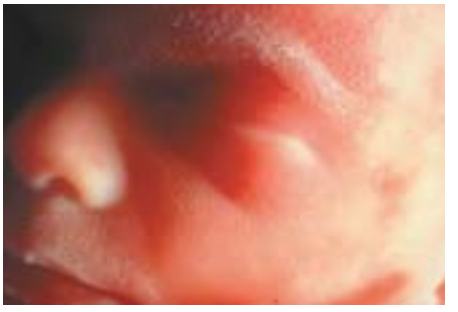
Number in human being: 30,000 to 40,000

SOURCES: NASA; Science, December 11, 1998; press release, Nobel Prize in Physiology or Medicine, 2002

BIOLOGY

Sugar and Spice and Everything Nice . . .

Mother Goose may be right. Experiments in mice reveal that meals high in sweets and low in fats led female rodents to produce twice as many female pups than males. The reverse was true for mothers on low-sugar, lard-filled diets. The experiments, done by a group at the University of Missouri–Columbia, support decades of anecdotal evidence connecting diet to the sex of offspring in mammals. Meals could be hormonally swaying the female reproductive tract so that embryos of one sex have a survival advantage over the other. Diet could also affect how X or Y chromosome–bearing sperm fertilize eggs. Or, as team member Cheryl S. Rosenfeld notes, the energy content of the food could have skewed the sex ratios, because the fatty diet was higher in calories. The findings appear in the April 15 *Proceedings of the National Academy of Sciences USA*. —*Charles Choi*



COULD A BABY'S SEX arise from the mother's diet? A mouse study raises the question.

TOXICOLOGY

Pass the Sushi

Mercury in fish eaten during pregnancy might not threaten children after all. University of Rochester researchers looked at the women of the Seychelles, who ate an average of 12 fish meals a week and had six times the mercury levels of a typical American, yet their children showed no meaningful cognitive problems. Past studies may have found a link because the women involved ate whale meat, which has five times the mercury concentrations of the more common ocean fish consumed in the Seychelles. The work appears in the May 16 *Lancet*. —*Philip Yam*

ENGINEERING

Sound Off on Tires

When the rubber hits the road, the ears inevitably pay a price. In the search for quieter highways, Purdue University engineers crafted a 19-ton, 12-foot-wide round apparatus that rolls tires over pavement while microphones record emanating tones and sound levels at several frequencies and distances. Other techniques either drag tires behind vehicles or ride stationary tires on motorized steel rollers, which never possess the exact traits of real pavement. The Purdue apparatus more completely mimicks how tires generate noise under many environmental conditions. The engineers tested smooth, porous and textured concrete surfaces with four tire designs. Preliminary findings show that pavement type, rather than tire design, dictated the level of noise, with porous surfaces generating the least amount. Researchers do not yet understand the precise roots of highway noise but suspect that tread grooves funnel air, thereby acting like tiny pipe organs, or that they vibrate at noise frequencies when they strike or peel away from pavement. —*Charles Choi*

FAST, FURIOUS—and loud. The type of concrete pavement affects road noise.

NEIL BROMHALL *Genesis Films/SPL (top)*; STEPHEN MALLON *Photonica (bottom)*; ILLUSTRATION BY MATT COLLINS

COPYRIGHT 2003 SCIENTIFIC AMERICAN, INC.

PHYSIOLOGY

Muscle Maintenance

MUSCLE USE tears membranes that must be repaired.

Everyday wear and tear ends up riddling muscle membrane with tiny holes. But healthy cells repair themselves quickly by releasing an armada of vesicles, which carry important chemicals to the site. The compounds patch the hole in just 10 to 30 seconds, and, according to a mouse study, the key repair substance is a protein called dysferlin. The absence of dysferlin has been known to result in two rare types of muscular dystrophy—Miyoshi myopathy and limb-girdle muscular dystrophy (type 2b). Evidently, in these dystrophies the muscles cannot be repaired as they get damaged over time, says principal investigator Kevin P. Campbell of Iowa University, who reports the findings in the May 8 *Nature*. Dysferlin may also help maintain cell health in other organs: it resides in heart, brain and ear tissues as well. *—Laura Wright*

PHYSICS

Mystery Meson

A strange and charming particle has been added to the subatomic zoo, but the newfound creature is far from what scientists expected. Named D_S (2317), the exotic particle was discovered by the BaBar detector, which inspects the debris of high-energy electron-positron collisions at the Stanford Linear Accelerator Collider. The new particle might be one of eight theorized pairs of charm and strange quarks and their antimatter counterparts (charm and strange are two of the six flavors of quarks). Quarks are the fundamental particles that in trios make up protons and neutrons; quark duos are called mesons. Four other charm-strange mesons were found before, all fitting predictions, but the just discovered particle is clearly and bafflingly some 10 percent lighter than expected. In their paper submitted to *Physical Review Letters,* the BaBar team has even proposed that D_S (2317) is a long-predicted, never-seen combination of four quarks. Data from the CLEO detector at Cornell University confirm BaBar's sighting and also suggest that the particle can exist at a higher energy level and thereby be slightly heavier. The results could rewrite what scientists know of the universe's most powerful fundamental force, the strong nuclear interaction. *—Charles Choi*

ENTOMOLOGY

Ant Thesis

SWARMING is a typical army ant trait.

In the 1954 movie *The Naked Jungle,* Charlton Heston tries to save his coffee plantation from marauding army ants, which thrive throughout the world's tropics. Heston's character was probably too preoccupied to assume, as entomologists have, that the typical army ant traits—nomadism, foraging without scouting and wingless queens producing up to four million eggs a month—evolved numerous times in species around the globe. But Cornell University's Sean Brady, while at the University of California at Davis, compared the DNA, morphology and fossils of some 30 army ant species and came to an unanticipated conclusion: a common army ant ancestor first emerged on the supercontinent Gondwana about 100 million years ago and spread as the continents separated. "This group represents an extraordinary case of long-term evolutionary stasis," Brady writes in the May 27 *Proceedings of the National Academy of Sciences USA*. In other words, if it ant broke, don't fix it. *—Steve Mirsky*

PURE/NONSTOCK (*top*); GEORGE BERNARD *SPL* (*bottom*)

BRIEF POINTS

- **Monarch butterflies on their autumn migration rely on internal clocks, which enable them to determine their flight path relative to the sun at any time of day. The study may mark the first direct evidence for circadian rhythms in celestial navigation.**

  Science, May 23, 2003

- **Personalities can change past the age of 30; in most people, the degree of conscientiousness and agreeableness increased, whereas anxiousness, openness and extraversion declined.**

  Journal of Personality and Social Psychology, May 2003; also, www.sciam.com/news_directory.cfm

- **Ebola to the rescue: part of the deadly virus could be used to create a hybrid virus that could efficiently deliver healthy genes to airway cells damaged by cystic fibrosis.**

  Journal of Virology, May 10, 2003

- **Smokers going cold turkey develop an altered sense of time. After abstaining for 24 hours, they estimated a 45-second interval to be 50 percent longer than nonsmokers did.**

  Psychopharmacology, May 2003

COPYRIGHT 2003 SCIENTIFIC AMERICAN, INC.

BY THE NUMBERS

Winners and Losers

THE BENEFITS OF GLOBALIZATION ARE SPREAD UNEVENLY **BY RODGER DOYLE**

Why are the biggest winners in the past decade of trade globalization mostly in South and East Asia, whereas the biggest losers are mostly in the former Soviet bloc and sub-Saharan Africa? History is a partial guide: East Asia has a millennia-old trading tradition, lately reinvigorated by the Chinese adoption of market economics. The Soviet Union, on the other hand, was sheltered from free-market forces for more than 70 years. In Africa, civil strife or inadequate infrastructure, which results in high transport costs, has hobbled a number of economies. Some are disadvantaged because they are landlocked; many have little to trade but commodities, prices of which have fallen in recent years.

In some regions, certain countries have suffered by adopting misguided policies, often under pressure from international institutions such as the International Monetary Fund. First among these is Russia, which in the early 1990s tried to embrace capitalism before first building the institutions that make capitalism work, such as an independent banking system, a system of business law, and an adequate method for collecting taxes. Encouraged by the IMF, the World Bank and the U.S. Department of the Treasury, President Boris Yeltsin's regime privatized the state-owned industrial sector, creating a class of oligarchs, who, knowing how unstable conditions were at home, sent their money abroad instead of investing it at home. Under IMF pressure, Russia imposed an overvalued exchange rate, a boon to those who imported luxury goods but a depressant for exporting industries. The result was a disaster for employees, who were frequently not paid or paid in goods, not rubles.

In contrast, China, the biggest winner from globalization, did not follow the IMF formula. Of the former states of the Soviet bloc, only a few, notably Poland and Hungary, managed to grow, which they did by ignoring IMF advice and adopting expansionary plans, including spending more than they collected in taxes. Botswana and Uganda are also success stories: despite their disadvantages, these countries achieved vigorous growth by creating stable civil societies, liberalizing trade and implementing reforms that ran counter to IMF prescriptions.

The IMF has, by its own admission, pursued failed policies in developing countries. Its original mission was to sustain the world economy by promoting full employment. But in the past few decades, according to Columbia University's Joseph E. Stiglitz, winner of the 2001 Nobel Prize in economics, the agency has come to be dominated by economists who are more attuned to the financial community than to the borrowing countries. Believing that fiscal austerity is beneficial, the IMF imposed counterproductive contractionary policies on borrowing countries as the price for loans. Stiglitz sees tentative signs that the IMF and other international institutions such as the World Bank are changing their approach.

If correct, Stiglitz's observation would be welcome news, for trade globalization has been a great force in reducing poverty. In China, for example, the number of people living in rural poverty went from 250 million in 1978 to 34 million in 1999. In the less globalized countries, poverty rose 4 percent from 1993 to 1998, and in Russia, it increased from 2 percent in 1989 to 24 percent in 1998.

Rodger Doyle can be reached at rdoyle2@adelphia.net

FURTHER READING

Globalization: A Critical Introduction. Jan Aart Scholte. Palgrave, Macmillan, 2000.

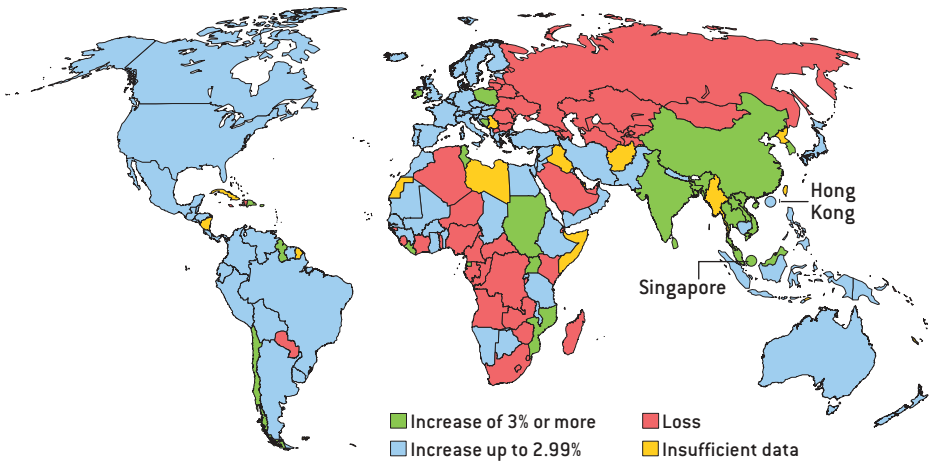
Globalization: Neoliberal Challenge, Radical Responses. Robert Went. Pluto Press and the International Institute for Research and Education, 2000.

Alternatives to Economic Globalization. International Forum on Globalization. Berrett-Koehler Publishers, 2002.

Globalization and Its Discontents. Joseph E. Stiglitz. W. W. Norton, 2002.

Globalization/Anti-Globalization. David Held and Anthony McGrew. Polity Press, 2002.

Average Annual Change in Per Capita Gross Domestic Product, 1990–2001



SOURCE: World Bank. Data are adjusted to constant purchasing power basis using 1995 U.S. dollars.

RODGER DOYLE

COPYRIGHT 2003 SCIENTIFIC AMERICAN, INC.

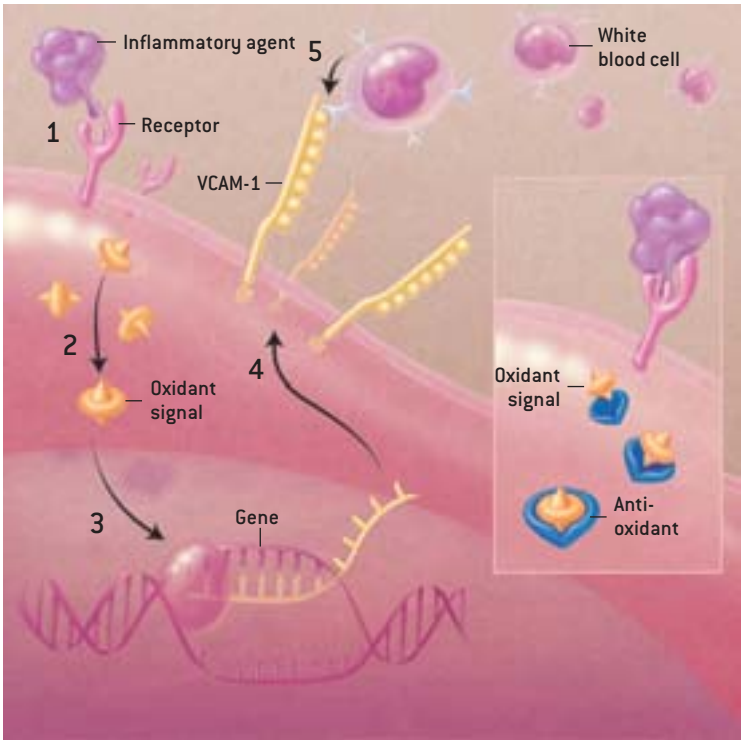
Signal Jammer

An academic experiment leads to a new class of drug for attacking heart disease

By GARY STIX

In the early 1990s the root causes of atherosclerosis had started to become clearer. Emerging research showed that the disease bore a direct relation to inflammation triggered by lipoproteins and other agents that spurred growth of atherosclerotic deposits.

Taking note of these discoveries, clinical researchers had begun to mull how they could intervene to block this process. Two professors at the Emory University School of Medicine—Russell M. Medford and R. Wayne Alexander, both cardiologists and biologists—were intrigued by findings that tied inflammation to oxidants, molecules that strip electrons from other molecules.



EARLY-STAGE ATHEROSCLEROSIS begins when an inflammatory agent attaches to an endothelial cell surface receptor (*1*). Resulting changes in the shape of the receptor generate an oxidant signal (*2*), which turns on a gene (*3*). The gene gives rise to a protein called vascular cell adhesion molecule-1 (VCAM-1) that migrates to the cell surface (*4*), where white blood cells attach to it (*5*), spurring events that lead to the buildup of atherosclerotic deposits. This process can be stopped if an antioxidant, such as AGI-1067, blocks the oxidant signal (*inset*).

ALICE Y. CHEN

Medford and Alexander theorized that oxidants might be involved in activating the genes that initiate the inflammatory process. An oxidant within one of the endothelial cells that make up the inner lining of a blood vessel might respond to the oxidized form of low-density lipoprotein (LDL, the "bad" cholesterol) by issuing an alarm that turns on the relevant genes that produce inflammation.

Studying cell cultures, the researchers found a type of oxidant, a lipid peroxide, that led to the activation of several genes, including one for vascular cellular adhesion molecule-1 (VCAM-1). Involved early on in the inflammatory process, VCAM-1 recruits white blood cells, including monocytes and lymphocytes, to the surface of the endothelial cell, initiating the chronic inflammatory reaction that ultimately results in atherosclerosis. In their experiments, Medford and Alexander observed that an antioxidant, pyrrolidine dithiocarbamate (PDTC), prevented the expression of the VCAM-1 genes.

The experiment suggested a new way to treat coronary artery disease. Statins, widely used anticholesterol drugs, work primarily by lowering levels of LDLs. But half the victims of heart attacks and angina that result from atherosclerosis do not experience elevated lipid levels. Thus, other triggers—including diabetes, high blood pressure, or chemicals in cigarette smoke—can also initiate the signals that lead to chronic inflammation. Medford and Alexander realized that they might have devised a method of treating coronary artery disease that dealt with more than a single risk factor (such as high cholesterol). "The notion of this pathway was that it may be fundamental signaling involved in atherosclerosis," Alexander says. "And, if so, it could attack the disease more directly and be more effective than

COPYRIGHT 2003 SCIENTIFIC AMERICAN, INC.

SCIENTIFIC AMERICAN
CUSTOM PUBLISHING

IT'S NOT THE MESSAGE.
IT'S NOT THE MEDIUM.

IT'S BOTH.

Only a sophisticated and innovative approach will reach a sophisticated and innovative audience.

Scientific American's Custom Publishing Program leverages the authority and expertise of Scientific American to meet YOUR organization's communications and marketing needs.

See what the Scientific American Custom Publishing Program can do for YOUR organization...

extend
or enhance the reach of your conference/symposium

deliver
in-depth coverage of a product launch or a new corporate initiative

build
your corporate image as a leader in cutting-edge technology

connect
and maintain an ongoing relationship with your customers

Contact us for a consultation
212.451.8859
or
jabbate@sciam.com

statins." Before these conclusions were published in the *Journal of Clinical Investigation* in 1993, Emory's technology transfer office filed for a patent.

Neither Alexander nor Medford had plans to start a company. But Michael A. Henos of Alliance Technology Ventures, a former Silicon Valley denizen with a specialty in biotechnology, persuaded them to do just that.

The first employees that chief executive Medford hired for the company, named Atherogenics, were medicinal chemists. PDTC, the compound previously assayed, was somewhat toxic and would clearly not make a suitable drug candidate. The new hires set about testing whether Probucol, an old-line anticholesterol medication, would prove to be a "chemical point of entry," a starting place for creating a drug.

In principle, Probucol's powerful antioxidant properties made it attractive, if its chemical structure could somehow be altered to eliminate its liabilities—among other things, it had trouble getting into endothelial cells and was sometimes linked to cardiac rhythm abnormalities. The molecule was symmetric in shape. The two phenol groups at each end of its linear molecular chain were what caused the drug to act as an antioxidant. The medicinal chemists removed one or both phenols, replaced them with a number of other side groups, and tested the activity of the altered molecules. One of the molecules tried, AGI-1067, had qualities superior to any other candidate tested. In AGI-1067, one phenol was replaced with another organic compound (an ester—more specifically, a succinate hydroxyl group). Leaving one phenol allowed it to remain as potent an antioxidant as Probucol, and the addition of the ester let it penetrate cells and obviated any safety concerns.

Preclinical testing during the mid-1990s in cell cultures and animals went well enough that Atherogenics could proceed with multiple rounds of venture financing and move into human clinical trials in 1998. The company chose to do a trial to counter restenosis—the narrowing of arteries after they have been propped open with wire-mesh tubes known as stents. It did so because it figured that a study for a narrowly focused condition might gain U.S. Food and Drug Administration approval more readily than a trial for atherosclerosis, the main objective of the drug developers. Atherogenics also took the essential step of partnering with a pharmaceutical manufacturer, Schering Plough, in October 1999.

AGI-1067 increased the volume of the artery's blood-flow channel—suggesting a reversal of atherosclerosis.

Ultimately, the partnership proved to be a poor match. Apparently distracted by internal drug-development programs, Schering did not want to proceed rapidly with AGI-1067. The arrangement collapsed two years later, sending Atherogenics's stock plummeting (the company had gone public in 2000). Six weeks after the split, Atherogenics received the results of an analysis done after its Phase II clinical trial, a six-week test of drug safety and effectiveness that encompassed 305 patients. From the Phase II results, Atherogenics already knew that the compound had shown good results in inhibiting restenosis. But the later analysis of a large subgroup from the Phase II trial showed that in a nearby, unstented part of an artery, the volume of the blood-flow channel had increased—suggesting a reversal of atherosclerosis.

These reports—which were followed by another Phase II trial—marked a turning point. The company got endorsement from the FDA to mount a large-scale Phase III trial recruiting 4,000 patients

COPYRIGHT 2003 SCIENTIFIC AMERICAN, INC.

with previous heart attacks or unstable angina. The intent of the new study is to determine what result AGI-1067 can achieve beyond existing treatments (such as statins) in patients who have had a heart attack or other coronary problems. The FDA gave Atherogenics approval to begin its Phase III trial to test the compound on atherosclerosis and to leave the more narrow indication of restenosis for later study.

Wall Street has taken notice. The company had no trouble issuing $50 million in new stock this past January, a time when the biotechnology industry was in a severe slump. Some analysts believe that AGI-1067 could draw at least $1 billion in annual revenues. But hurdles may remain. The company will be looking at the compound's effects on the good cholesterol (HDL), which was lowered in one trial.

Atherogenics is now planning to enter into a partnership with a pharmaceutical house capable of marketing the drug. The trials so far could translate into better royalties and upfront payments than the now defunct early-stage agreement struck with Schering would have. The development of AGI-1067 occurred in tandem with a growing understanding that antioxidants that quell inflammation might treat conditions ranging from rheumatoid arthritis to asthma to organ transplant rejection. To diversify beyond atherosclerosis, the company has more preliminary drug-development efforts for these maladies. And it has licensed a patent to create drugs for another inflammatory pathway.

The next two to three years will be crucial for Atherogenics. Any setbacks with AGI-1067 would cause a steep drop in the stock price and make it much harder to raise additional capital to move ahead with alternative drug candidates. The Phase III test of the compound will be the critical link along the path to what could become the first of a wholly new class of cardiovascular drugs. SA

FINALLY, BUSINESS SOLUTIONS THAT WORK WITH EXISTING TECHNOLOGIES AND NONEXISTENT BUDGETS.

You need to get more out of what you have. We have just the thing: solutions based on our open technology platform, SAP NetWeaver.™ Because it's preconfigured to work with your current IT investments – and it's fully operable with .NET and J2EE – SAP NetWeaver reduces the need for custom integration. That lowers your total cost of ownership for your entire IT landscape and gets you quicker ROI. Everything a CIO wants (and a CFO didn't think was possible). Visit sap.com/netweaver or call 800 880 1727 for details.

THE BEST-RUN BUSINESSES RUN SAP

SAP

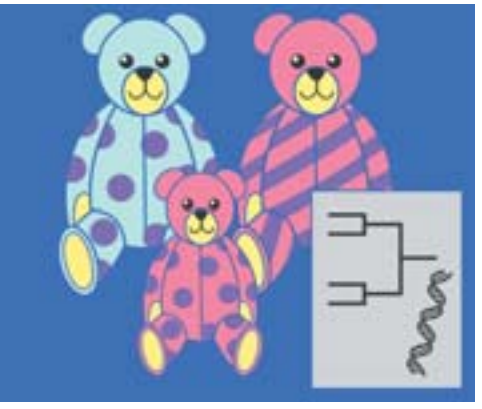
©2003 SAP AG. SAP and the SAP logo are registered trademarks of SAP AG in Germany and several other countries. Other product or service names mentioned herein are the trademarks of their respective owners.

COPYRIGHT 2003 SCIENTIFIC AMERICAN, INC.

You Can Patent *That?*

A selection of recently issued intellectual-property gems By GARY STIX

The U.S. Patent and Trademark Office issues several thousand new patents a week. Not everything that originates from the patent office is just another variation on a rotary valve or a mobile communications device. In each weekly batch, a number of issuances demonstrate both the scope of human ingenuity and the expansive breadth of what the patent office considers novel, useful and inventive. What follows are a few recent highlights that, if nothing else, transcend the mundane.



Sildenafil citrate chewing gum formulations and methods of using the same, patent 6,531,114, assigned to Wm. Wrigley Jr. Company. According to the text of the patent: "We claim a method for treating erectile dysfunction in an individual comprising the steps of providing a chewing gum composition that includes a therapeutically effective amount of sildenafil citrate in the chewing gum composition." Chewing causes the drug "to be released from the chewing gum composition into the oral cavity of the individual." Sildenafil citrate is better known as Viagra.

Warren portal identification and tunnel resident disgorger system, patent 6,474,601, Richard Krobusek and David H. Hitt of Plano, Tex. A jet engine is aimed at the mouth of a cave. "Running the jet engine at idle … creates a significant volumetric flow of exhaust gases, including significant quantities of carbon dioxide and carbon monoxide," the patent states. "These gases displace the oxygen that the terrorists require to breathe." The engine can also be run at cruise power, which "causes significant airflow and force to be applied to those persons and objects in the warren. Therefore, the terrorists are assaulted through their sense of touch as they are blown about in the warren."

Registered pedigree stuffed animals, patent 6,482,067, David L. Pickens of Honolulu. From the patent: "A pair of opposite sex 'parent' toy animals are sold together with a serial number by which the parent's genotype and phenotype may be identified. The owner or owners of the 'parent' toy animals may register the parents with the manufacturer and subsequently request 'breeding' of the animals, whereupon the manufacturer makes at least one 'offspring' toy animal randomly selected from a litter having phenotypes [traits] determined according to the registered genotypes of the parents and the Mendelian laws of inheritance."

Nervous system manipulation by electromagnetic fields from monitors, patent 6,506,148, Hendricus G. Loos of Laguna Beach, Calif. A pulsed electromagnetic field, from either a television set or a computer, can be created to manipulate the human nervous system, inducing sensations that range from relaxation to a "tonic smile," to sexual excitement or "sudden loose stool." Sometimes the pulses cannot be seen on the monitor. "This is unfortunate," the patent notes, "since it opens a way for mischievous application of the invention, whereby people are exposed unknowingly to manipulation of their nervous systems for someone else's purposes. Such application would be unethical and is of course not advocated. It is mentioned here in order to alert the public to the possibility of covert abuse that may occur while being online…."

Semen taste-enhancement dietary supplement, patent 6,485,773, Lois Kay Myers and Brent Richard Myers of Apache Junction, Ariz. Details can be sought by getting in touch with the U.S. government. Go to the patent and trademark office site (www.uspto.gov) and type in the patent number in the "search patents" section. Then read about a formulation that could complement the aforementioned Wrigley patent. SA

More offbeat patents will be included in next month's Staking Claims column.

JENNIFER KANE

COPYRIGHT 2003 SCIENTIFIC AMERICAN, INC.

Bottled Twaddle

Is bottled water tapped out? By MICHAEL SHERMER

In 1979 I started drinking bottled water. My bottles, however, contained tap water and were nestled in small cages on the frame of my racing bicycle.

Tap water was good enough then because we did not know how much healthier and tastier bottled water is. It must be, because Americans today spend more than $7 billion a year on it, paying 120 to 7,500 times as much per gallon for bottled water as for tap. Bottled prices range from 75 cents to $6 a gallon, versus tap prices that vary from about 80 cents to $6.40 per *1,000* gallons. We wouldn't invest that for nothing, would we?

Apparently we would. In March 1999 the Natural Resources Defense Council (NRDC) published the results of a four-year study in which they tested more than 1,000 samples of 103 brands of bottled water, finding that "an estimated 25 percent or more of bottled water is really just tap water in a bottle—sometimes further treated, sometimes not." If the label says "from a municipal source" or "from a community water system," it's tap water.

Some bottled water is really just tap water in a bottle.

Even more disturbing, the NRDC found that 18 of the 103 brands tested had, in at least one sample, "more bacteria than allowed under microbiological-purity guidelines." About one fifth of the waters "contained synthetic organic chemicals—such as industrial chemicals (e.g., toluene or xylene) or chemicals used in manufacturing plastic (e.g., phthalate, adipate, or styrene)," but these were "generally at levels below state and federal standards." The International Bottled Water Association issued a response to the NRDC study in which it states, "Close scrutiny of the water quality standards for chemical contaminants reveals that [the U.S. Food and Drug Administration's] bottled water quality standards are the same as [the Environmental Protection Agency's] tap water standards." Well, that's a relief, but in paying exceptional prices one might hope for exceptional quality.

One problem is that bottled water is subject to less rigorous purity standards and less frequent tests for bacteria and chemical contaminants than those required of tap water. For example, bottled-water plants must test for coliform bacteria once a week; city tap water must be tested 100 or more times a month.

If bottled water is not safer (a 2001 World Wildlife Fund study corroborated the general findings of the NRDC), then surely it tastes better? It does … as long as you believe in your brand. Enter the water-wars hype. Pepsi introduced Aquafina, so Coke countered with Dasani, a brand that included a "Wellness Team" (meet Susie, Jonny and Ellie, the "stress relief facilitator," "fitness trainer" and "lifestyle counselor," respectively) on its Web site. Both companies charge more for their plain water than for their sugar water.

When the test is blind, however, the hype falls on deaf taste buds. In May 2001 ABC's *Good Morning America* found viewers' preferences to be Evian (12 percent), O-2 (19 percent), Poland Spring (24 percent) and good old New York City tap (45 percent). In July 2001 the *Cincinnati Enquirer* discovered that on a 1-to-10 scale, that city's tap water rated an 8.2, compared with Dannon's 8.3 and Evian's 7.2. In 2001 the Yorkshire, England, water company found that 60 percent of 2,800 people surveyed could not tell the difference between the local tap water and the U.K.'s bottled waters.

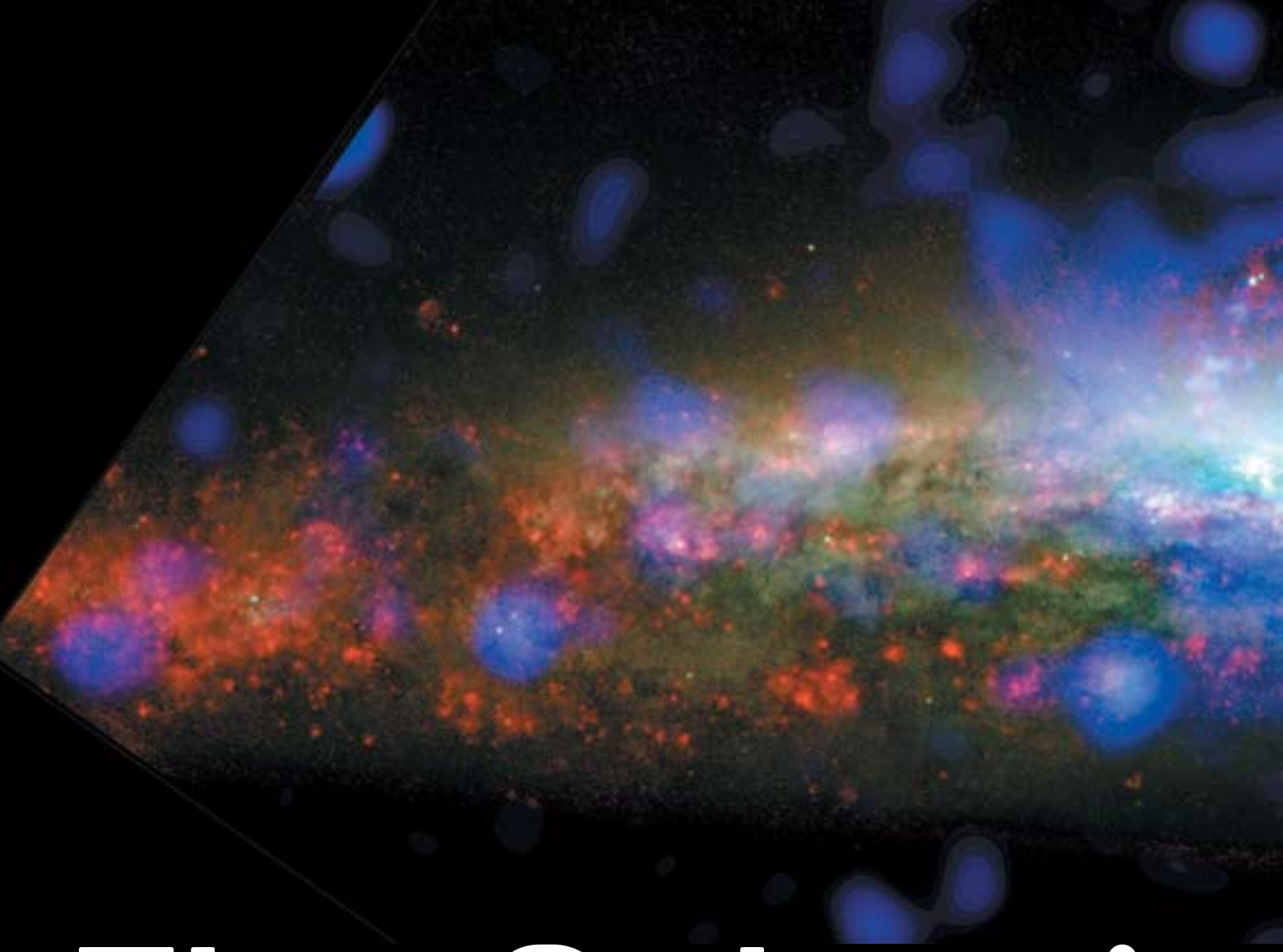
The most telling taste test was conducted by the Showtime television series *Penn & Teller: Bullshit!* The hosts began with a blind comparison in which 75 percent of New Yorkers preferred city tap to bottled waters. They then went to the Left Coast and set up a hidden camera at a trendy southern California restaurant that featured a water sommelier who dispensed elegant water menus to the patrons. All bottles were filled out of the same hose in the back of the restaurant; nevertheless, Angelenos were willing to plunk down nearly $7 a bottle for L'eau Du Robinet (French for "faucet water"), Agua de Culo (Spanish for "ass water") and Amazone ("filtered through the Brazilian rain forest's natural filtration system"), declaring them all to be far superior to tap water. There's no accounting for taste.

Bottled water does have one advantage over tap: you can take it with you wherever you go. So why not buy one bottle of each desirable size and refill it with your city's finest unnaturally filtered yet salubriously delicious tap water? SA

Michael Shermer is publisher of Skeptic *(www.skeptic.com) and author of* Why People Believe Weird Things.

BRAD HINES

COPYRIGHT 2003 SCIENTIFIC AMERICAN, INC.

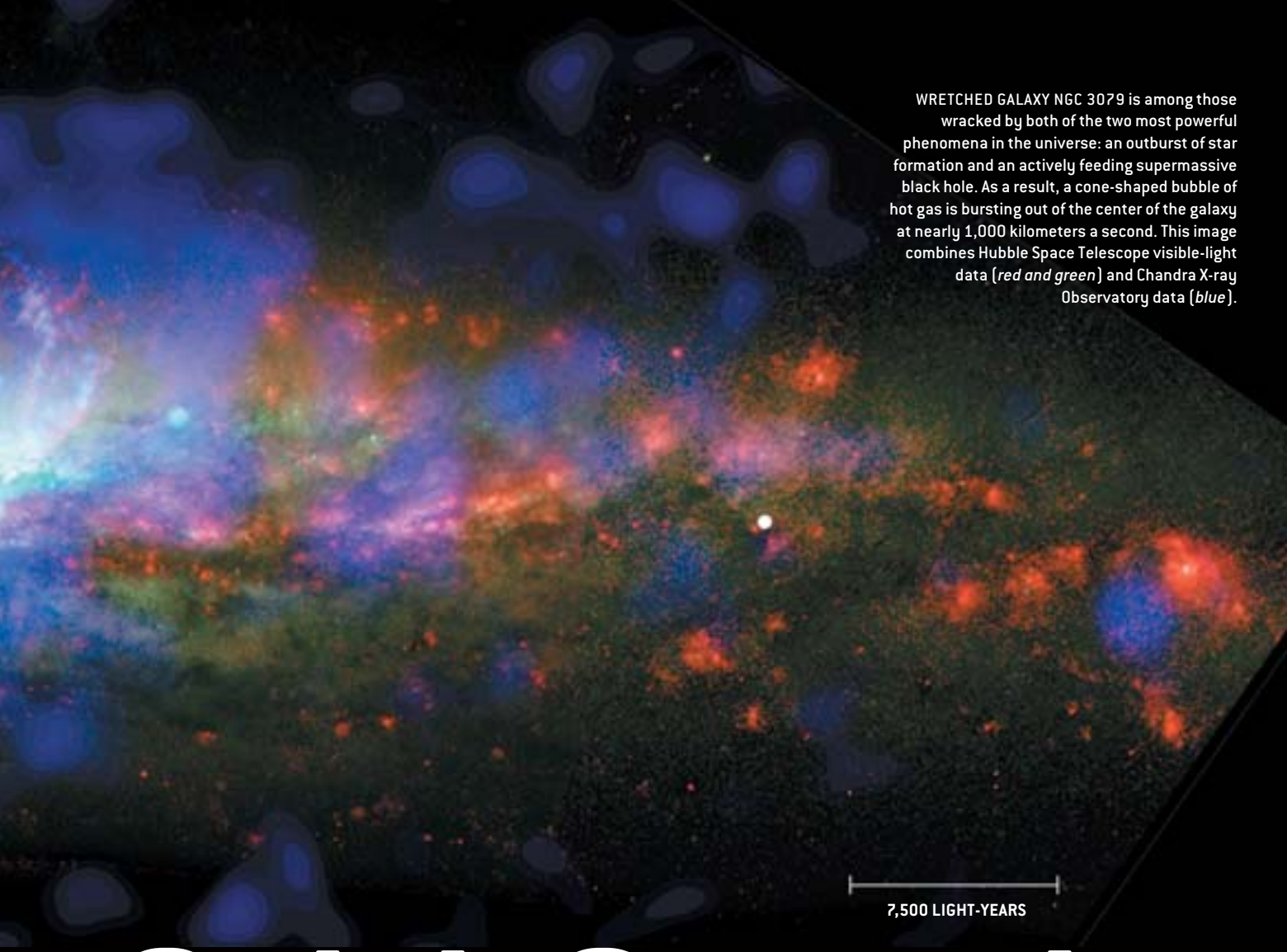


The Galactic

Why do GIANT BLACK HOLES and STELLAR BABY BOOMS, two phenomena with little in common, so often go together?

BY KIMBERLY WEAVER

COPYRIGHT 2003 SCIENTIFIC AMERICAN, INC.



WRETCHED GALAXY NGC 3079 is among those wracked by both of the two most powerful phenomena in the universe: an outburst of star formation and an actively feeding supermassive black hole. As a result, a cone-shaped bubble of hot gas is bursting out of the center of the galaxy at nearly 1,000 kilometers a second. This image combines Hubble Space Telescope visible-light data (*red and green*) and Chandra X-ray Observatory data (*blue*).

Odd Couple

Black holes have a bad reputation. In many ways, it is deserved.

They are the most efficient engines of destruction known to humanity. Their intense gravity is a one-way ticket to oblivion for anything that strays too close; inside them is undiscovered country from whose bourn no traveler returns. We see them only because the victims do not go quietly to their doom. Material spiraling into a black hole can heat up to millions of degrees and glow brightly. Some of its kinetic energy and momentum may be transferred to a jet of particles flowing outward at close to the speed of light. Black holes of varying sizes take the rap for fusillades of radiation and plasma that astronomers observe all over the cosmos.

COPYRIGHT 2003 SCIENTIFIC AMERICAN, INC.

Yet black holes are not all-powerful. Even those found at the centers of many galaxies, supermassive black holes—whose very name connotes a voracious monster that rules its galactic roost—are minuscule by cosmic standards. They typically account for less than a percent of their galaxy's mass, and their gravity is highly concentrated. Accordingly, astronomers long assumed that supermassive holes, let alone their smaller cousins, would have little effect beyond their immediate neighborhoods. Star formation farther out in the galaxy was thought to march to the beat of a different drummer.

So it has come as a surprise over the past decade that black hole activity and star formation are closely intertwined. In many galaxies where black holes devour material greedily—generating a phenomenon that astronomers call an active galactic nucleus (AGN)—stars form at a precipitous rate in episodes known as starbursts. How can these two seemingly disconnected processes be so intimately related?

Today the AGN-starburst connection is a revolutionary area of research. Beautiful Hubble Space Telescope images are allowing astronomers to pick apart the complex events at the hearts of galaxies, the Chandra X-ray Observatory is peering into places hidden to Hubble, and theorists are trying to make sense of it all. This research bears on some of the most basic questions in astronomy: How did the dark early universe come to light up with billions of stars? Did supermassive black holes need a helping hand to grow to be so big? Could they be agents of creation as well as destruction?

Galaxies on Steroids

BOTH ACTIVE GALACTIC nuclei and starbursts are among the most spectacular phenomena in the universe. An AGN is a luminous and compact source of light at the center of a galaxy. Quasars are the most extreme example. Pumping out as much power as a billion to a trillion suns, AGNs can outshine the rest of their host galaxies. The supermassive black holes that are thought to power them pack a million to a billion times the sun's mass inside a region smaller than a thousand times the sun's diameter. Like a falling rock, material spiraling toward the hole picks up speed and releases energy as it collides with other material. In so doing, it gives off radiation at all wavelengths: radio, infrared, optical, ultraviolet, x-ray, gamma-ray.

Starburst galaxies rival the brilliance of AGNs. They are places where gas condenses into stars at a rate equivalent to producing up to 1,000 suns a year—1,000 times faster than stars currently form in our own galaxy. Some starbursts are confined to comparatively small regions, only hundreds of light-years across, located near the center of a galaxy; others occur on much larger scales, sometimes tens of thousands of light-years across. Starbursts often take place in galaxies that are going through, or have recently undergone, a close encounter or merger with a neighboring galaxy. The tidal forces between the two galaxies disrupt gas and cause it to fall inward, greatly accelerating the normal process by which interstellar clouds collapse and form stars. A starburst typically lasts about 10 million years before running out of gas (literally).

Like AGNs, starburst galaxies shine at a wide range of wavelengths. Much of their power output is simply the light of the stars that have been formed. Starbursts tend to be especially bright sources of infrared radiation, which is produced when interstellar dust absorbs and reradiates starlight. Starbursts also produce a lot of x-rays, which pour forth from massive stars, especially as they die. A massive star goes out with a bang: a supernova explosion, which generates x-rays directly, scatters hot x-ray-emitting debris, and leaves behind a neutron star or a smallish black hole, capable of cannibalizing a companion star and spewing x-rays. The surrounding interstellar gas, heated by all the stellar activity, gives off x-rays, too.

The idea that AGNs are somehow linked to starbursts was not sparked by a single earthshaking discovery but has evolved slowly. It goes back to a time when astronomers were still debating what powered AGNs. Although today nearly all attribute AGNs to supermassive black holes, the situation was not so clear as recently as 15 years ago. Researchers including Roberto Terlevich of the University of Cambridge and Jorge Melnick of the European Southern Observatory argued that AGNs were a type of starburst. To the telescopes of the day, a tight knot of young stars and supernova debris would look just like a supermassive black hole.

Overview/*AGNs and Starbursts*

- The two most powerful phenomena in galaxies are active galactic nuclei (AGNs) and starbursts. The former are intense, concentrated sources of light—probably matter falling into a supermassive black hole. (Quasars are the best-known example.) Starbursts are galactic fireworks shows during which stars form at a frenetic pace.
- Astronomers used to think that AGNs and starbursts, which are often separated by vast distances, had nothing to do with each other. But they have found that the two phenomena tend to occur hand in hand.
- Does an AGN cause the starburst? Or vice versa? Or are they both caused by some underlying process? The answer will be crucial to understanding the evolution of galaxies.

The Case for a Connection

THE NOTION FELL FROM FAVOR only in the late 1980s, as higher-resolution telescopes operating at multiple wavelengths began to reveal just how compact AGNs are: at most a few light-years across and probably a matter of light-minutes across, far too small to encompass a starburst. Even if an entire cluster of stars could fit into such a small space, the stars would rapidly merge together and collapse into a black hole anyway. In addition, AGNs tend to be accompanied by fast-moving jets of material—as a black hole, but not a starburst, would naturally produce [see "Black Holes in Galactic Centers," by Martin J. Rees; SCIENTIFIC AMERICAN, November 1990].

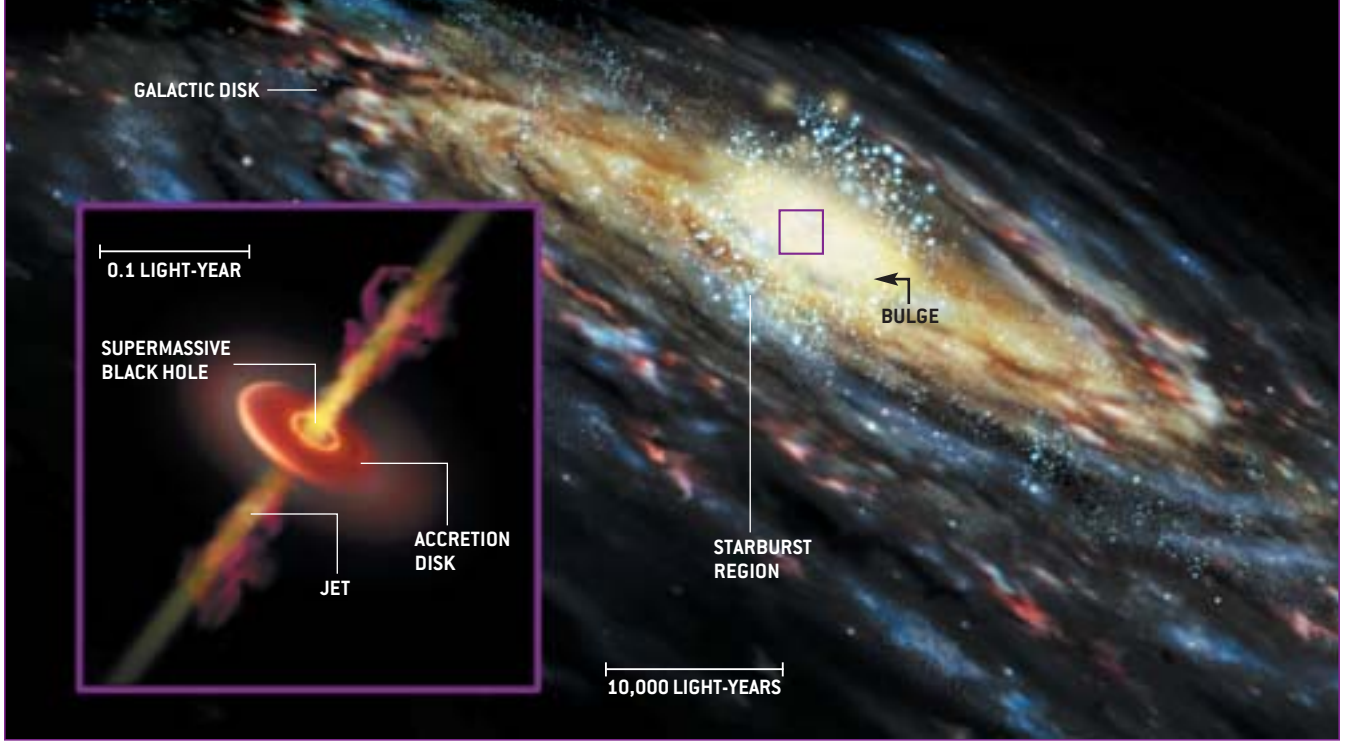
Although AGNs and starbursts proved to be distinct phenomena, these discussions primed astronomers to accept that they might be related in some way [see "Colossal Galactic Ex-

NASA/CXC/GERALD CECIL *University of North Carolina (preceding pages)*

COPYRIGHT 2003 SCIENTIFIC AMERICAN, INC.

ANATOMY OF A GALAXY

A TYPICAL SPIRAL GALAXY contains 100 billion stars, most in a flattened disk. Toward the center is a bulge of stars, and at the very center is usually a supermassive black hole. If the hole is actively feeding, infalling matter forms an accretion disk or is shot back out as a jet. If the galaxy is undergoing a starburst, loose gas turns into stars at a high rate. For years, astronomers thought that the hole and the starburst were unrelated. They were wrong.



plosions," by Sylvain Veilleux, Gerald Cecil and Jonathan Bland-Hawthorn; Scientific American, February 1996]. Several pillars of observational evidence now point to just such a relation. The findings come in a bewildering variety, suggesting that the connection has had a pervasive effect on the universe.

The first piece of evidence is the most direct. Telescopes have seen AGNs alongside starbursts in nearby galaxies. These observations have been tricky to make because galactic cores are filled with gas and dust, obstructing our view. This is where x-ray astronomy comes in. X-rays can penetrate dense gas. Even though current x-ray telescopes lack the resolution of Hubble, they often produce clearer pictures of the dusty centers of galaxies.

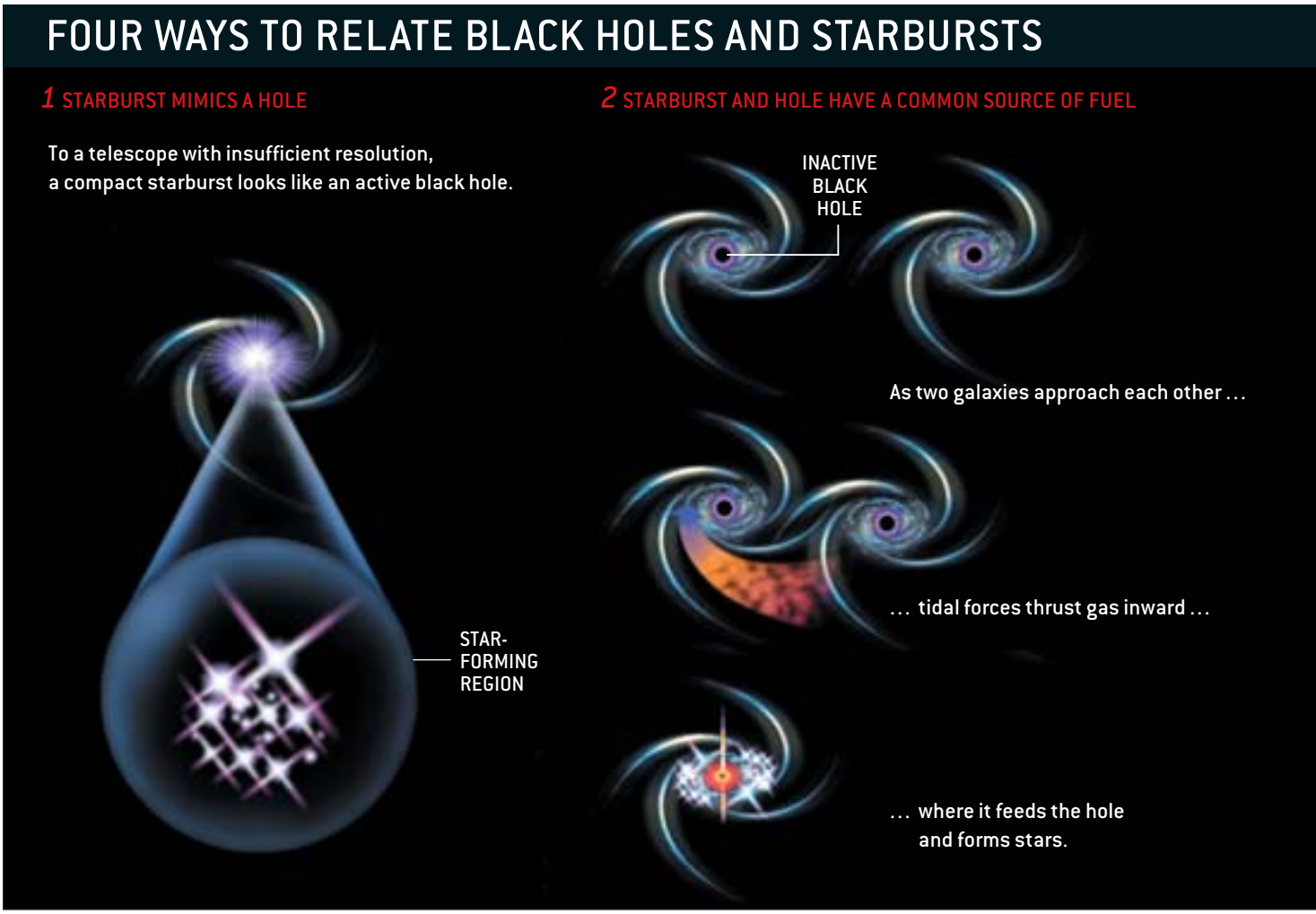
A second line of evidence comes from a recent survey of nearly 23,000 AGNs by Timothy Heckman of Johns Hopkins University and his colleagues. Rather than scrutinize images of all those galaxies, the researchers inferred the presence of AGNs or starbursts from the strength of particular spectral lines, taking highly ionized oxygen as a sign of an AGN and strong hydrogen absorption as indicative of a starburst. The main conclusion was that galaxies with powerful AGNs had many more young stars than did similar galaxies without AGNs. The more powerful the AGN, the more likely it was that the galaxy had experienced a major starburst not long ago. In short, this study verified that the AGN-starburst connection is not merely anecdotal.

Third, AGN galaxies are not the only ones to be blessed with supermassive black holes. Astronomers have detected them at the centers of inactive galaxies as well. It seems that giant holes are everywhere. Most of the time, they lie dormant and invisible; they produce AGNs only when material falls into them at a large and sustained rate. John Kormendy of the University of Texas at Austin, Douglas O. Richstone of the University of Michigan at Ann Arbor and others have demonstrated a correlation between the mass of these holes and the total mass of stars in the galactic centers: the black hole mass is about 0.1 percent of the stellar mass. The same correlation applies to most (though not all) AGN galaxies. Some process, therefore, has linked central black holes to star formation. Lingering discrepancies show that researchers do not fully understand the link.

An AGN-starburst connection might even lurk a mere 24,000 light-years away—at the core of our own galaxy. Rapid motions of stars and gas around the galaxy's center betray the presence of a concentrated mass equal to that of 2.6 million suns. The radio and x-ray emission from this location indicates that the mass is a supermassive black hole—not a truly active hole but one that does feed occasionally. Some have hypothesized that it operates like a mini AGN, slurping up surrounding material at one ten-millionth the rate of a true AGN. Although it is not currently accompanied by a starburst, bright clusters

DON DIXON

COPYRIGHT 2003 SCIENTIFIC AMERICAN, INC.



of stars do reside nearby. They could be left over from a burst of star formation several million years ago.

Two other forms of evidence come from looking back in time. Observers have noticed that AGNs and star formation were even more closely related when the universe was a tenth of its current age. Back then, two types of galaxies were more common: ultraluminous infrared galaxies (called ULIRGs) and radio galaxies, which appear to be galaxies either in an early stage of formation or in the process of a galaxy merger. Their cores contain huge amounts—billions of solar masses—of cold, dense gas. And they host both AGNs and intense starbursts. The other historical approach concerns distant and luminous AGNs—specifically, quasars. They frequently live in messy galaxies, whose distorted shapes and unusual colors suggest that they are merging and forming stars at a high rate.

A final line of evidence derives from the x-ray background radiation, a lesser-known cousin of the cosmic microwave background radiation. Studies of the background have unveiled a population of AGNs hidden from optical telescopes. This obscuration has a natural explanation: the AGNs were accompanied by starbursts, which choked the galaxies with dust [see "The Cosmic Reality Check," by Günther Hasinger and Roberto Gilli; SCIENTIFIC AMERICAN, March 2002].

Chicken or Egg?

THE AGN-STARBURST CONNECTION could have come about in four broad ways: the starburst and AGN are one and the same; some third process caused both the AGN and the starburst; the AGN caused the starburst; or the starburst caused the AGN.

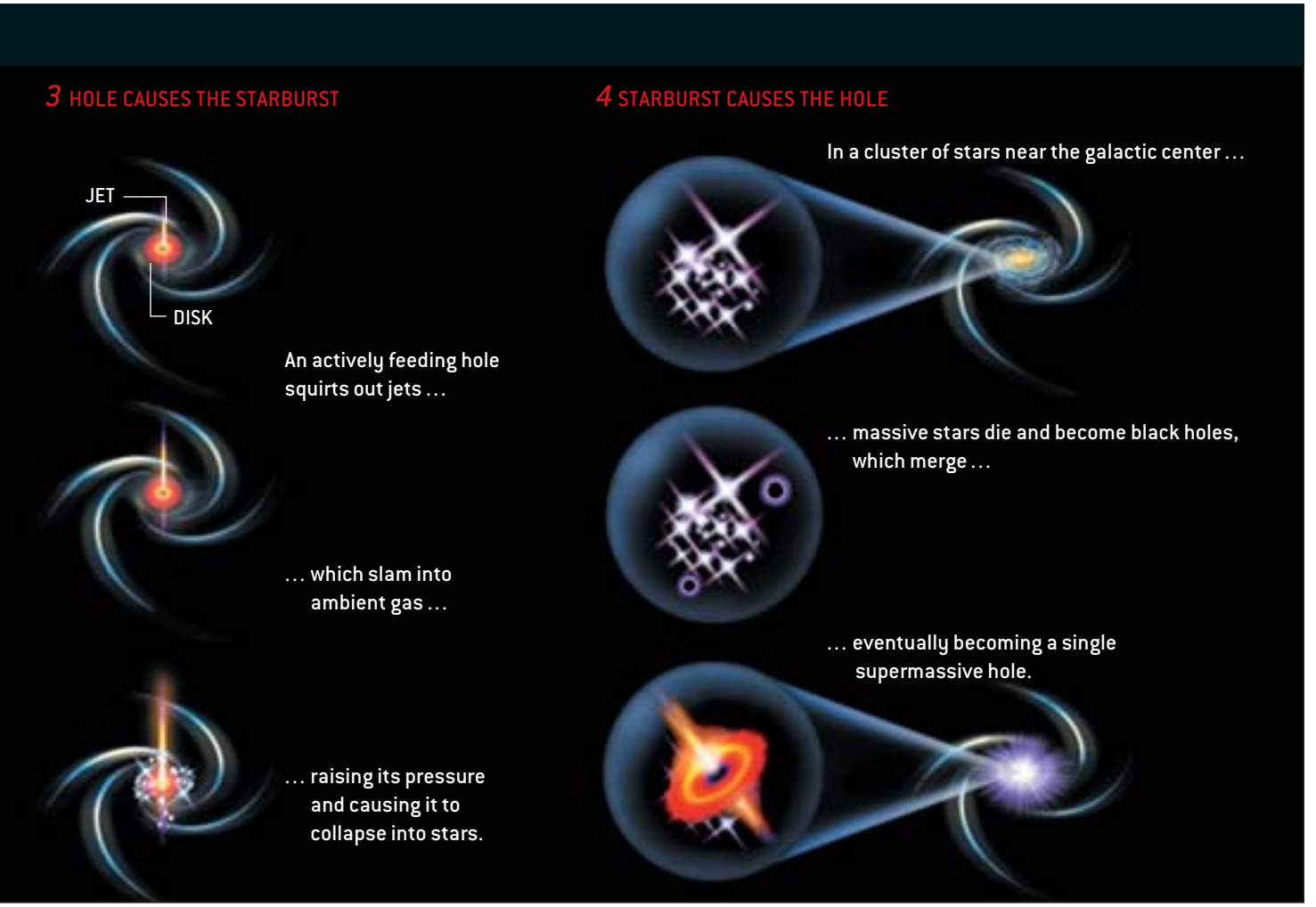
The first scenario is a limited version of the older idea that AGNs are just a type of starburst. Although that idea proved to be wrong for most AGNs, it might work for some of them. Weak AGNs could conceivably be produced by extreme stellar activity rather than a supermassive hole. The activity would occur in such a small region that telescopes might mistake it for a hole. The jury is still out on this possibility.

The second scenario is that the "connection" is merely coincidence. The same processes could set the stage for both starbursts and AGNs. For instance, a galaxy merger could shove gas toward the center of the newly formed entity, inducing a starburst and, by providing fuel for a hole, triggering an AGN. Interestingly, theory predicts that the time it takes for a black hole to grow to supermassive proportions (about 10 million years) is similar to the typical lifetime of a starburst, which is also similar to the time it takes for two galaxies to merge together.

Most researchers, however, have gravitated to the remaining two scenarios, in which AGNs and starbursts are causally

DON DIXON

COPYRIGHT 2003 SCIENTIFIC AMERICAN, INC.



related. The third scenario posits that an existing supermassive black hole, contrary to expectation, exerts a strong influence on its host galaxy. Perhaps the hole pulls material toward the galactic center, enabling star formation. Françoise Combes of the Astronomical Observatory of Paris has championed this model. She argues that once a hole is in place, gas naturally flows into the galaxy core, fueling an AGN. As gas collects, it serves as the raw material for a starburst. The theory is quite plausible: many nearby galaxies that host AGNs also contain dusty structures within their cores, which could be material drawn in from outside. On the other hand, not all these structures have the theoretically predicted shape.

Instead of resulting from an inflow of material into the hole, a starburst might be set off by an outflow of energy from the hole. When the supermassive black hole starts to devour material and produce an AGN, shock waves and jets may rip through the galaxy. Gas piles up along shock fronts and condenses into stars. Chandra observations of the Centaurus A galaxy, where the star formation rate is extremely high, suggest that a massive AGN outburst occurred about 10 million years ago. In the outskirts of the galaxy lies a ring of x-ray emission about 25,000 light-years across, which may have resulted from the shock waves of this explosion. The explosion coincided with an episode of star formation, and the x-ray ring overlaps with arcs of young stars.

The black-hole-comes-first scenario has interesting implications. Black holes, rather than stars, may have been the first beacons in the utter blackness of the early universe. Moreover, some astronomers have suggested that the sun was born during a starburst. If this event was triggered by an AGN in the Milky Way, we may owe our existence to a black hole.

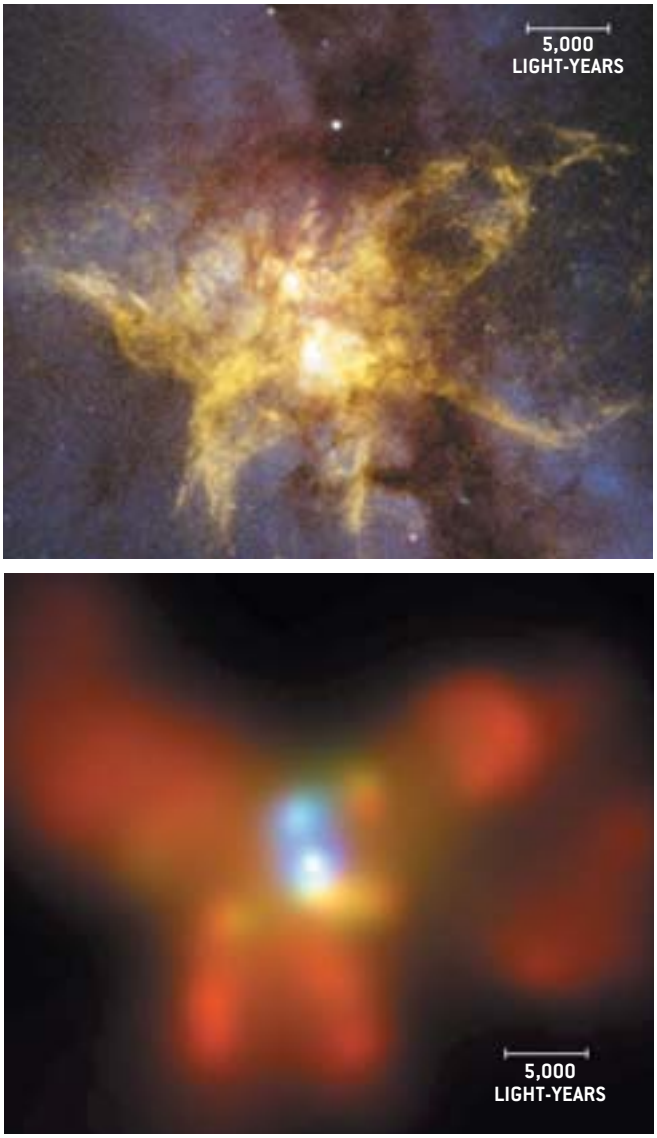
Digging a Hole

THE STARBURST-COMES-FIRST scenario, though, has the most theoretical and empirical support. The connection can result naturally from normal stellar evolution. A starburst creates dense clusters of stars, within which stellar collisions are common [see "When Stars Collide," by Michael Shara; SCIENTIFIC

THE AUTHOR

KIMBERLY WEAVER is an astrophysicist at the Laboratory for High Energy Astrophysics at the NASA Goddard Space Flight Center and an adjunct professor at Johns Hopkins University. She specializes in x-ray studies of black holes, active galactic nuclei and starburst galaxies. In 1996 Weaver received a Presidential Early Career Award for Scientists and Engineers. She also loves the arts; her hobbies include singing, dancing and acting in community theater.

COPYRIGHT 2003 SCIENTIFIC AMERICAN, INC.



DOUBLE TROUBLE: As its strange butterfly shape suggests (*top*), NGC 6240 is not one galaxy but a pair of galaxies that recently merged. The system appears to have not one but two supermassive black holes, which show up as distinct sources of x-rays (*blue circles on bottom image*). Diffuse x-ray-emitting gas (*red*) is a sign of rapid star formation. NGC 6240 is a classic example of how holes, starbursts and galaxy mergers occur together.

NASA/ROELAND P. VAN DER MAREL AND JORIS GERSSEN *Space Telescope Science Institute (optical)*, NASA/CXC/STEFANIE KOMASSA *Max Planck Institute for Extraterrestrial Physics (x-ray)*

American, November 2002]. Massive stars in the cluster quickly die and become neutron stars or stellar-mass black holes, and these bodies agglomerate together. Over tens of millions to hundreds of millions of years, they build up a more massive black hole.

Alternatively, a large black hole could arise from lightweight stars similar to our sun, which do not normally turn into holes. In a dense cluster, these stars could undergo a runaway process of mergers, in which the stars collide and form massive stars, which then join further into megastars a few hundred to a few thousand times as heavy as our sun. Those megastars then collapse to form black holes of similar mass. This process would also take about 100 million years—much less than the lifetime of a galaxy and fast enough to account for the earliest quasars.

No matter how they are created, the black holes would tend to sink into the center of the galaxy. Several could merge to form a supermassive one. This idea has been bolstered by observations of the galaxy NGC 6240, in which a pair of supermassive black holes are circling each other, destined to merge [*see illustration at left*]. Supermassive black holes can continue to grow by feasting on surrounding material. Even star clusters that form in distant reaches of a galaxy can contribute mass to the central hole. Those clusters slowly lose kinetic energy and angular momentum because of friction on a galactic scale, caused by dynamical and gravitational interactions with the rest of the galaxy. They spiral inward and eventually get torn apart by tidal forces. Over the course of billions of years, this process could inject into the central black hole a mass equivalent to tens of millions of suns. Disturbances of the galaxy disk, such as an interaction or merger, could likewise pour fuel into the black hole.

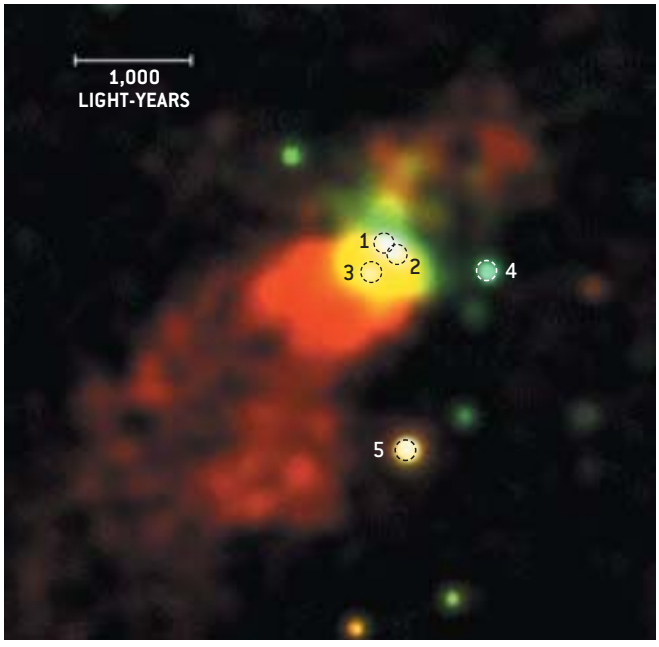
Middleweights

THE STARBURST-COMES-FIRST model predicts an entirely new population of black holes, intermediate between stellar-mass black holes and supermassive ones. Over the past 10 years, circumstantial evidence for these midsize holes has emerged in the form of so-called ultraluminous x-ray sources. Found in several nearby galaxies, these sources emit 10 times to several hundred times as much x-ray power as neutron stars or stellar-mass black holes [see "Hole in the Middle," by George Musser; News Scan, Scientific American, April 2001]. They might be neutron stars whose light is beamed in our direction, making them appear abnormally powerful. But evidence is accumulating that they are in fact black holes with a mass of up to several hundred times the mass of the sun.

Last year two teams of astronomers, led respectively by Roeland P. van der Marel of the Space Telescope Science Institute in Baltimore and Michael Rich of the University of California at Los Angeles, found hints of intermediate-mass holes at the centers of two dense star clusters, M15 and M31-G1. Stars in these clusters are moving so quickly that it would take bodies of 2,000 and 20,000 solar masses, respectively, to confine them. The "bodies" do not have to be large black holes—they could be a batch of neutron stars or small black holes. But even if that is the case, those objects should eventually merge and produce a large black hole.

Tod Strohmayer and Richard Mushotzky of the NASA Goddard Space Flight Center recently discovered that one of the ultraluminous sources, located near the center of the starburst galaxy M82, flickers with a period of about 18 seconds. The flickering is too slow and irregular to come from the surface of a neutron star and too intense to come from material in orbit around such a star. If it comes instead from material in orbit around a black hole, the hole could have a mass of several thousand suns. In the spiral galaxy NGC 1313, Jon Miller of the Harvard-Smithsonian Center for Astrophysics and his colleagues found two ultraluminous x-ray sources that are cooler than stellar-mass black holes. Theory predicts that the temperatures near black holes decrease as their mass increases, so the holes in NGC 1313 must be more massive than stellar-mass holes.

COPYRIGHT 2003 SCIENTIFIC AMERICAN, INC.



SMOKING GUN? The central region of galaxy NGC 253 (*left*) suggests that starbursts can build up supermassive black holes. Five x-ray sources (*circles on right image*) are brighter than stellar-mass black holes but dimmer than supermassive ones. They could be medium-size black holes, an intermediate step in the process of creating big holes from mergers of dead stars. Fuzz in the x-ray image is gas associated with star formation.

These candidate middleweight holes are not located at the centers of their host galaxies, so their relevance to the AGN-starburst connection is not firmly established. But my studies of one nearby starburst galaxy, NGC 253, have provided some direct clues. Before 1995, astronomers believed that the energetic x-rays from this galaxy originated in the hot gas associated with the starburst. In that year, I found hints of black holes in the x-ray spectrum. It was not until 2001, however, that my colleagues and I obtained an x-ray image of this galaxy with Chandra [*see illustration above*].

We found five ultraluminous x-ray sources within the inner 3,000 light-years of NGC 253. One of them, located exactly at the center of the galaxy, is about 100 times as bright as a neutron star or stellar-mass hole, suggesting that it has a mass equivalent to about 100 suns. It could be a black hole caught in the act of developing into a full-fledged AGN. The sequence of events might go as follows: A starburst takes place near the center of the galaxy. The massive stars thus formed collapse and merge to form lightweight black holes, which then spiral to the galactic center and merge, forming the seed for a supermassive hole. As the starburst winds down, the supermassive hole starts to power an AGN.

Studying how starburst activity affects the fueling and growth of a supermassive hole should offer insight into the birth of the most powerful of all AGNs, quasars. Astronomers have wondered why quasars in the early universe were much more powerful than present-day AGNs. The reason may be simply that the early universe had more frequent episodes of star formation, which triggered more intense AGNs.

To be sure, the situation may be more complicated than a straightforward triggering of one type of activity by the other. Galaxies could cycle between an AGN phase and a starburst phase. When the cycles overlapped, astronomers would see both phenomena together. AGNs and starbursts may even evolve in unison. Current observations are not able to tell whether the AGN comes first, the starburst comes first, or they both occur together. This fascinating question should be answered with the next generation of telescopes.

Observations with the Space Infrared Telescope Facility, which NASA plans to launch this year, will illuminate the AGN-starburst connection in the earliest galaxies. Scientists will be able to compare infrared, visible-light and x-ray data to determine whether AGNs or starbursts dominate activity during galaxy formation, which could determine which came first. It is also important to find more nearby galaxies like NGC 253.

The AGN-starburst connection is perhaps the ultimate intergenerational link in the universe. Black holes represent the coalesced embers of bygone stars; starbursts represent the birth of vibrant young stars. It may have taken a partnership of the old and the new to shape galaxies, including ours. SA

MORE TO EXPLORE

Relationships between Active Galactic Nuclei and Starburst Galaxies. Edited by Alexei V. Filippenko. ASP Conference Series, Vol. 31; 1992.

An Introduction to Active Galactic Nuclei. Bradley M. Peterson. Cambridge University Press, 1997.

Starburst Galaxies: Near and Far. Edited by Linda Tacconi and Dieter Lutz. Springer Verlag, 2001.

X-ray Properties of the Central kpc of AGN and Starbursts: The Latest News from Chandra. Kimberly A. Weaver in *The Central Kiloparsec of Starbursts and AGN: The La Palma Connection.* Edited by J. H. Knapen, J. E. Beckman, I. Shlosman and T. J. Mahoney. ASP Conference Series, Vol. 249; 2001. Available at **arXiv.org/abs/astro-ph/0108481**

Starburst-AGN Connection from High Redshift to the Present Day. Yoshiaki Taniguchi in *Proceedings of the IAU 8th Asian-Pacific Regional Meeting,* Vol. 1. Edited by Satoru Ikeuchi, John Hearnshaw and Tomoyuki Hanawa. ASP Conference Series, Vol. 289; 2003. Available at **arXiv.org/abs/astro-ph/0211161**

NASA/SAO/CXC (*x-ray*); EUROPEAN SOUTHERN OBSERVATORY (*optical*)

COPYRIGHT 2003 SCIENTIFIC AMERICAN, INC.

COUNTING

COPYRIGHT 2003 SCIENTIFIC AMERICAN, INC.

THE Last Fish

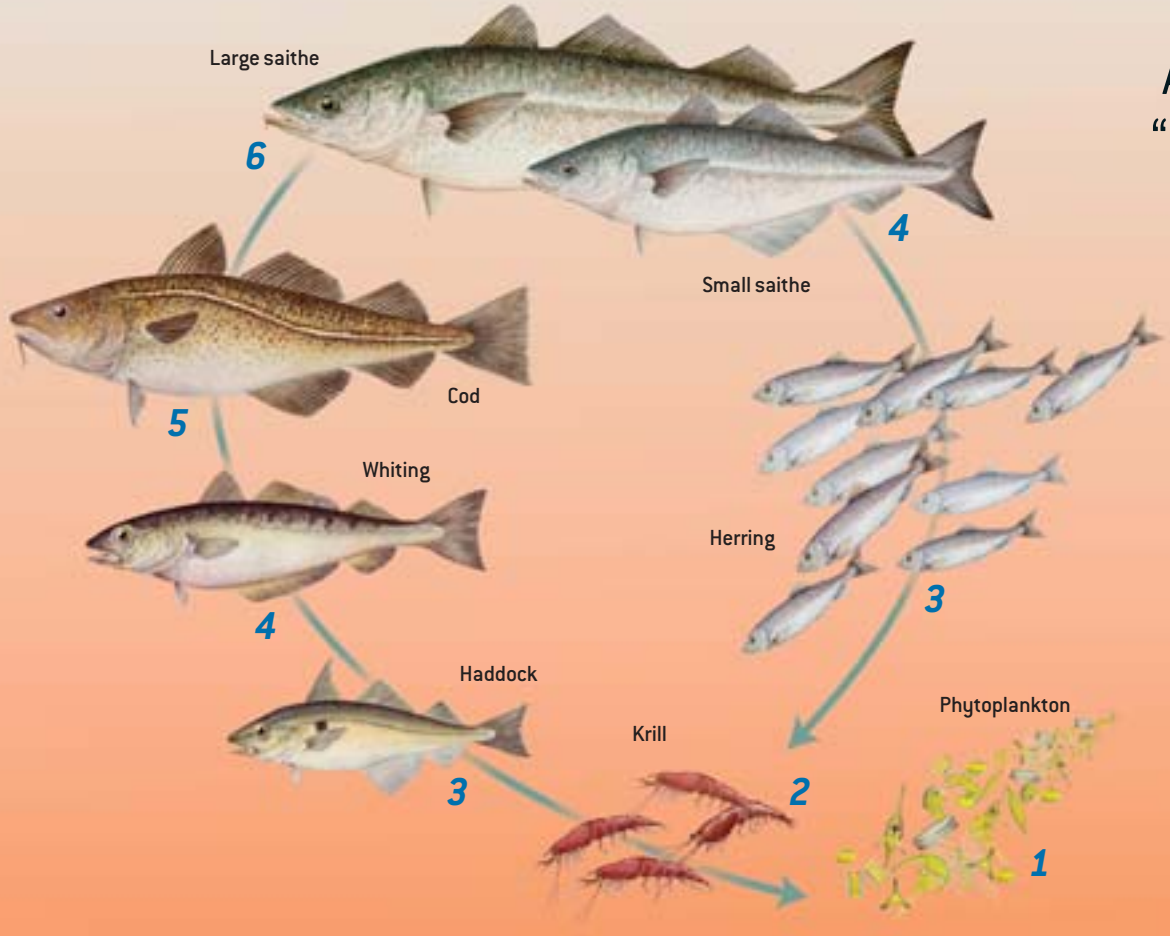
JEAN GAUMY *Magnum*

OVERFISHING HAS SLASHED STOCKS—ESPECIALLY OF LARGE PREDATOR SPECIES—TO AN ALL-TIME LOW WORLDWIDE, ACCORDING TO NEW DATA. IF WE DON'T MANAGE THIS RESOURCE, WE WILL BE LEFT WITH A DIET OF JELLYFISH AND PLANKTON STEW

By Daniel Pauly and Reg Watson

Georges Bank—the patch of relatively shallow ocean just off the coast of Nova Scotia, Canada—used to teem with fish. Writings from the 17th century record that boats were often surrounded by huge schools of cod, salmon, striped bass and sturgeon. Today it is a very different story. Trawlers trailing dredges the size of football fields have literally scraped the bottom clean, harvesting an entire ecosystem—including supporting substrates such as sponges—along with the catch of the day. Farther up the water column, longlines and drift nets are snagging the last sharks, swordfish and tuna. The hauls of these commercially desirable species

COPYRIGHT 2003 SCIENTIFIC AMERICAN, INC.



An Example of "Fishing Down"

FOOD WEBS contain fewer steps, or trophic levels, when overfishing occurs. After fishers have taken the largest members of a slow-growing predatory species—such as saithe—they must turn to smaller individuals that have not yet achieved full size. Unlike older saithe, these younger fish are not large enough to catch cod, which normally consume whiting, which in turn usually eat krill-grazing haddock (*left*). Instead the small saithe must eat even smaller fish, such as herring, which feed directly on krill (*right*). Wiping out larger saithe therefore shortens the food web to four levels instead of six, disrupting ecosystems. Note that actual trophic levels rarely reach six because large fish eat a variety of other fish.

are dwindling, and the sizes of individual fish being taken are getting smaller; a large number are even captured before they have time to mature. The phenomenon is not restricted to the North Atlantic but is occurring across the globe.

Many people are under the mistaken impression that pollution is responsible for declines in marine species. Others may find it hard to believe that a shortage of desirable food fish even exists, because they still notice piles of Chilean sea bass and tuna fillets in their local fish markets. Why is commercial fishing seen as having little if any effect on the species that are being fished? We suspect that this perception persists from another age, when fishing was a matter of wresting sustenance from a hostile sea using tiny boats and simple gear.

Our recent studies demonstrate that we can no longer think of the sea as a bounteous provider whose mysterious depths contain an inexhaustible resource. Over the past several years we have gathered and analyzed data on the world's fisheries, compiling the first comprehensive look at the state of the marine food resource. We have found that some countries, particularly China, have overreported their catches, obscuring a downward trend in fish caught worldwide. In general, fishers must work farther offshore and at greater depths in an effort to keep up with the catches of yesteryear and to try to meet the burgeoning demand for fish. We contend that overfishing and the fishing of these distant stocks are unsustainable practices and are causing the depletion of important species. But it is not too late to implement policies to protect the world's fisheries for future generations.

Overview/*Fish Declines*

- New analyses show that fisheries worldwide are in danger of collapsing from overfishing, yet many people still view the ocean as a limitless resource whose bounty humanity has just begun to tap.
- Overfishing results from booms in human populations, increases in the demand for fish as a nutritious food, improvements in commercial fishing technology, and global and national policies that fail to encourage the sustainable management of fisheries.
- Solutions to the problem include banning fishing gear such as dredges that damage ecosystems; establishing marine reserves to allow fisheries to recover; and abolishing government subsidies that keep too many boats on the seas chasing too few fish.

The Law of the Sea

EXPLAINING HOW THE SEA got into its current state requires relating a bit of history. The ocean used to be a free-for-all, with fleets flying the flags of various countries competing for fish thousands of miles from home. In 1982 the United Nations adopted the Convention on the Law of the Sea, which allows countries bordering the ocean to claim exclusive economic zones reaching 200 nautical miles into open waters. These areas include the highly productive continental shelves of roughly 200 meters in depth where most fish live out their lives.

The convention ended decades—and, in some instances, even

CLEO VILETT

COPYRIGHT 2003 SCIENTIFIC AMERICAN, INC.

centuries—of fighting over coastal fishing grounds, but it placed the responsibility for managing marine fisheries squarely on maritime countries. Unfortunately, we cannot point to any example of a nation that has stepped up to its duties in this regard.

The U.S. and Canadian governments have subsidized the growth of domestic fishing fleets to supplant those of now excluded foreign countries. Canada, for instance, built new offshore fleets to replace those of foreign nations pushed out by the convention, effectively substituting foreign boats with even larger fleets of more modern vessels that fish year-round on the same stocks that the domestic, inshore fleet was already targeting. In an effort to ensure that there is no opportunity for foreign fleets to fish the excess allotment—as provided for in the convention—these nations have also begun to fish more extensively than they would have otherwise. And some states, such as those in West Africa, have been pressured by others to accept agreements that allow foreign fleets to fish their waters, as sanctioned by the convention. The end result has been more fishing than ever, because foreign fleets have no incentive to preserve local marine resources long-term—and, in fact, are subsidized by their own countries to garner as much fish as they can.

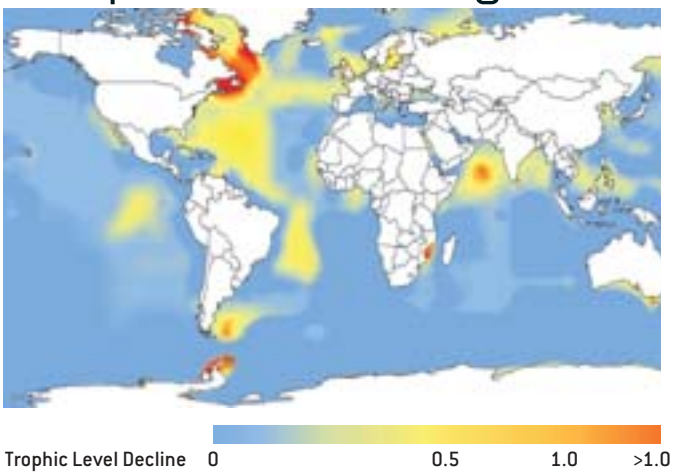
The expansion made possible by the Convention on the Law of the Sea and technological improvements in commercial fishing gear (such as acoustic fish finders) temporarily boosted fish catches. But by the late 1980s the upward trend began to reverse, despite overreporting by China, which, in order to meet politically driven "productivity increases," was stating that it was taking nearly twice the amount of fish that it actually was.

In 2001 we presented a statistical model that allowed us to examine where catches differed significantly from those taken from similarly productive waters at the same depths and latitudes elsewhere in the world. The figures from Chinese waters—about 1 percent of the world's oceans—were much higher than predicted, accounting for more than 40 percent of the deviations from the statistical model. When we readjusted the worldwide fisheries data for China's misrepresentations, we concluded that world fish landings have been declining slowly since the late 1980s, by about 700,000 metric tons a year. China's overreporting skewed global fisheries statistics so significantly because of the country's large size and the degree of its overreporting. Other nations also submit inaccurate fisheries statistics—with a few overreporting their catches and most underreporting them—but those numbers tend to cancel one another out.

Nations gather statistics on fish landings in a variety of ways, including surveys, censuses and logbooks. In some countries, such as China, these data are forwarded to regional offices and on up through the government hierarchy until they arrive at the national offices. At each step, officials may manipulate the statistics to meet mandatory production targets. Other countries have systems for cross-checking the fish landings against import/export data and information on local consumption.

The most persuasive evidence, in our opinion, that fishing is wreaking havoc on marine ecosystems is the phenomenon that one of us (Pauly) has dubbed "fishing down the food web." This describes what occurs when fishers deplete large predator fish at the top of the food chain, such as tuna and swordfish, until they become rare, and then begin to target smaller species that would usually be eaten by the large fish [*see illustration on opposite page*].

Hot Spots of Overfishing



OVERFISHING caused the complexity of the food chains in important fisheries to drop by more than one trophic level between the years 1950 and 2000. The open ocean usually has few fish.

Fishing Down

THE POSITION A PARTICULAR ANIMAL occupies in the strata of a food web is determined by its size, the anatomy of its mouthparts and its feeding preferences. The various layers of the food web, called trophic levels, are ranked according to how many steps they are removed from the primary producers at the base of the web, which generally consists of phytoplanktonic algae. These microscopic organisms are assigned a trophic level (TL) of 1.

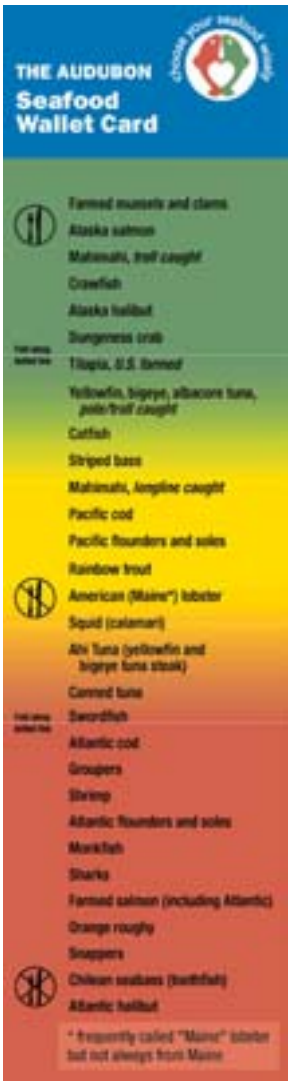
Phytoplankton are grazed mostly by small zooplankton—mainly tiny crustaceans of between 0.5 and two millimeters in size, both of which thus have a TL of 2. (This size hierarchy stands in stark contrast to terrestrial food chains, in which herbivores are often very large; consider moose or elephants, for instance.) TL 3 consists of small fishes between 20 and 50 cen-

THE AUTHORS

DANIEL PAULY and *REG WATSON* are fisheries researchers at the Sea Around Us Project in Vancouver, where Pauly is the principal investigator and Watson is a senior scientist. The project, which was initiated and funded by the Pew Charitable Trusts, is based at the Fisheries Center at the University of British Columbia and is devoted to studying the impact of fishing on marine ecosystems. Pauly's early career centered on formulating new approaches for fisheries research and management in tropical developing countries. He has designed software programs for evaluating fish stocks and initiated FishBase, the online encyclopedia of fishes of the world. Watson's interests include fisheries modeling, data visualization and computer mapping. His current research focuses on mapping the effects of global fisheries, modeling underwater visual census techniques and using computer simulations to optimize fisheries.

NINA FINKEL

COPYRIGHT 2003 SCIENTIFIC AMERICAN, INC.



POPULAR FISH—including many of the fillets and steaks that can be found in piles at fish markets (*above*)—have been decimated by overfishing. Fishers must use increasingly complex technology and fish farther offshore and at greater depths to catch such fish. The National Audubon Society and other organizations have issued wallet cards (*right*) so that consumers can avoid overfished species (*red*) or those whose status is cause for concern (*yellow*). The entire card can be downloaded at www.audubon.org/campaign/lo/seafood/cards.html

timeters in length, such as sardines, herring and anchovies. These small pelagic fishes live in open waters and usually consume a variable mix of phytoplankton and both herbivorous and carnivorous zooplankton. They are caught in enormous quantities by fisheries: 41 million metric tons were landed in 2000, a number that corresponds to 49 percent of the reported global marine fish catch. Most are either destined for human consumption, such as canned sardines, or reduced to fish meal and oil to serve as feed for chickens, pigs and farmed salmon or other carnivorous fish.

The typical table fish—the cod, snapper, tuna and halibut that restaurants serve whole or as steaks or fillets—are predators of the small pelagics and other small fishes and invertebrates; they tend to have a TL of between 3.5 and 4.5. (Their TLs are not whole numbers because they can consume prey on several trophic levels.)

The increased popularity in the U.S. of such fish as nutritious foods has undoubtedly contributed to the decline in their stocks. We suggest that the health and sustainability of fisheries can be assessed by monitoring the trends of average TLs. When those numbers begin to drop, it indicates that fishers are relying on ever smaller fish and that stocks of the larger predatory fish are beginning to collapse.

In 1998 we presented the first evidence that "fishing down" was already occurring in some fishing grounds, particularly in the North Atlantic, off the Patagonian coast of South America and nearby Antarctica, in the Arabian Sea, and around parts of Africa and Australia. These areas experienced TL declines of 1 or greater between 1950 and 2000, according to our calculations [*see map on preceding page*]. Off the west coast of Newfoundland, for instance, the average TL went from a maximum of 3.65 in 1957 to 2.6 in 2000. Average sizes of fish landed in those regions dropped by one meter during that period.

Our conclusions are based on an analysis of the global database of marine fish landings that is created and maintained by the U.N. Food and Agriculture Organization, which is in turn derived from data provided by member countries. Because this data set has problems—such as overreporting and the lumping of various species into a category called "mixed"—we had to incorporate information on the global distribution of fishes from FishBase, the online encyclopedia of fishes pioneered by Pauly, as well as information on the fishing patterns and access rights of countries reporting catches.

Research by some other groups—notably those led by Jeremy B. C. Jackson of the Scripps Institution of Oceanography in San Diego and Ransom A. Myers of Dalhousie University in Halifax—suggests that our results, dire as they might seem, in fact underestimate the seriousness of the effects that marine fisheries have on their underlying resources. Jackson and his colleagues have shown that massive declines in populations of marine mammals, turtles and large fishes occurred along all coastlines where people lived long before the post–World War II period we examined. The extent of these depletions was not recognized until recently because biologists did not consult historians or collaborate with archaeologists, who study evidence of fish consumption in middens (ancient trash dumps).

JASON ESKENAZI (*photograph*); LIVING OCEANS PROGRAM, AUDUBON (*wallet card*)

COPYRIGHT 2003 SCIENTIFIC AMERICAN, INC.

Myers and his co-workers used data from a wide range of fisheries throughout the world to demonstrate that industrial fleets generally take only a few decades to reduce the biomass of a previously unfished stock by a factor of 10. Because it often takes much longer for a regulatory regime to be established to manage a marine resource, the sustainability levels set are most likely to be based on numbers that already reflect population declines. Myers's group documents this process particularly well for the Japanese longline fishery, which in 1952 burst out of the small area around Japan—to which it was confined until the end of the Korean War—and expanded across the Pacific and into the Atlantic and Indian oceans. The expansion decimated tuna populations worldwide. Indeed, Myers and his colleague Boris Worm recently reported that the world's oceans have lost 90 percent of large predatory fish.

Changing the Future

WHAT CAN BE DONE? Many believe that fish farming will relieve the pressure on stocks, but it can do so only if the farmed organisms do not consume fish meal. (Mussels, clams and tilapia, an herbivorous fish, can be farmed without fish meal.) When fish are fed fish meal, as in the case of salmon and various carnivores, farming makes the problem worse, turning small pelagics—including fish that are otherwise perfectly fit for human consumption, such as herring, sardines, anchovies and mackerels—into animal fodder. In fact, salmon farms consume more fish than they produce: it can take three pounds of fish meal to yield one pound of salmon.

One approach to resolving the difficulties now besetting the world's fisheries is ecosystem-based management, which would seek to maintain—or, where necessary, reestablish—the structure and function of the ecosystems within which fisheries are embedded. This would involve considering the food requirements of key species in ecosystems (notably those of marine mammals), phasing out fishing gear that destroys the sea bottom, and implementing marine reserves, or "no-take zones," to mitigate the effects of fishing. Such strategies are compatible with the set of reforms that have been proposed for years by various fisheries scientists and economists: radically reducing global fleet capacity; abolishing government subsidies that keep otherwise unprofitable fishing fleets afloat; and strictly enforcing restrictions on gear that harm habitats or that capture "bycatch," species that will ultimately be thrown away.

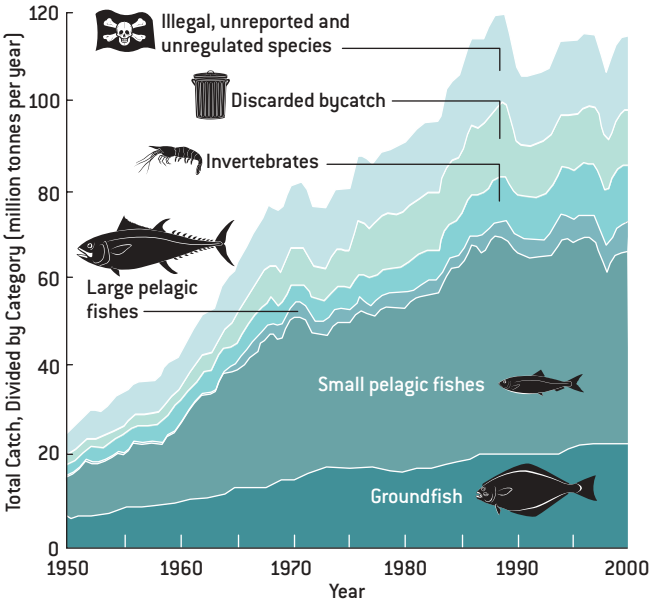
Creating no-take zones will be key to preserving the world's fisheries. Some refuges should be close to shore, to protect coastal species; others must be large and offshore, to shield oceanic fishes. No-take zones now exist, but they are small and scattered. Indeed, the total area protected from any form of fishing constitutes a mere 0.01 percent of the ocean surface. Reserves are now viewed by fishers—and even by governments—as necessary concessions to conservationist pressure, but they must become management tools for protecting exploited species from overfishing.

A major goal should be to conserve species that once maintained themselves at deeper depths and farther offshore, before fishers developed improved gear for going after them. This type of fishing is similar to a nonrenewable mining operation because fishes are very vulnerable, typically long-lived, and have very low productivity in the dark, cold depths. These measures would enable fisheries, for the first time, to become sustainable. SA

NATIONAL GEOGRAPHIC CHANNEL
always wonder

A segment based on this article will air June 26 on *National Geographic Today,* a program on the National Geographic Channel. Please check your local listings.

Catching More Fish



AMOUNT OF FISH LANDED has more than quintupled over the past 50 years. As the world's population has grown, commercial fishing technology has advanced, and demand for fish in some countries has surged.

NINA FINKEL (*graph*) AND CLEO VILLETT (*illustrations*)

MORE TO EXPLORE

Effect of Aquaculture on World Fish Supplies. Rosamond L. Naylor, Rebecca J. Goldburg, Jurgenne H. Primavera, Nils Kautsky, Malcolm C. M. Beveridge, Jason Clay, Carl Folke, Jane Lubchenco, Harold Mooney and Max Troell in *Nature,* Vol. 405, pages 1017–1024; June 29, 2000.

Historical Overfishing and the Recent Collapse of Coastal Ecosystems. Jeremy B. C. Jackson et al. in *Science,* Vol. 293, pages 629–638; July 27, 2001.

Systematic Distortion in World Fisheries Catch Trends. Reg Watson and Daniel Pauly in *Nature,* Vol. 414, pages 534–536; November 29, 2001.

In a Perfect Ocean: The State of Fisheries and Ecosystems in the North Atlantic Ocean. Daniel Pauly and Jay Maclean. Island Press, 2003.

Rapid Worldwide Depletion of Predatory Fish Communities. Ransom A. Myers and Boris Worm in *Nature,* Vol. 423, pages 280–283; May 15, 2003.

More information on the state of world fisheries can be found on the Web sites of the Sea Around Us Project at **www.saup.fisheries.ubc.ca** and of FishBase at **www.fishbase.org**

COPYRIGHT 2003 SCIENTIFIC AMERICAN, INC.

ADAPTIVE ANTENNA ARRAY focuses radio waves on a "personal cell" surrounding each mobile user. This smart-antenna technology can increase the range of wireless voice and data networks and allow several users in the same coverage area to communicate on the same frequency.

COPYRIGHT 2003 SCIENTIFIC AMERICAN, INC.

By Martin Cooper

ANTENNAS GET SMART

Adaptive antenna arrays can vastly improve wireless communications by connecting mobile users with virtual wires

Each of us is immersed in a sea of radio-frequency waves. The invisible electromagnetic energy comes from many sources: broadcast towers, cellular-phone networks and police radio transmissions, among others. Although this radiation may be harmless to our bodies, it can severely inhibit our ability to receive and transmit information. Excess radio energy is a kind of pollution, because it can disrupt useful communications. As the intensity of radio-frequency interference in our environment grows, we have to raise the volume of radio signals so that they can be heard over the electromagnetic background noise. And as our electronic communications become more intense, they simply add to the din of radio interference.

One solution to this problem lies in a new class of radio antennas that could dramatically reduce man-made interference. Instead of wastefully broadcasting personal communications—such as cell-phone calls—in all directions, these innovative antennas track the positions of mobile users and deliver radio signals directly to them. These antenna systems also maximize the reception of an individual cell-phone user's signal while minimizing the interference from other users. In effect, the antennas create a virtual wire extending to each mobile phone.

These systems are generically referred to as smart antennas, but the smartest members of the class are called adaptive antenna arrays. In 1992 I co-founded ArrayComm, a San Jose, Calif., company focused on developing adaptive arrays that can be incorporated into both new and existing wireless networks. Each of our arrays consists of up to a dozen antennas and a powerful digital processor that can combine and manipulate the incoming and outgoing signals. The technology, which is also being pursued by Lucent Technologies, Nortel Networks and other firms, promises to decrease the cost and improve the quality of wireless communications. Adaptive antenna arrays

KENN BROWN

COPYRIGHT 2003 SCIENTIFIC AMERICAN, INC.

are already providing these benefits to millions of cell-phone users. Moreover, these smart antennas may become the linchpins of the wireless Internet because they are ideally suited to transmitting and receiving large amounts of data.

The Physics of Antennas

TO UNDERSTAND HOW smart antennas operate, it helps to know how ordinary, "dumb" antennas work. A radio antenna converts electric currents and voltages created by a transmitter into electromagnetic waves that radiate into space. Antennas also intercept such waves and convert them back into currents and voltages that can be processed by a receiver. The simplest and most common radio antennas, called dipoles, are merely rods of very specific lengths that radiate energy in all directions [*see top illustration on opposite page*]. Radio waves get weaker as they spread through space and are absorbed by obstacles such as air, trees and buildings.

Commercial radio and television stations need to reach geographically dispersed audiences, so it is logical for them to broadcast signals in all directions. A cell-phone call, though, is usually aimed at just one user. In a cellular network, users communicate with the nearest base station, a set of antennas that handle all the wireless service's signals in the surrounding area (called the cell). The base stations are located so that the entire coverage area can be divided into cells; when a user moves from one cell to another, the system automatically hands off the call to the appropriate base station. In this situation, it would be far preferable to focus the radio energy on each user, much as the reflector in a flashlight focuses light into a beam. A radio beam would extend much farther than a signal of equivalent power that is broadcast in all directions. And because the radio beams transmitted by the cellular base station to different users would be spatially separated, interference would be reduced.

Reflectors can focus radio waves into beams, but they are cumbersome and costly. So engineers have developed tricks to create radio beams without reflectors. If we stand two antennas side by side, with the distance between them equal to one half the wavelength of the radio signal, the radiated energy from this simple array assumes the pattern of a figure eight when viewed from above [*see middle illustration on opposite page*]. The radio waves travel farthest in the two directions perpendicular to the array (that is, perpendicular to the line connecting the antennas), because in these directions the user would receive both antenna signals at the exact same time (in other words, the two signals would be in phase). When two identical signals are in phase, they combine to form a signal that is twice as strong as either one alone. But in the directions parallel to the array, the user would receive the two antenna signals 180 degrees out of phase. The wave peaks from one antenna would arrive at the same time as the wave troughs from the other, so the two signals would cancel each other out. This phenomenon creates a null, an area where the signal cannot be detected.

The beam generated by the two-antenna array is a fairly broad one, and it extends in opposite directions. But engineers can progressively narrow the beam by adding more antennas. Such phased-array antennas have been used to focus radar beams since World War II. Although increasing the number of antennas makes the beam narrower, it also produces smaller beams, called lobes, on both sides of the main beam [*see bottom illustration on opposite page*]. Depending on the user's direction from the antenna array, the signal can be either stronger than the signal radiated by a single antenna ("gain") or weaker because of cancellation effects ("rejection").

Aiming the Beam

RADIO BEAMS ARE OF LITTLE USE, however, if they cannot be pointed at their intended recipients. The most obvious solution is to physically turn the antenna array, but this method is very awkward and expensive. It is much easier to steer the radio beams electronically. Using one technique, called beam switching, antenna arrays create a group of overlapping radio beams that together cover the surrounding area [*see top illustration on page 52*]. When a cell-phone user makes a call, the radio receiver determines which beam is providing the strongest signal from the user. The array transmitter then "talks back" using the same beam from which the signal was received. If the user walks out of the original beam into an adjacent one, the radio's control system switches to the new beam, employing it for both reception and transmission.

Beam switching, though, does not work well in the real world of wireless communications. For a beam to be most effective, the mobile user has to stand in the center of the beam [*see bottom illustration on page 52*]. As the user moves off center, the signal fades, just as the light from a flashlight gets dimmer as you step away from the direction in which it is pointed. When the user approaches the far edge of the beam, the signal strength can degrade rapidly before the system switches to the adjacent beam. And what if there is another user who is trying to use the same radio channel but from a different direction? If the second user is standing in a null, there would be no interference, but if the interloper happens to be in the center of a lobe, the second signal may well block or distort the first.

Another challenge for beam-switching systems is the fact that in most environments, radio signals rarely travel in direct paths. The signal you receive on your cell phone is usually a combination of reflections off natural and man-made objects—buildings, mountains, vehicles, trees and so on. And these re-

Overview/*Smart Antennas*

- In wireless networks, it is often useful to employ an array of antennas at each cellular base station.
- Adaptive arrays include powerful processors that manipulate the antenna signals to enhance communication with one user while minimizing interference from all others.
- The companies involved in smart-antenna development include ArrayComm, Navini Networks, Lucent Technologies and Nortel Networks.

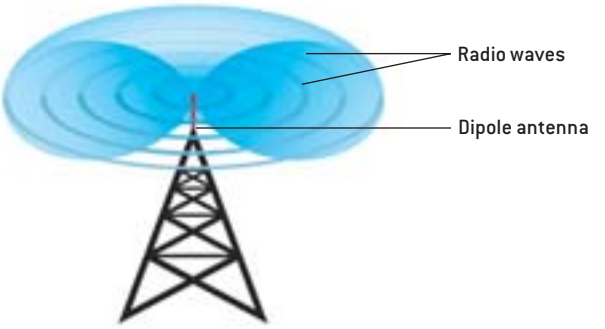
COPYRIGHT 2003 SCIENTIFIC AMERICAN, INC.

GENERATING A RADIO BEAM

SMART-ANTENNA TECHNOLOGY is based on the ability to shape the coverage patterns of radio waves. By focusing the waves on individual cell-phone users, smart antennas can extend the range of cellular systems and minimize interference.

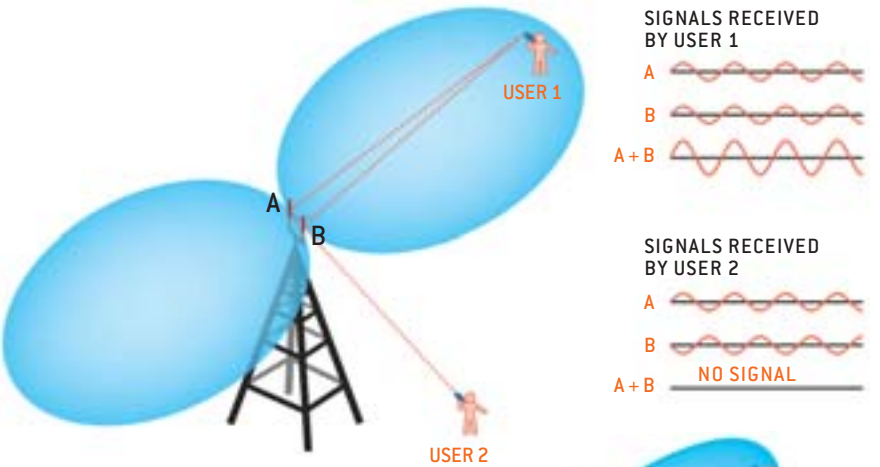
ORDINARY ANTENNA

The simplest antenna is a dipole, a metal rod that radiates energy in all directions. The radio waves get weaker as they propagate through space and are absorbed by obstacles. These antennas are suitable for radio and television broadcasts, which need to reach geographically dispersed audiences.



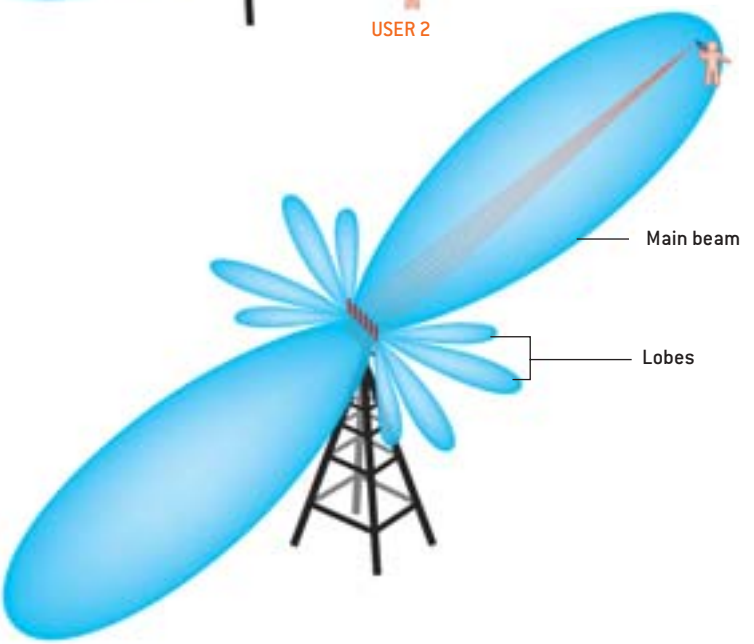
TWO-ANTENNA ARRAY

Now consider antennas A and B placed side by side, with the distance between them equal to half the wavelength of the radio signal that each is emitting. (For cellular systems, wavelengths range from 12 to 37 centimeters.) The cell-phone user standing perpendicular to the array (user 1) receives both signals at the same time, doubling their intensity. But the user aligned with the array (user 2) receives the signal from antenna B half a cycle before receiving the signal from antenna A. The two signals are 180 degrees out of phase, so they cancel each other out. Thus, the two-antenna array creates areas, called nulls, where the signal cannot be heard.



SIX-ANTENNA ARRAY

An array with six antennas generates a radio beam that is narrower and extends farther than the beam produced by the two-antenna array. The six-antenna array also creates smaller beams called lobes.



SARA CHEN

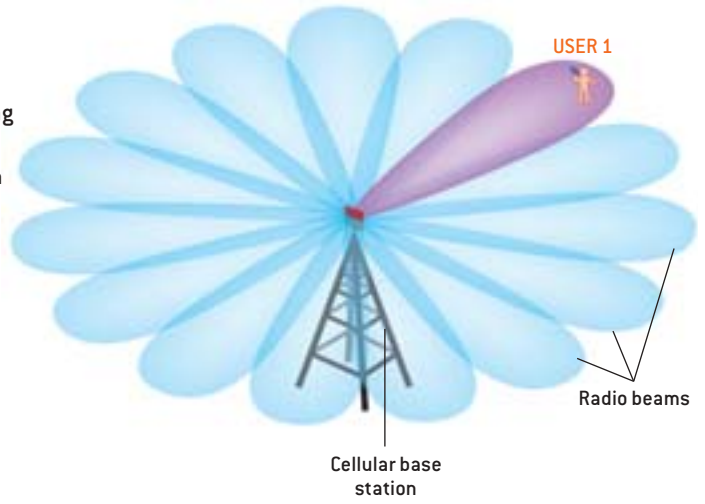
COPYRIGHT 2003 SCIENTIFIC AMERICAN, INC.

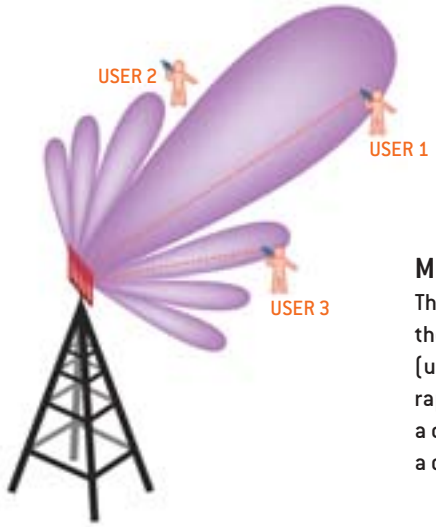
BEAM SWITCHING

A SMART ANTENNA must be able to point its radio beams at their intended recipients. One approach, called beam switching, can be easily incorporated into wireless networks, but the technique has some serious drawbacks.

SELECTING A BEAM

In a beam-switching system, an antenna array generates overlapping radio beams that together cover the surrounding area (*right*). When a cell-phone user makes a call, the radio receiver at the cellular base station determines which beam is best aligned with the user's signal. The array uses the same beam (*purple*) to transmit signals back to the user.





MIXED SIGNALS

The chief problem with beam switching is that the array does not always aim the beam directly at the user (*left*). As the intended recipient of a cell-phone call (user 1) approaches the edge of the beam, the signal strength can degrade rapidly. Other callers can communicate on the same frequency (but using a different beam) if they are in a null (user 2). But signals traveling to and from a cell-phone user in a lobe (user 3) may interfere with user 1's signals.

flections are constantly changing, especially the ones caused by large vehicles such as buses. This phenomenon, known as multipath, also affects the signals sent from the cell phone to the base station. If a user in a beam-switching network is near the edge of a beam, the signal that he or she transmits may bounce into the adjacent beam before reaching the beam-switching array. In that case, the array would transmit to the wrong beam, and the user might miss the return signal entirely.

For practical applications, beam-switching systems are obviously inadequate. An antenna array that is truly smart would be able to aim a radio beam directly at a mobile user instead of choosing a beam that comes relatively close. The ideal array would also be able to adjust the beam pattern so that it minimized interference from other users who were communicating on the same frequency channel. And last, the array must be able to adapt quickly to changes in the user's position and shifting reflections. That is where adaptive antenna arrays come in.

The Cocktail Party Effect

WHAT MAKES THE ADAPTIVE ARRAY so smart? The key step is processing the information received by its antennas. A good analogy is the way the brain processes acoustic information from the ears. A person with normal hearing can usually locate the source of a sound even with his or her eyes closed. The convoluted folds of the outer ear produce differing resonances depending on the angle of the incoming sound. And unless the sound is coming from directly ahead or behind (or directly above or below), it reaches one ear before the other, so there is a time lag between the two signals. The brain receives this information and rapidly computes the location of the source.

What is more, people with normal hearing can pick up relatively quiet sounds—say, an interesting conversation—amid loud background noise. This phenomenon is commonly known as the cocktail party effect. Researchers have shown that the ability to focus on a specific sound partly depends on the ability to locate the sound's source. In an experiment that tested how well people can hear a signal while being blasted with background noise, subjects listening with both ears were able to detect much softer sounds than subjects listening with only one ear. Once the brain has determined the position of the acoustic source, it can focus on the sound and tune out unwanted noise coming from other directions.

Similarly, an adaptive antenna array can pinpoint the source of a radio signal and selectively amplify it while canceling out competing signals. The array's brain is a digital processor that can manipulate the signals coming down the wires from the antennas. A typical adaptive array contains four to 12 antennas, but for simplicity's sake let us consider an array of two antennas, separated by a distance equal to half the wavelength of the radio signal. In an ordinary array the signals from the two an-

SARA CHEN

COPYRIGHT 2003 SCIENTIFIC AMERICAN, INC.

tennas are just added together, but in an adaptive array the signals are sent to the adjoining processor, which can perform any number of mathematical operations on them.

For example, suppose that the array is aligned north to south and a signal from a cell-phone user comes in from the east [*see top illustration on next page*]. The processor can quickly determine the direction of the signal: because the radio waves reach both antennas at the same time, they must be coming from a direction perpendicular to the array. To maximize reception, the processor adds the signals together, doubling their intensity. When transmitting back to the user, the array emits identical signals from both antennas.

But now suppose that another cell-phone user sends a signal from the south [*see middle illustration on next page*]. Because the radio waves hitting the north antenna are 180 degrees out of phase from the waves striking the south antenna, the processor can tell that the signal is coming from a direction parallel to the array. So the processor now *subtracts* one signal from the other—that is, it inverts the signal from the north (or south) antenna, turning wave peaks into wave troughs and vice versa, and adds this mirror image to the signal from the south (or north) antenna. Again, the signal's intensity is doubled. And when the array transmits back to the cell-phone user, the processor sends an out-of-phase signal to one of the antennas, generating a radio beam that runs from north to south.

Notice that in both these examples the radio beam generated for one cell-phone user does not reach the other. The two users could be communicating with the adaptive array at the same time and on the same frequency channel, but their signals would not interfere with each other. The array's processor can create radio beams pointing in other directions as well by performing more complex mathematical operations on the signals from the two antennas. The task of selective transmission and reception is thus reduced to solving a series of simultaneous equations. To direct beams at users who are moving around, the processor must repeatedly solve the equations with constantly updated information from the antenna array.

Adding more antennas to the adaptive array increases the locating precision and the gain of the signal [*see bottom illustration on next page*]. An array with 12 antennas can hear signals a dozen times as faint as those that can be heard by a single antenna. The array can also transmit 12 times as loudly and much more directly. And the processor can juggle the antenna signals to create beam patterns that ensure the greatest possible gain for a desired signal and the greatest possible rejection for other signals on the same frequency.

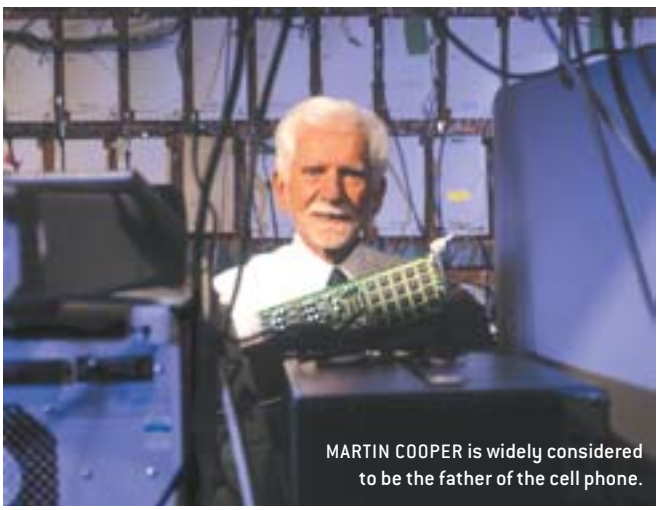
Because the processor is fast enough to perform this task many times a second, the array can continually readjust the radio beam as the cell-phone user walks or drives across the array's coverage area. The system is designed so that stray reflections of the user's signal against vehicles or buildings do not trigger abrupt changes in the direction of the beam. By tracking the user's route, the array can estimate the likelihood of future movements and discard erroneous information indicating sudden leaps in position.

Furthermore, the most advanced adaptive arrays can take advantage of the multipath phenomenon to focus radio signals still further. The processors in these arrays are so powerful that they can handle information from all the reflected signals that bounce along various routes between the cell phone and the adaptive array. By including the multipath reflections in the mathematical equations, the processor can extrapolate not only the direction of the signal but also the exact position of the user's cell phone. In an urban environment where there are nu-

An adaptive array can pinpoint the source of a radio signal and selectively amplify it.

THE AUTHOR

MARTIN COOPER is perhaps most famous for leading the development of the portable cellular phone. On April 3, 1973, Cooper—then a 44-year-old vice president and division manager at Motorola—used a prototype cellular system to call a rival at AT&T Bell Labs from a street corner in Manhattan. Ten years later Motorola introduced the first commercial cell phone, the DynaTAC 8000X, which weighed nearly two pounds and cost $3,995. During 29 years with Motorola, Cooper built and managed both its paging and cellular businesses and served as corporate director of research and development. After leaving Motorola in 1983, Cooper co-founded Cellular Business Systems, which became the dominant firm in the cellular billing industry before the company was sold to Cincinnati Bell. In 1992 he co-founded ArrayComm in San Jose, Calif., now the world leader in smart-antenna technology. Cooper holds bachelor's and master's degrees in electrical engineering from the Illinois Institute of Technology.



MARTIN COOPER is widely considered to be the father of the cell phone.

ANDY FREEBERG

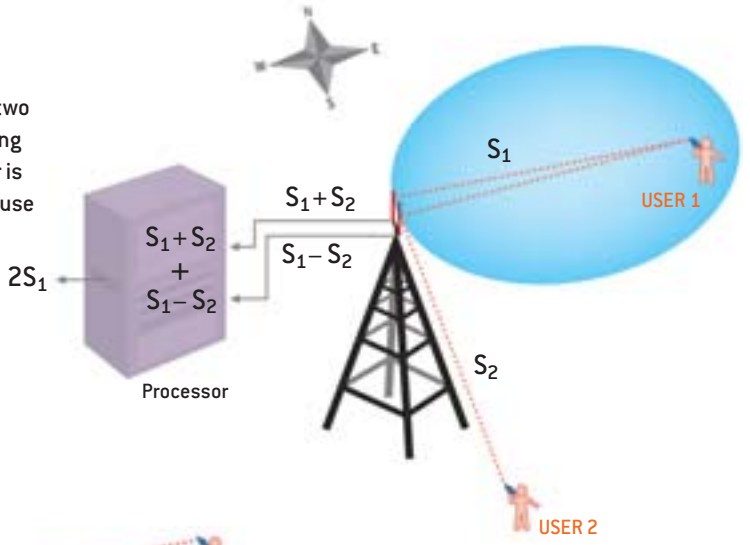
COPYRIGHT 2003 SCIENTIFIC AMERICAN, INC.

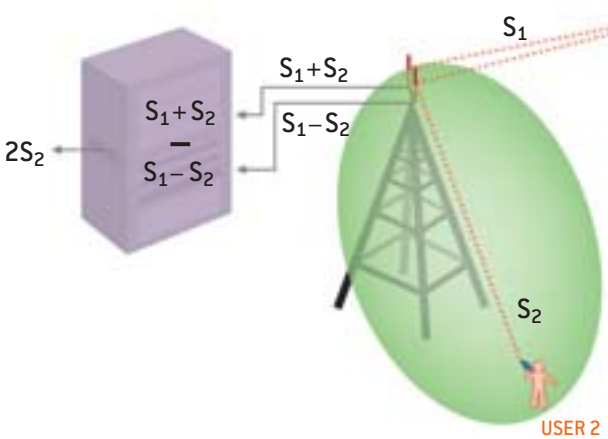
ADAPTIVE ANTENNA ARRAYS

THE SMARTEST ANTENNAS employ digital processors to manipulate incoming and outgoing signals. Called adaptive antenna arrays, these systems enhance reception from and transmission to one user while minimizing interference from others.

POINTING THE BEAM

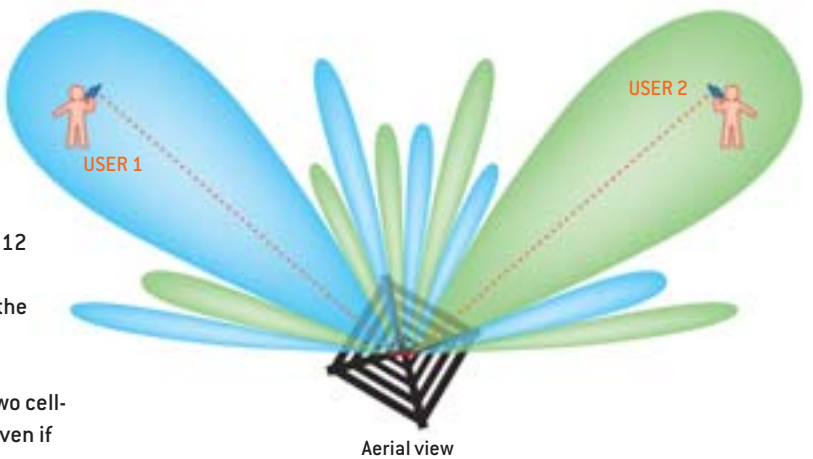
The simplest example of an adaptive array contains two antennas spaced half a wavelength apart. On receiving a call from user 1, the array determines that the user is standing due east (perpendicular to the array), because the signal (S_1) arrives at both antennas at the same time. For user 1, the processor adds the signals from the two antennas to maximize the intensity of S_1 and to eliminate interference from user 2.





CHANGING DIRECTION

On receiving a call from user 2, the processor determines that the user is standing due south (aligned with the array), because the signal arriving at one antenna (S_2) is 180 degrees out of phase from the signal arriving at the other antenna ($-S_2$). For user 2, the processor therefore subtracts one signal from the other, maximizing the intensity of S_2 and eliminating the signal from user 1.



TWO CALLERS AT ONCE

In practice, an adaptive array contains four to 12 antennas. The processor performs complex mathematical operations on the signals from the antennas to ensure that the beam and lobes generated for user 1 (*blue*) do not overlap the pattern created for user 2 (*green*). Thus, the two cell-phone users do not interfere with each other even if they communicate on the same frequency.

SARA CHEN

COPYRIGHT 2003 SCIENTIFIC AMERICAN, INC.

merous reflections, the adaptive array can receive from and transmit to a small area surrounding the phone. Instead of generating a radio beam, the array creates a "personal cell" that can be only centimeters in radius. And because the array is constantly recalculating the phone's position, this personal cell can follow the user as he or she moves about.

Benefits and Applications

WIRELESS NETWORKS THAT EMPLOY adaptive antenna arrays have several advantages over conventional cellular networks. Because a base station equipped with an adaptive array has a far greater range than an ordinary station transmitting at the same power, fewer stations are needed to cover a given area. Although adaptive arrays may be more expensive than traditional antennas, reducing the number of base stations dramatically cuts the overall cost of deploying and operating a wireless network.

Adaptive arrays also enable a cellular service company to make better use of a scarce resource: the spectrum of frequencies allotted to the company for its radio signals. Many cellular systems are becoming overloaded with customers—in certain congested sectors, the barrage of signals sometimes exceeds the amount that can be carried on the limited number of radio channels. Customers feel the crunch when their calls are dropped or they hear poor-quality signals. But because adaptive arrays allow several cell-phone users within a base station's coverage area to share the same radio channel, the technology increases the capacity of the spectrum. The improvement over ordinary antennas is significant: base stations outfitted with adaptive arrays can serve about six times as many people for voice communications and up to 40 times as many for data transmission. The result is better service and less interference, not to mention less wasted energy and radio pollution.

It is not surprising, then, that adaptive antenna arrays are already in commercial use. Arrays using technology created by ArrayComm have been mounted on more than 150,000 cellular base stations in Japan, China, Thailand and other countries in Asia and Africa. All told, the arrays provide phone service to more than 15 million people. Commercial adoption has been slower in the U.S. and Europe, partly because the telecommunications industry's economic slump has curtailed new investment in cellular networks. But one U.S. manufacturer, Airnet in Melbourne, Fla., is currently making cellular base stations that employ ArrayComm's technology. And Marconi, a British telecommunications company, is developing an advanced base station that will contain adaptive arrays.

Adaptive arrays are also a boon to wireless data networks. Because the arrays minimize interference, they can receive and transmit more data to users in a given chunk of frequency spectrum. A base station equipped with an adaptive array could deliver data to as many as 40 concurrent users at a rate of one megabit a second, which is about 20 times as fast as the typical data rate for existing long-range wireless networks. Because all the users in such a network do not usually require peak data rates at the same time, one station with an adaptive array could serve several thousand people. Users with laptops or other portable devices would be able to get uninterrupted high-speed access to the Internet while walking or driving across the coverage area.

Since the late 1990s the telecommunications industry has been heralding the advent of the wireless Internet [see "The Wireless Web: Special Report"; SCIENTIFIC AMERICAN, October 2000]. The new networks have been developing more slowly than originally predicted, but work is nonetheless progressing. As wireless carriers continue to pursue 3G networks—next-generation cellular systems that transmit data in packets—other companies are offering a variety of competing solutions for high-speed data transmission. Smart antennas have been incorporated into some of these solutions and can be put to use in existing networks as well. A data network based on ArrayComm's technology is now operating in Sydney, Australia, and similar networks may soon be built in the U.S. and South Korea. Adaptive arrays developed by Navini Networks in Richardson, Tex., are also being tested by wireless carriers. Several major manufacturers of telecommunications equipment plan to incorporate smart-antenna technology into their next generation of products.

All told, adaptive arrays provide cellular-phone service to more than 15 million people.

For almost 100 years after Alexander Graham Bell invented the telephone, voice communications relied on a physical connection—a copper wire or a coaxial cable—between the caller and the network. Over the past 30 years, though, cellular phones have given us a taste of the freedom to communicate without wires. With the help of adaptive-array technology, wireless carriers will be able to offer far better performance, at a much lower cost, than wired networks do. Only then will we rid ourselves of the copper cage. SA

MORE TO EXPLORE

A Review of the Cocktail Party Effect. Barry Arons in *Journal of the American Voice I/O Society,* Vol. 12; July 1992. Available at **xenia.media.mit.edu/~barons/cocktail.html**

Intelligent Antennas: Spatial Division Multiple Access. Martin Cooper and Marc Goldburg in *1996 Annual Review of Communications,* pages 999–1002. Available at **www.arraycomm.com/Company/spatial_division.pdf**

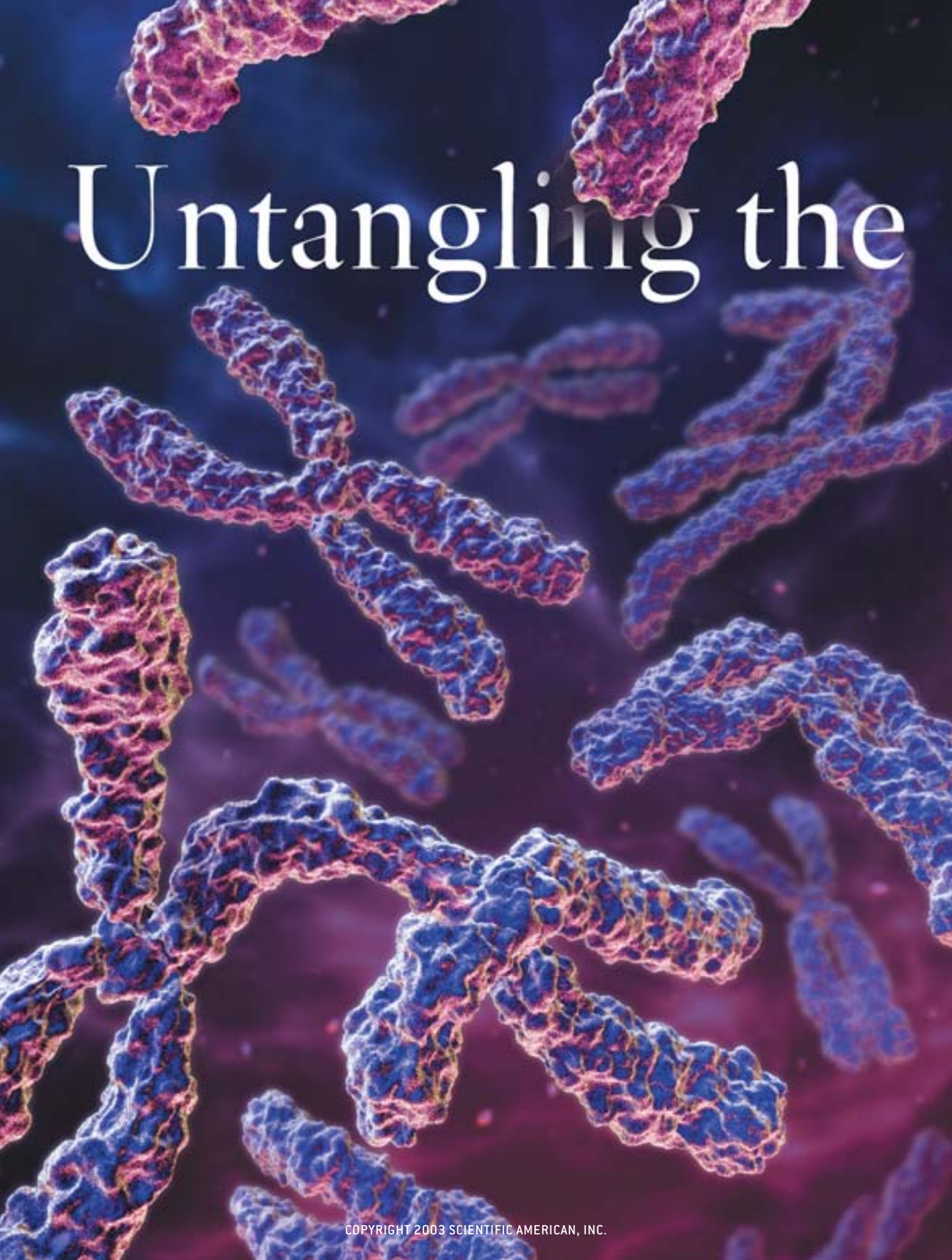
A Layman's Guide to Cellular. Martin Cooper in *1996 Annual Review of Communications,* pages 993–997. Available at **www.arraycomm.com/Company/laymans_guide.pdf**

Smart Antennas for Wireless Communications: IS-95 and Third Generation CDMA Applications. Joseph C. Liberti and Theodore S. Rappaport. Prentice Hall, 1999.

More information about smart antennas can be found at **www.arraycomm.com/Company/white_papers.html**

COPYRIGHT 2003 SCIENTIFIC AMERICAN, INC.

Untangling the



COPYRIGHT 2003 SCIENTIFIC AMERICAN, INC.

Roots of Cancer

RECENT EVIDENCE CHALLENGES LONG-HELD THEORIES OF HOW CELLS TURN MALIGNANT—AND SUGGESTS NEW WAYS TO STOP TUMORS BEFORE THEY SPREAD

By W. Wayt Gibbs

What causes cancer?

Tobacco smoke, most people would say. Probably too much alcohol, sunshine or grilled meat; infection with cervical papillomaviruses; asbestos. All have strong links to cancer, certainly. But they cannot be root causes. Much of the population is exposed to these carcinogens, yet only a tiny minority suffers dangerous tumors as a consequence.

A cause, by definition, leads invariably to its effect. The immediate cause of cancer must be some combination of insults and accidents that induces normal cells in a healthy human body to turn malignant, growing like weeds and sprouting in unnatural places.

At this level, the cause of cancer is not entirely a mystery. In fact, a decade ago many geneticists were confident that science was homing in on a final answer: cancer is the result of cumulative mutations that alter specific locations in a cell's DNA and thus change the particular proteins encoded by cancer-related genes at those spots. The mutations affect two kinds of cancer genes. The first are called tumor suppressors. They normally restrain cells' ability to divide, and mutations permanently disable the genes. The second variety, known as oncogenes, stimulate growth—in other words, cell division. Mutations lock oncogenes into an active state. Some researchers still take it as axiomatic that such growth-promoting changes to a small number of cancer genes are the initial event and root cause of every human cancer.

Others, however, including a few very prominent oncologists, are increasingly challenging that theory. No one questions that cancer is ultimately a disease of the DNA. But as biologists trace tumors to their roots, they have discovered many other abnormalities at work inside the nuclei of cells that, though not yet cancerous, are headed that way. Whole chromosomes, each containing 1,000 or more genes, are often lost or duplicated in their entirety. Pieces of chromosomes are frequently scrambled, truncated or fused together. Chemical additions to the DNA, or to the histone proteins around which it coils, somehow silence important genes, but in a reversible process quite different from mutation.

The accumulating evidence has spawned at least three hypotheses that compete with the standard dogma to explain what changes come first and which aberrations matter most in the decade-long transformation of a cell and its descendants from well-behaved tissue to invasive tumor. The challengers dispute the dominant view of the disease as the product of a defined genetic state. They argue that it is more useful to think of cancer as the consequence of a chaotic process, a combination of Murphy's Law and Darwin's Law: anything that can go wrong will, and in a competitive environment, the best adapted survive and prosper.

Despite that shared underlying principle, the new theories make different predictions about what kind of treatments will work best.

JEFF JOHNSON *Hybrid Medical Animation*

CAREFULLY CHOREOGRAPHED dance of chromosomes occurs during cell division. Missteps that mangle chromosomes or that send the wrong number to each daughter cell may be critical events early in the development of cancer, according to new theories.

COPYRIGHT 2003 SCIENTIFIC AMERICAN, INC.

Some suggest that many cancers could be prevented altogether by better screening, changes in diet, and new drugs—or even by old drugs, such as aspirin. Other theories cast doubt on that hope.

Marks of Malignancy

A WORKABLE THEORY of cancer has to explain both why it is predominantly a disease of old age and why we do not all die from it. A 70-year-old is roughly 100 times as likely to be diagnosed with a malignancy as a 19-year-old is. Yet most people make it to old age without getting cancer.

Biologists estimate that more than 10 million billion cells must cooperate to keep a human being healthy over the course of an 80-year life span. If any one of those myriad cells could give rise to a tumor, why is it that fewer than half the population will ever contract a cancer serious enough to catch a doctor's attention?

One explanation is that a cell must acquire several extraordinary skills to be malignant. "Five or six different regulatory systems must be perturbed in order for a normal cell to grow as a cancer," asserts Robert A. Weinberg of the Whitehead Institute at the Massachusetts Institute of Technology. In a review paper last November, he and William C. Hahn of the Dana-Farber Cancer Institute in Boston argued that all life-threatening cancers manifest at least six special abilities. (Although Weinberg is one of the founding proponents of the standard paradigm, even those who challenge that theory tend to agree with this view.)

For example, cancer cells continue dividing in situations in which normal cells would quietly wait for a special chemical signal—say, from an injured neighbor. Somehow they counterfeit these pro-growth messages. Conversely, tumor cells must ignore "stop dividing" commands that are sent out by the adjacent tissues they squeeze and by their own internal aging mechanisms.

All cancerous cells have serious problems of some sort with their DNA, and as they double again and again, many cells in the resulting colony end up far from the blood vessels that supply oxygen and nutrients. Such stresses trigger autodestruct mechanisms in healthy cells. Tumor cells find some way to avoid this kind of suicide. Then they have to persuade nearby blood vessels to build the infrastructure they need to thrive.

A fifth superpower that almost all cancers acquire is immortality. A culture of normal human cells stops dividing after 50 to 70 generations. That is more than enough doublings to sustain a person through even a century of healthy life. But the great majority of cells in tumors quickly die of their genetic defects, so those that survive must reproduce indefinitely if the tumor is to grow. The survivors do so in part by manipulating their telomeres, gene-free complexes of DNA and protein that protect the ends of each chromosome.

Tumors that develop these five faculties are trouble, but they are probably not deadly. It is the sixth property, the ability to invade nearby tissue and then metastasize to distant parts of the body, that gives cancer its lethal character. Local invasions can usually be removed surgically. But nine of every 10 deaths from the disease are the result of metastases.

Only an elite few cells in a tumor seem to acquire this ability to detach from the initial mass, float through the circulation and start a new colony in a different organ from the one that gave birth to them. Unfortunately, by the time they are discovered, many cancers have already metastasized—including, in the U.S., 72 percent of lung cancers, 57 percent of colorectal, and 34 percent of breast. By then the prognosis is frequently grim.

> **"If you look at most solid tumors in adults, it looks like someone SET OFF A BOMB in the nucleus."**
>
> *—William C. Hahn, Dana-Farber Cancer Institute*

Overview/*How Cancer Arises*

- Cancer is a genetic disease. Alterations to the DNA inside cells can endow cells with morbid "superpowers," such as the ability to grow anywhere and to continue dividing indefinitely.
- Most cancer researchers have long focused on mutations to a relatively small set of cancer-related genes as the decisive events in the transformation of healthy cells to malignant tumors.
- Recently, however, other theories have emerged to challenge this view. One hypothesizes that a breakdown in DNA duplication or repair leads to many thousands of random mutations in cells. Another suggests that damage to a few "master" genes mangles the chromosomes, which then become dangerous. A third challenger proposes that abnormal numbers of chromosomes in a cell may be the first milestone on the road to cancer.

The Order of Disorder

DOCTORS COULD CATCH incipient tumors sooner if scientists could trace the steps that cells take down the road to cancer after the initial assault to their DNA by a carcinogen or some random biochemical mishap. Researchers broadly agree on the traits of the diseased cells that emerge from the journey. It is the propelling force and the order of each milestone that are under active debate.

The dominant paradigm for 25 years has been that tumors grow in spurts of mutation and expansion. Genetic dam-

COPYRIGHT 2003 SCIENTIFIC AMERICAN, INC.

SIX DIABOLICAL SUPERPOWERS OF CANCER

1. GROWTH EVEN IN THE ABSENCE OF NORMAL "GO" SIGNALS

Most normal cells wait for an external message before dividing. Cancer cells (*image*) often counterfeit their own pro-growth messages.

2. GROWTH DESPITE "STOP" COMMANDS ISSUED BY NEIGHBORING CELLS

As the tumor (*yellow*) expands, it squeezes adjacent tissue, which sends out chemical messages that would normally bring cell division to a halt. Malignant cells ignore the commands.

3. EVASION OF BUILT-IN AUTODESTRUCT MECHANISMS

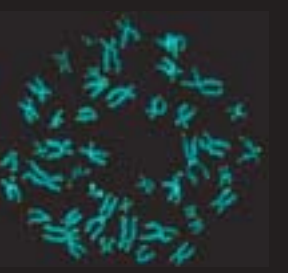
In healthy cells, genetic damage above a critical level usually activates a suicide program. Cancerous cells (*magenta*) bypass this mechanism, although agents of the immune system (*orange*) can sometimes successfully order the cancer cells to self-destruct.

4. ABILITY TO STIMULATE BLOOD VESSEL CONSTRUCTION

Tumors need oxygen and nutrients to survive. They obtain them by co-opting nearby blood vessels to form new branches (*brown streaks*) that run throughout the growing mass.

5. EFFECTIVE IMMORTALITY

Healthy cells can divide no more than 70 times. Malignant cells need more than that to make tumors. So they work around systems—such as the telomeres (*yellow*) at the end of chromosomes (*blue*)—that enforce the reproductive limit.



6. POWER TO INVADE OTHER TISSUES AND SPREAD TO OTHER ORGANS

Cancers usually become life-threatening only after they somehow disable the cellular circuitry that confines them to a specific part of the particular organ in which they arose. New growths (*orange* and *yellow*) appear and eventually interfere with vital systems.

Clockwise from top right: CHRIS JONES *Corbis;* PETER LANSDORP *University of British Columbia;* SCIENCE PHOTO LIBRARY; FRANK LYNCH *QualTek Molecular Laboratories;* ANDREJS LIEPINS/SPL; CNRI/SPL; SPL

age to a cell deletes or disrupts a tumor suppressor gene—*RB*, *p53* and *APC* are among the best known—thereby suppressing proteins that normally ensure the integrity of the genome and the process of cell division. Alternatively, a mutation may increase the activity of an oncogene—such as *BRAF*, *c-fos* or *c-erbb3*—whose proteins then stimulate the cell to reproduce.

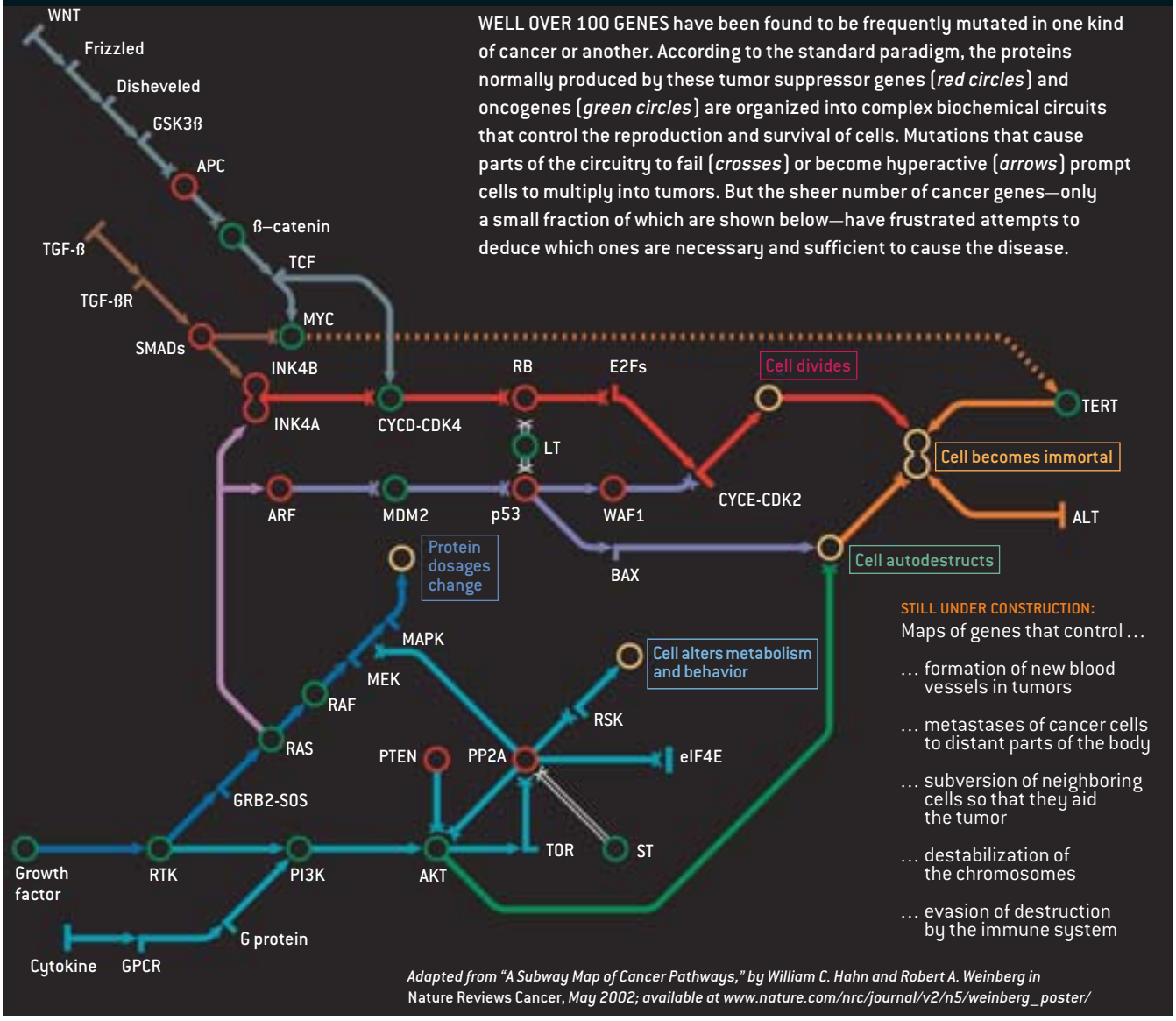
Changes to cancer genes endow the cell with one or more superpowers, allowing it to outbreed its neighbors. The cell passes abnormalities in its DNA sequence on to its descendants, which become a kind of clone army that grows to the limits of its capacity. Eventually another random mutation to a cancer gene knocks down another obstacle, initiating another burst of growth.

Cells normally have two copies of every chromosome—one from the mother, the other from the father—and thus two copies, or alleles, of every gene. (In males, the single X and Y chromosomes are notable exceptions.) A mutation to just one allele is enough to activate an oncogene permanently. But it takes two hits to knock out both alleles of a tumor suppressor gene. Four to 10 mutations in the right genes can transform any cell. Or so the theory goes.

The mutant-gene paradigm gained almost universal acceptance because it explained very well what scientists saw in their experiments on genetically engineered mice and human cell cultures. But new technologies now allow researchers to study the genomes of cancerous and

COPYRIGHT 2003 SCIENTIFIC AMERICAN, INC.

MALIGNANT MUTATIONS: A PARTIAL MAP



Adapted from "A Subway Map of Cancer Pathways," by William C. Hahn and Robert A. Weinberg in Nature Reviews Cancer, *May 2002; available at www.nature.com/nrc/journal/v2/n5/weinberg_poster/*

precancerous cells taken directly from people. Many recent observations seem to contradict the idea that mutations to a few specific genes lie at the root of all cancers.

Unexplained Phenomena

IN APRIL, FOR EXAMPLE, Muhammad Al-Hajj of the University of Michigan at Ann Arbor and his colleagues reported that they had identified distinguishing marks for a rare subset of cells within human breast cancers that can form new tumors. As few as 100 cells of this type quickly spawned disease when injected into mice lacking an immune system. Tens of thousands of other cells, harvested from the same nine breast malignancies but lacking the telltale marks, failed to do so. "This is the first tumor-initiating cell anyone has isolated for solid tumors," says John E. Dick, a biologist at the University of Toronto who has identified similar cells for leukemia.

The tantalizing implication, Dick says, is that just a small fraction of the cells in a tumor are responsible for its growth and metastasis—and so for the illness and death of the patient. If that is shown to be true for humans as well as mice, it could pose a problem for the mutant-gene theory of cancer. If mutations, which are copied from a cell to its progeny, give tumor cells their powers, then shouldn't all clones in the army be equally powerful?

In fact, most tumors are not masses of identical clones. On the contrary, closer examination has revealed amazing genetic diversity among their cells, some of which are so different from normal human cells (and from one another) that they might fairly be called new species.

A few cancer-related genes, such as *p53*, do seem to be mutated in the majority of tumors. But many other cancer genes are changed in only a small fraction of cancer types, a minority of patients, or a sprinkling of cells within a tumor. David Sidransky of the Johns Hopkins University School of Medicine and his coworkers tested DNA from 476 tumors of various kinds. They reported in April that

© COPYRIGHT 2002 MACMILLAN MAGAZINES LTD. REPRINTED BY PERMISSION

COPYRIGHT 2003 SCIENTIFIC AMERICAN, INC.

the oncogene *BRAF* was altered in two thirds of papillary thyroid cancers. But *BRAF* was not mutated in any of several other kinds of thyroid cancers.

Moreover, some of the most commonly altered cancer genes have oddly inconsistent effects. Bert E. Vogelstein's group at Johns Hopkins found that the much studied oncogenes *c-fos* and *c-erbb3* are curiously less active in tumors than they are in nearby normal tissues. The tumor suppressor gene *RB* was recently shown to be hyperactive—not disabled—in some colon cancers, and, perversely, it appears to protect those tumors from their autodestruct mechanisms.

The "two hit" hypothesis—that both alleles of a tumor suppressor gene must be deactivated—has also been upended by the discovery of a phenomenon called haploinsufficiency. In some cancers, it turns out, tumor suppressors are not mutated at all. Their output is simply reduced, and that seems to be enough to push cells toward malignancy. This effect has now been seen for more than a dozen tumor suppressor genes, and investigators expect to find many more like them. Searching for the mere presence or absence of a gene's protein is too simplistic. Dosage matters.

Beyond Mutation

RESEARCHERS ARE NOW looking more closely at other phenomena, besides errors in a gene's DNA sequence, that can dramatically alter the dosage of a protein in a cell. Candidates include the loss or gain of a chromosome (or part of one) containing the gene; changes in the concentration of other proteins that regulate how the gene is transcribed from DNA to RNA and translated into a protein; even so-called epigenetic phenomena that alter gene activity by reversible means. All these changes are nearly ubiquitous in established cancers.

"If you look at most solid tumors in adults, it looks like someone set off a bomb in the nucleus," Hahn says. "In most cells, there are big pieces of chromosomes hooked together and duplications or losses of whole chromosomes."

Scientists have yet to settle on a term for the suite of chromosomal aberrations seen in cancer. The word "aneuploidy" once referred specifically to an abnormal number of chromosomes. But more recently, it has been used in a broader sense that also encompasses chromosomes with truncations, extensions or swapped segments. That more inclusive definition serves our purposes here.

Almost a century ago German biologist Theodor Boveri noticed the strange imbalance in cancer cells between the numbers of maternal versus paternal chromosomes. He even suggested that aneuploid cells might cause the disease. But scientists could find no recurrent pattern to the chromosomal chaos—indeed, the genome of a typical cancer cell is not merely aneuploid but is unstable as well, changing every few generations. So Boveri's idea was dropped as the search for oncogenes started to bear fruit. The aneuploidy and massive genomic instability inside tumor cells were dismissed as side effects of cancer, not prerequisites.

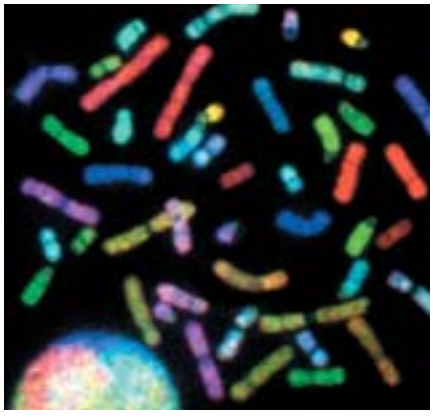
But the oncogene/tumor suppressor gene hypothesis has also failed, despite two decades of effort, to identify a particular set of gene mutations that occurs in every instance of any of the most common and deadly kinds of human cancer. The list of cancer-related mutations has grown to more than 100 oncogenes and 15 tumor suppressor genes. "The rate at which these molecular markers are being identified continues to increase rapidly," Weinberg and Hahn lamented in their November review. "As a consequence," they added, "it remains possible that each tumor is unique" in the pattern of its genetic disarray.

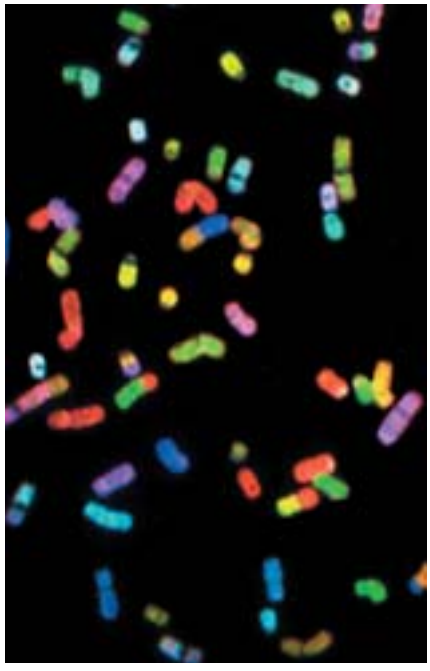
Hahn reflected on this possibility in his Boston office in January. Along with Weinberg, he has pioneered the construction of artificial tumors using mutant cancer genes. But he acknowledged that they cannot be the whole story. "The question is which comes first," he said. "Mutations or aneuploidy?"

There are at least three competing answers. Let us call them the modified dogma, the early instability theory and the all-aneuploidy theory. Encouragingly, the theories seem to be converging as they bend to accommodate new experimental results.

The modified form of the standard dogma revives an idea proposed in 1974 by Lawrence A. Loeb, now at the University of Washington. He and other geneticists have estimated that, on average, random mutation will affect just one gene in any given cell over the course of a lifetime. Something—a carcinogen, reactive oxidants, or perhaps a malfunction in the DNA duplication and repair machinery of the cell—must dramatically accelerate

ABERRANT CHROMOSOMES IN A CANCER CELL can alter the dosage of thousands of genes at once. A healthy cell (*below*) contains one pair of each of the 22 kinds of chromosomes (*distinct colors*), plus two sex chromosomes. In a malignant cell (*right*), some chromosomes contain arms of different types (*multicolored, at left edge*). Others are missing limbs (*royal blue*) or are present in the wrong number (*lime green*).



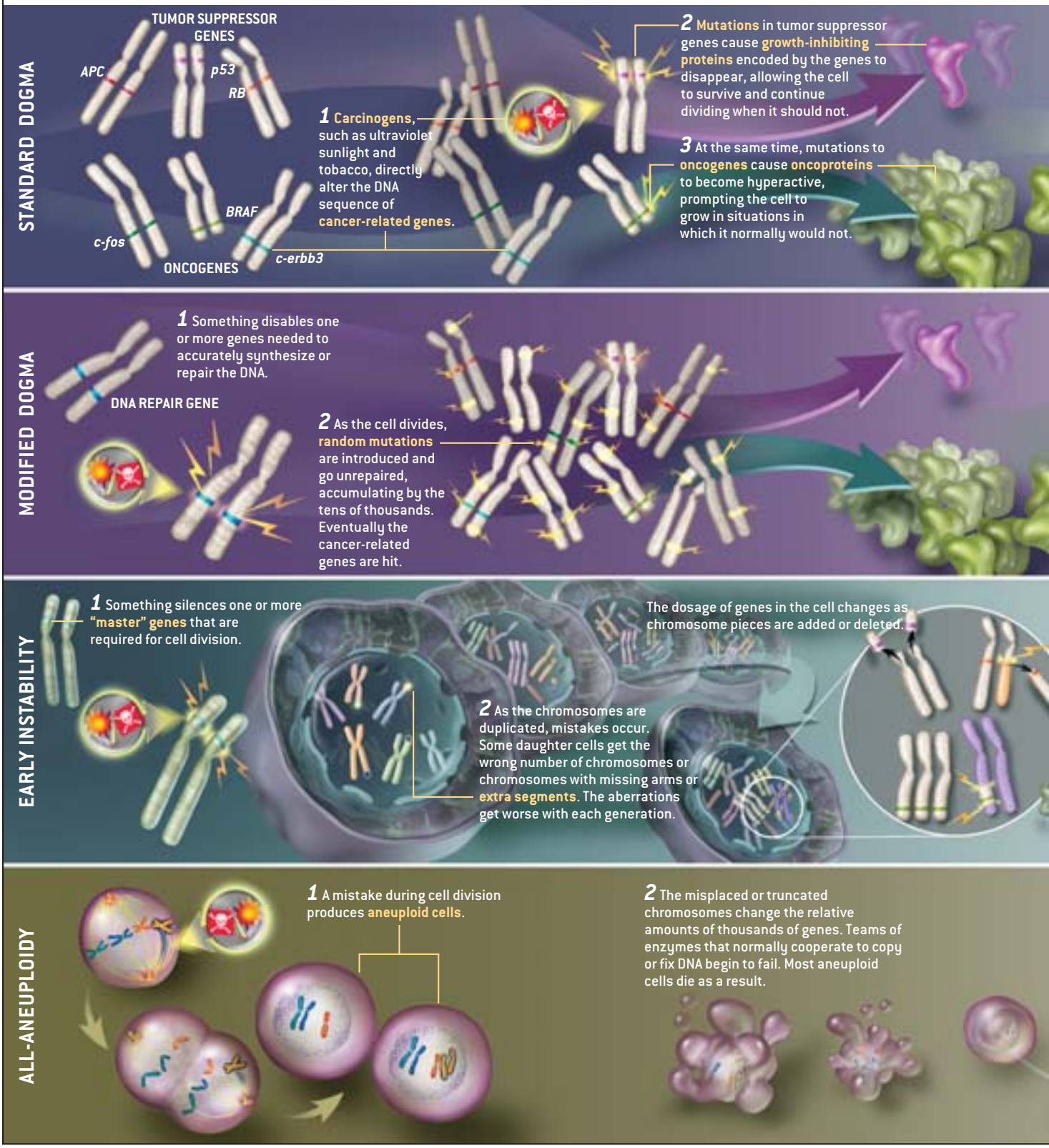


HESED M. PADILLA-NASH AND THOMAS REID *National Cancer Institute*

COPYRIGHT 2003 SCIENTIFIC AMERICAN, INC.

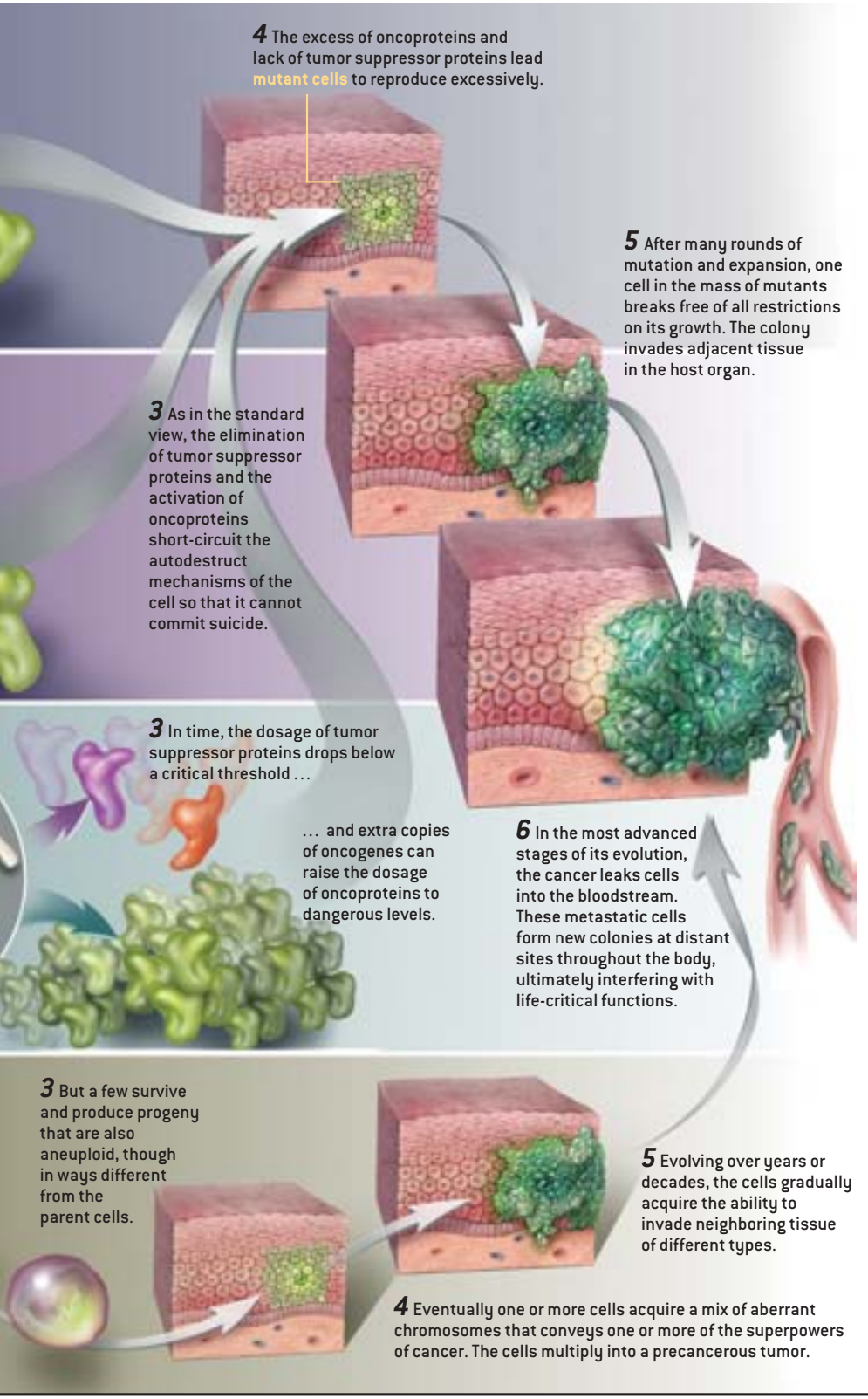
THE GENESIS OF CANCER: FOUR THEORIES

FOR DECADES, the most widely accepted view of how cancer begins has been that mutations to a handful of special genes eliminate tumor suppressor proteins and activate oncoproteins. More recently, three alternative theories have gained currency. One modifies the standard paradigm by postulating a dramatic increase in the accumulation of random mutations throughout the genomes of precancerous cells. Two other theories focus on the role of aneuploidy:



COPYRIGHT 2003 SCIENTIFIC AMERICAN, INC.

large-scale aberrations in the chromosomes. Aneuploidy could lead to genomic instability early on and later mutate known cancer genes. Or it may form tumors through an almost infinite variety of genetic changes.



the mutation rate, Loeb argues. "I think that is probably right," Hahn concurs. Otherwise, he says, "cells wouldn't accumulate a sufficient number of mutations to form a tumor."

Loeb believes that "early during the genesis of cancer there are enormous numbers of random mutations—10,000 to 100,000 per cell." Evidence for the theory is still slim, he acknowledges. Counting random mutations is hard; scientists must compare the genomes of individual cells letter by letter. Advances in biotechnology have only recently made that feasible.

The modified dogma would thus add a prologue to the long-accepted life history of cancer. But the most important plot points in that story are still mutations to genes that serve to increase the reproductive success of cells. Mangled and ever changing chromosomes are but fortuitous by-products.

Unstable from the Outset

CRISTOPH LENGAUER and Vogelstein of Johns Hopkins, both well-known colon cancer specialists, have proposed an alternative theory in which chromosomal instability can occur early on. The genetic flux then combines forces with natural selection to produce a benign growth that may later be converted to an invasive malignancy and life-threatening metastases.

In their hypothesis, there are several "master" genes whose function is critical for a cell to reproduce correctly. If as few as one of these genes is disabled, either by mutation or epigenetically, the cell stumbles each time it attempts the carefully choreographed dance of cell division, muddling some of the chromosomes into an aneuploid state. One result is to increase 100,000-fold the rate at which cells randomly lose one of the two alleles of their genes. For a tumor suppressor gene, a lost allele may effectively put the gene out of commission, either because the remaining copy is already mutated or because of the haploinsufficiency effect. Lengauer and Vogelstein still assume that some cancer genes must be altered before a malignancy can erupt.

In December 2002, together with Martin A. Nowak and Natalia L. Komarova

CHRISTY KRAMES

COPYRIGHT 2003 SCIENTIFIC AMERICAN, INC.

of the Institute for Advanced Study in Princeton, N.J., Lengauer and Vogelstein published a mathematical analysis that applied this theory to nonhereditary colon cancer. Even if there are as few as half a dozen master genes in the human genome, they calculated, it is very likely that a master gene will be disabled before a particular cancer gene is hit.

Calculations are fine, but only empirical evidence is persuasive. Some recent studies do support the early instability theory. In 2000 Lengauer's laboratory examined colon adenomas—benign polyps that occasionally turn malignant—and observed that more than 90 percent had extra or missing pieces of at least one chromosome. More than half had lost the long arm of chromosome 5, home to the *APC* tumor suppressor gene, long implicated in the formation of colon cancer. Other researchers have discovered similarly aberrant chromosomes in precancerous growths taken from the stomach, esophagus and breast.

The early instability theory still has some loose ends, however. How can cells with shifty chromosomes outcompete their stable counterparts? Under normal conditions, they probably do not, suggests immunologist Jarle Breivik of the University of Oslo. But in a "war zone," where a carcinogen or other stressor is continually inflicting damage to cells, normal cells stop dividing until they have completed repairs to their DNA. Genetically unstable cells get that way because their DNA repair systems are already broken. So they simply ignore the damage, keep on proliferating, and thus pull ahead, Breivik hypothesizes.

He cites an experiment in which Lengauer and his colleagues exposed human cell lines to toxic levels of a carcinogen in broiled meat. Only a few cells developed resistance and survived. All were genetically unstable before exposure to the toxin.

But what jumbles the chromosomes in the first place? No genes have yet been conclusively identified as master genes, although several strong suspects have surfaced. German A. Pihan of the University of Massachusetts Medical School and his co-workers may have uncovered a clue in their study, published in March, of 116 premalignant tumors caught before they had invaded neighboring tissues of the cervix, prostate and breast. Thirty to 72 percent of the growths contained defective centrosomes, structures that appear during cell division to help separate the newly duplicated chromosomes from the originals. Unsurprisingly, most of those cells were aneuploid. Scientists are still working out all the genes that control centrosome formation and function; any of them might be a master gene.

Aneuploidy All the Way Down

ON THE OTHER HAND, maybe cells can become malignant even before any master genes, oncogenes or tumor suppressor genes are mutated. Peter H. Duesberg and Ruhong Li of the University of California at Berkeley have put forth a third theory: nearly all cancer cells are aneuploid (leukemia being one exception) because they start that way. Lots of things can interfere with a dividing cell so that one of its daughter cells is cheated of its normal complement of 46 chromosomes and the other daughter is endowed with a bonus. Asbestos fibers, Duesberg notes, can physically disrupt the process.

Most aneuploid cells are stillborn or growth-retarded. But in the rare survivor, he suggests, the dosage of thousands of genes is altered. That corrupts teams of enzymes that synthesize and maintain DNA. Breaks appear in the double helix, destabilizing the genome further. "The more aneuploid the cell is, the more unstable it is, and the more likely it will produce new combinations of chromosomes that will allow it to grow anywhere," Duesberg explains.

Unlike the three other theories, the all-aneuploidy hypothesis predicts that the emergence and progress of a tumor are more closely connected to the assortment of chromosomes in its cells than to the mutations in the genes on those chromosomes. Some observations do seem to corroborate the idea.

In May, for instance, Duesberg, working with scientists at the University of Heidelberg, reported on experiments with normal and aneuploid hamster embryos. The more the cells deviated from the correct number of chromosomes, the faster aberrations accumulated in their chromosomes. Genomic instability rose exponentially with the degree of aneuploidy.

Thomas Reid, chief of cancer genomics at the National Cancer Institute, has obtained supporting evidence in humans from his investigation of aneuploidy in cervical and colorectal cancers. "Unequivocally, there are recurrent patterns of genomic imbalances," Reid avers. "Every single case of [nonhereditary] colorectal cancer, for example, has gains of chromosomes 7, 8, 13 or 20 or a loss of 18. In cervical cancer, aneuploidy of chromosome 3 happens very early, and those cells seem to have a selective advantage." Reid finds the average number of abnormal chromosomes increasing gradually from 0.2 in a normal cell to 12 in the cells of metastatic colon tumors.

"So I actually think Duesberg is right that aneuploidy can be the first genetic aberration in cancer cells," Reid says. "But he also argues that no gene mutations are required. This is simply not true."

Stopping Cancer at Its Roots

NEITHER THE STANDARD dogma nor any of the new theories that challenge it can fully untangle the knotted roots of

BRANCHING POINTS IN THE

1914 Boveri suggests that aberrant chromosomes may cause cancer

1927 Hermann J. Muller observes that radiation mutates cells

1951 Muller proposes theory that multiple mutations turn a cell malignant

1915 1920 1925 1930 1935

COPYRIGHT 2003 SCIENTIFIC AMERICAN, INC.

EVOLUTION OF CANCER THEORY


Muller


Vogelstein


Duesberg

1960 Discovery that an exchange of DNA between chromosomes 9 and 22 leads to chronic myelogenous leukemia

1971 Alfred G. Knudson explains different rates of inherited and spontaneous retinal cancer with the hypothesis that two "hits," or damaging mutations, are needed to disable both alleles of the *RB* gene and that one mutation can be inherited

1974 Loeb argues that random mutations must accumulate much faster than is normal inside cells that become malignant

1986 Weinberg and colleagues isolate *RB*, the first tumor suppressor gene

1990 Vogelstein and Eric R. Fearon publish a model of sequential gene mutations that lead to colon cancer

1997 Lengauer, Vogelstein and co-workers demonstrate dramatic increase in gain and loss of chromosomes in colon tumor cells and propose that chromosomal instability is a critical early event that leads to the mutation of oncogenes and tumor suppressor genes

1999 Duesberg and collaborators publish detailed theory of how aneuploidy may be sufficient to cause cancer itself, even without mutations to any particular set of genes

2002 Reid identifies recurrent patterns of aneuploidy in human cervical and colon cancers

2003 The number of identified cancer genes, now well over 100, continues to grow rapidly

1940 1945 1950 1955 1960 1965 1970 1975 1980 1985 1990 1995 2000

From left to right: SPL (*Boveri*); HULTON-DEUTSCH COLLECTION/CORBIS (*Muller*); ALEX WONG Getty Images (*Vogelstein*); MARC GELLER (*Duesberg*)

the 100-odd diseases we call cancer and explain them as variations of a single principle. And all the theories will need to be expanded to incorporate the role of epigenetic phenomena, which may be pivotal but is still rather mysterious.

It is important to determine which of the ideas is more right than the others, because they each make different predictions about the kinds of therapy that will succeed best against the most common and lethal cancers. In the standard view, tumors are in effect addicted to the proteins produced by oncogenes and are poisoned by tumor suppressor proteins. Medicines should therefore be designed to break the addiction or supply the poison. Indeed, this strategy is exploited by some newer drugs, such as Gleevec (for rare forms of leukemia and stomach cancer) and Herceptin (for one variety of advanced breast cancer).

But all existing therapies, including Gleevec and Herceptin, fail in some patients because their tumors evolve into a resistant strain. Loeb fears that there may be no easy way around that problem. "If I am right, then within any given tumor, which contains roughly 100 million cells, there will be cells with random mutations that protect them from any treatment you can conceive," Loeb says. "So the best you can hope for is to delay the tumor's growth. You are not going to cure it."

For the elderly—who, after all, are the main victims of cancer—a sufficient delay may be as good as a cure. And even better than slowing the growth of a tumor would be to delay its formation in the first place. If Lengauer and other adherents of the early instability theory succeed in identifying master genes, then it should also be possible to make drugs that protect or restore their function. Lengauer says his group has already licensed cell lines to the pharmaceutical industry to use in drug screening.

Screening of a different kind may be the best approach if the all-aneuploidy theory is correct. There are no known means of selectively killing cells with abnormal chromosomes. But a biopsy that turns up a surfeit of aneuploid cells might warrant careful monitoring or even preventive surgery in certain cases. And Duesberg suggests that foods, drugs and chemicals should be tested to identify compounds that cause aneuploidy.

One day science will produce a definitive answer to the question of what causes cancer. It will probably be a very complicated answer, and it may force us to shift our hope from drugs that cure the disease to medicines that prevent it. Even without a clear understanding of why, doctors have discovered that a daily baby aspirin seems to prevent colon adenomas in some adults. The effect is small. But it is a step from chemotherapy toward a better alternative: chemoprevention. SA

W. Wayt Gibbs is senior writer.

MORE TO EXPLORE

Aneuploidy Precedes and Segregates with Chemical Carcinogenesis. Peter Duesberg, Ruhong Li, David Rasnick, Charlotte Rausch, Andreas Willer, Alwin Kraemer, George Yerganian and Ruediger Hehlmann in *Cancer Genetics and Cytogenetics,* Vol. 119, No. 2, pages 83–93; June 2000.

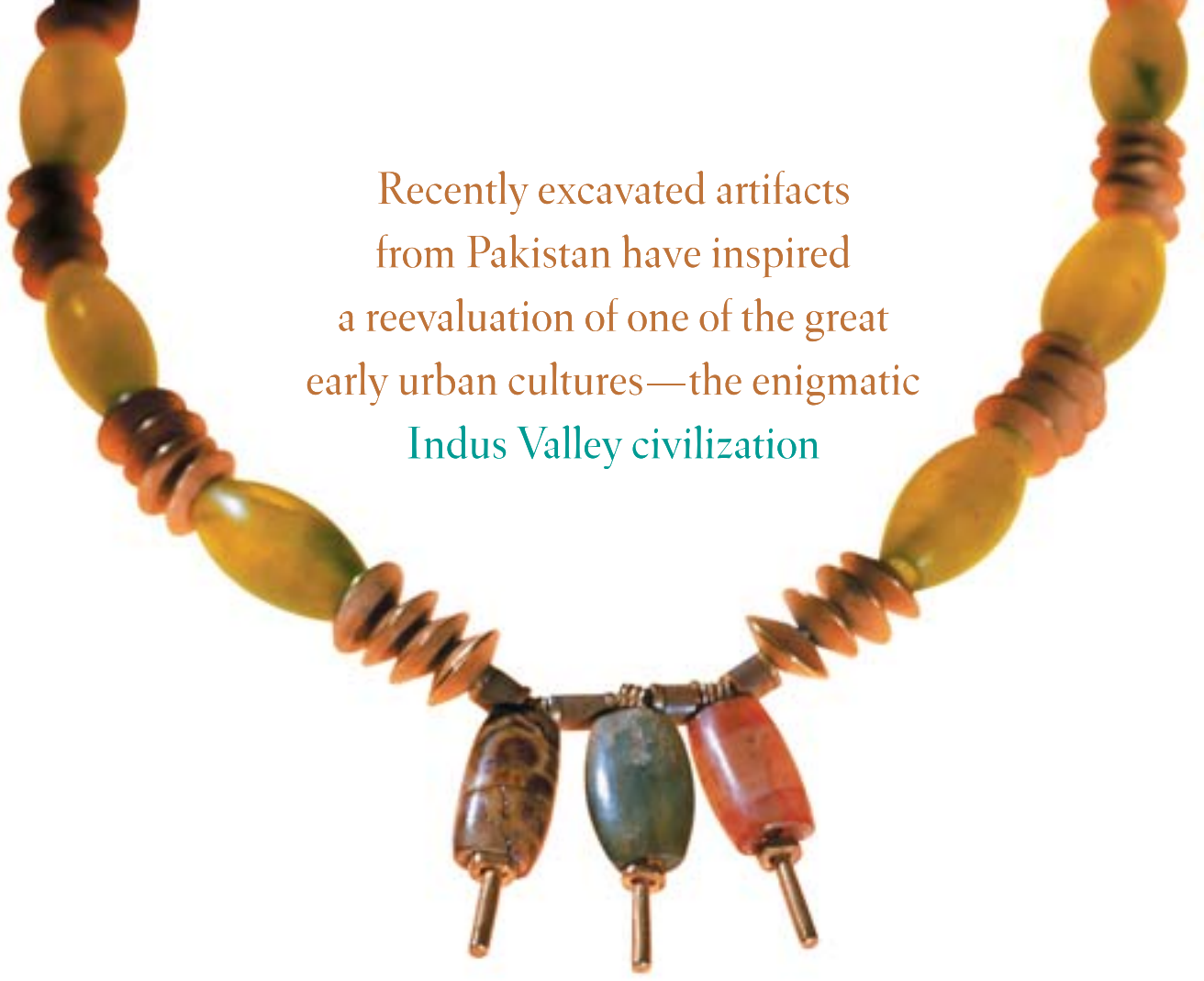
Chromosome Segregation and Cancer: Cutting through the Mystery. Prasad V. Jallepalli and Cristoph Lengauer in *Nature Reviews Cancer,* Vol. 1, No. 2, pages 109–117; November 2001.

Rules for Making Human Tumor Cells. William C. Hahn and Robert A. Weinberg in *New England Journal of Medicine,* Vol. 347, No. 20, pages 1593–1603; November 14, 2002.

Multiple Mutations and Cancer. Lawrence A. Loeb, Keith R. Loeb and Jon P. Anderson in *Proceedings of the National Academy of Sciences USA,* Vol. 100, No. 3, pages 776–781; February 4, 2003.

COPYRIGHT 2003 SCIENTIFIC AMERICAN, INC.

COPYRIGHT 2003 SCIENTIFIC AMERICAN, INC.



Recently excavated artifacts from Pakistan have inspired a reevaluation of one of the great early urban cultures—the enigmatic Indus Valley civilization

UNCOVERING THE KEYS TO THE LOST INDUS CITIES

By Jonathan Mark Kenoyer

COLORFUL STONE BEADS such as those from Mohenjo Daro (*above*) were more than mere baubles for the ancient peoples of the Indus River Valley civilization; they served as symbols of wealth, status and power. Skilled artisans developed sophisticated craft technologies to fabricate these and other kinds of ornaments for the elite classes of this early urban society. Archaeologists interpret the strikingly carved soapstone figure, the so-called priest-king (*left*), to be a member of the Indus Valley ruling class.

RANDY OLSON *Aurora Photos*; DEPARTMENT OF ARCHAEOLOGY AND MUSEUMS, GOVERNMENT OF PAKISTAN (*left* and *above*)

COPYRIGHT 2003 SCIENTIFIC AMERICAN, INC.

INDUS MERCHANTS stamped glazed seals bearing animal images and the undeciphered Indus writing symbols into soft clay tags attached to trade goods to indicate ownership and accounting information.

In the mid-1980s, during our first few seasons excavating the long-dead city archaeologists call Harappa, my colleagues and I watched the passage of the annual spring fairs without realizing their implications for our studies of the ancient Indus civilization. Every year in Pakistan's Indus River Valley, people living in villages travel to larger towns to attend festivals called *sang*, "gathering fairs" where musicians, performers and circus troupes entertain the crowds while itinerant merchants and traders hawk their wares. During the single-day event, women present religious offerings to professional pilgrims, who, on the women's behalf, will later petition the Sufi saint Sakhi Sarwar for healthy children, especially sons. At day's end the holy sojourners and their colorful secular entourage make their way by foot and donkey cart to the next town along the region's age-old trade routes [*see illustration on page 70*].

Eventually various groups of devout wayfarers arriving from different parts of the country converge on the saint's tomb hundreds of kilometers away in the Baluchistan hills. Here they deliver the alms and prayers they have collected during their treks to the saint's precinct. Making their way along the same path the next year, the pilgrims bear salt and sacred earth from the distant shrine, tokens of favor for the fortunate mothers and their precious newborns.

Our view of the *sang* changed significantly when we began digging inside the southern gateway of Harappa, just adjacent to the traditional fairgrounds. As we sifted the uppermost soil from the newly opened trenches, we recovered modern pottery shards, hundreds of pieces of recently manufactured glass bangles, some contemporary coins, lead pellets from the air guns fired at balloons during the festival, fragments of plastic and metal toys, even a gold earring. Just below the surface levels, however, we began finding ancient Harappan artifacts that were surprisingly similar to the modern flotsam: broken pottery, terra-cotta bangles, clay marbles, toy carts, figurine fragments, and, occasionally, inscribed tablets and weights that were probably used in trade and taxation at the city's entrance.

The buried remains suggested that a thriving market existed in the vicinity and that the activities conducted there some four millennia ago were not unlike those occurring even now. Our further work has confirmed that belief. As today, nearby villagers would come to the big city on special market or festival days to participate in ceremonies, to renew family or clan ties, or to buy (or barter for) specially crafted items such as ornaments and pottery. Along the very same pathways trod by present-day pilgrims and their followers, raw materials from the hinterlands once arrived at the gates of Harappa, where the artisans in the city's many workshops transformed them into the finished luxury products that were later purchased by local elite personages and shipped far afield for export markets. Then, as now, people in the Indus Valley wore ornaments and jewelry to demonstrate their wealth and status. To modern observers, the continuity of day-to-day life in this region from deep in the past to the present is rather astounding. As archaeologists, we can try to determine whether these similarities result from cultural choices or from the fact that the available materials and technologies have not changed much over the millennia.

The enigmatic Indus Valley civilization was one of the four great early Old World state-cultures, along with Mesopotamia, Egypt and China's Yellow River civilization. But much less is known about it because, unlike the other ancient

Overview/*Crafts as Cultural Clues*

- The ancient Indus Valley civilization, which first arose in what is now Pakistan and western India after 2600 B.C.E., is the least known of the four major early urban cultures in the Old World, which included Mesopotamia, Egypt and China.
- With no "Rosetta stone" to rely on, linguists have failed to decipher the ancient Indus writing, so researchers must study stone beads and other surviving evidence to try to understand the social, economic and political structure of this once extensive state.
- Through careful investigation, scientists have reproduced the techniques master craftsmen used to make the many ornamental, commercial and ritual artifacts excavators have recovered from the long-buried ruins of the Indus cities. Reading their fine handicrafts for clues, archaeologists are at last developing a more detailed portrait of this extinct civilization.

RANDY OLSON *Aurora Photos*; HARAPPA ARCHAEOLOGICAL RESEARCH PROJECT AND DEPARTMENT OF ARCHAEOLOGY AND MUSEUMS, GOVERNMENT OF PAKISTAN

COPYRIGHT 2003 SCIENTIFIC AMERICAN, INC.

urban cultures, linguists have yet to decipher the Harappan script we see on recovered seals, amulets and pottery vessels. In our ongoing attempt to understand how the now vanished people of the Indus culture ordered their society and to determine the sources of political, economic, military and ideological (religious) power in this remarkably extensive and urbanized state, my co-workers and I have to draw clues from the miscellaneous material we dig up and from the layout and architecture of the cities and settlements we excavate.

The Harappan writings have not been totally useless, however. Although our inability to translate the symbols that artisans and others inscribed on objects has prevented us from learning directly how certain individuals and communities gained and maintained power, we have gleaned insights from examining the context of the writing's use. These studies, together with recent analyses of the advanced crafts that have survived the centuries at Harappa, have begun to yield a new understanding of the social power structure in the Indus civilization.

Hidden Cities

IN THE 1920S archaeologists excavating old mounds of soil and refuse that covered the two large Bronze Age cities of Harappa and Mohenjo Daro ("Mound of the Dead" or "Mound of Mohen") in what are now the Pakistani provinces of Punjab and Sindh brought the Indus civilization to the world's attention. That a major state had flourished on the rich floodplains of the great trans-Himalayan river was unexpected. Subsequent surveys and excavations in western India and Pakistan have uncovered more than 1,500 additional settlements distributed over an area the size of western Europe and twice that of Mesopotamia or ancient Egypt. Although the Indus Valley people did not produce monumental stone carvings and did not bury their dead with their wealth, they constructed large, well-planned cities and made exquisite luxury items that were traded and exported to distant markets in the Persian Gulf, Central Asia and Mesopotamia. The similarities in site layout and artifact style throughout the Indus region reflect a surprisingly uniform economic and social structure.

In 1986 the late George F. Dales of the University of California at Berkeley established the Harappa Archaeological Research Project, a long-term multidisciplinary study effort that I now co-direct with Richard H. Meadow of Harvard University and Rita Wright of New York University, in collaboration with Pakistan's Federal Department of Archaeology and Museums. We conduct on-site investigations and laboratory research to study the original Harappan settlement and to trace the evolution of the larger city that emerged at the site. That work has revealed several phases of development.

The Indus cities established their economic base on agricultural produce and livestock, supplemented by fishing and hunting. Both the common people and the elite classes derived additional income from the production and trade of commodities, including cotton and woolen textiles as well as a variety of craft items.

The earliest village settlement at Harappa (called the Ravi phase) dates from before 3300 B.C.E. to around 2800 B.C.E., a time when the Sumerians were building their first ziggurats and elaborately decorated temples and the Egyptians were burying their rulers and vast hoards of wealth in mud-brick tombs. Farming an environment similar to the agricultural lands of the Fertile Crescent in the Middle East, the ancient Indus peoples herded cattle and cultivated wheat, barley, legumes and sesame. Specialized craft technologies spread among the early settlements along trade networks, which likewise disseminated a shared set of religious symbols and artifact styles throughout the region.

Archaeologists have found other small farming communities from this period to the north and south of Harappa along the Ravi River, but none of these hamlets expanded into a larger town. In the limited exposed area of the Ravi-period Harappa, investigators have turned up signs of the production of both terra-cotta and stone beads and bangles. The terra-cotta items were probably worn by children or commoners, or both, whereas the more exotic stone and seashell ornaments most likely adorned local upper-class populations. Through careful analysis of the raw materials and comparison

INDUS VALLEY CIVILIZATION, which arose some 4,800 years ago, eventually included more than 1,500 cities and settlements spread over an area the size of western Europe in present-day Pakistan and western India.

EDWARD BELL

COPYRIGHT 2003 SCIENTIFIC AMERICAN, INC.

to known source regions, archaeologists have shown that some of the raw materials used by the early Ravi craftsmen arrived at the site from 300 to 800 kilometers away. Impressions of woven fabric on small terra-cotta beads provide evidence of textile production, possibly of both cotton and wool.

This early site also contains the first indication of abstract symbols, or pictographs, scratched onto pottery. Current studies suggest that some of these symbols were retained in the later formalized Indus script, much as ancient Mesopotamian and Egyptian symbols on pottery and clay tablets from around 3500 B.C.E. and 3200 B.C.E., respectively, later found their way into cuneiform and hieroglyphic writing.

Between 2800 and 2600 B.C.E., Harappa became a thriving economic center. Physically, it expanded into a substantial town containing two walled sectors covering an area of more than 25 hectares—about the size of several large shopping malls. In the meantime, many related villages grew up in scattered sites nearby. During this period, called the Kot Dijian (after the site of Kot Diji to the south), artisans developed new craft technologies for making widely distributed products such as gray fired bangles and faience (a form of glazed pottery), the fabrication of which usually involved the use of high-temperature kilns [*see box on page 74*].

At the same time, increasing quantities of stone and other raw materials were being transported to Harappa, probably by oxcarts and flat-bottomed river boats. Toy bullock carts and wheels made of terra-cotta have been found at Harappa from this period, and the later use of carts and boats is well documented. Archaeologists have also recovered figurines of cattle and humans painted with what appear to be woven fabrics, a sign of growing textile production and the importance of clothing, not only for functionality but for public display.

A formal writing system, known as the Early Indus script, emerged in this phase, as evinced by its appearance on numerous pottery fragments and in impressions that a seal, or stamp, made in clay. Merchants employed seals to indicate ownership of storerooms or bundles of goods by stamping clay tags, or bullae, over a cord or a secured door. These square seals, carved in intaglio with geometric or animal motifs, served as economic documentation. Because only a few seals have been discovered, it is likely that they were used by individuals or communities with considerable power, such as landowners, merchants and religious leaders.

Excavations have turned up other signs of economic evolution during the Kot Dijian period. In particular, scientists unearthed a tiny cubical limestone weight. The stone cube weighs 1.13 grams, which corresponds directly with the standardized weight system of the later Indus cities. Its recovery indicates that a system of determining value by weight, possibly for tax or tribute, was established in Harappa two centuries before it became widespread throughout the Indus Valley.

Many religious symbols of horned human forms and ritual designs on pottery began to appear at Harappa and in far-flung corners of the Indus region during the Ravi and Kot Dijian periods, indicating the spread and synthesis of religious and cultural ideas. Whereas in ancient Egypt and Mesopotamia military conquest achieved the integration of distinct regions into single states, this pattern has not been seen in early Indus settlements. The first settlers at Harappa seem to have exploited the rich agricultural and grazing lands along the Ravi River to sustain themselves as they built economic and political power through craft production and trade and then legitimized their standing through religious practices rather than warfare.

The Indus people constructed extensive, **well-planned cities** and made **exquisite** luxury items that were exported to distant markets.

The City Blossoms

THE FULLY URBAN PHASE of Harappa (termed the Harappa phase) began around 2600 B.C.E. and continued until around 1900 B.C.E. For seven centuries Harappa was one of the largest and most powerful economic and political centers

EXULTANT RELIGIOUS PILGRIMS make their way along the age-old Indus Valley trade routes one town a day, bringing prayers and offerings of the faithful to a distant holy site.

RANDY OLSON *Aurora Photos*

COPYRIGHT 2003 SCIENTIFIC AMERICAN, INC.

BIT BY BIT, Pakistani excavators uncover the massive walls of the lost city of Harappa as archaeologist J. Mark Kenoyer (*far right, in group*) explains the site's significance to visitors.

in the Indus Valley, despite the seeming lack of an army. During the spring and late-summer trading seasons, the city would have hosted hundreds of traders who attracted thousands of people from the surrounding rural areas. Depending on the time of year, 40,000 to 80,000 people may have lived in the city. Wealthy patrons and entrepreneurial competition stimulated the development of new technologies and more extensive trade networks. Excavators have found distinctive pottery with widely used ritual motifs at all settlements throughout the greater Indus Valley, along with unique objects such as cubical stone weights, and seals inscribed with Indus writing and a motif depicting a mythical unicorn.

Although most traders operated within the Indus Valley region, some materials were available only from more distant locales. The presence of raw materials and finished goods from Afghanistan and Central Asia indicates that merchants from these areas came to the city bringing lapis lazuli, tin, gold, silver and, perhaps, fine woolen textiles (which have since vanished). Traders would have carried back to the highlands cereal grains and livestock, as well as fine cottons and possibly even silks, but these items are not well preserved. Other nonperishable objects, such as long, elegant beads made of carnelian (a form of red agate) and shell bangles from the Indus Valley, have been dug up in Central Asia and Mesopotamia.

In its prime, Harappa measured more than 150 hectares in area—more than five kilometers in circuit, encompassing three large raised mounds and associated suburbs. The modern town of Harappa, with a population of around 20,000, still occupies a third of the ancient site. The city's architecture and street layout were organized to facilitate access to the different neighborhoods and to segregate the public and private areas. Massive mud-brick walls enclosed each of the raised mounds [*see illustration above*], and narrow gates limited access, permitting only a single oxcart to pass at a time. Masons employed kiln-fired brick to build multistory houses that were placed along north-south and east-west street grids. Major avenues spanned more than eight meters, and some featured central dividers that may have regulated two-way bullock-cart traffic.

Builders dug drinking-water wells in and around the city, and Harappan houses were equipped with bathing areas, latrines and sewage drains. Linked to larger mains, which eventually emptied outside the city walls, the sewers at Harappa would have removed wastewater from the habitation areas, depositing fertile sludge on the surrounding agricultural fields. Save for the Indus cities, no other city in the ancient world featured such a sophisticated water and waste management system. Even during the Roman Empire, some 2,000 years later, these kinds of facilities were limited to upper-class neighborhoods.

During this period, scribes developed a sophisticated writing system comprising more than 400 symbols that the society's elite classes—including traders, landowners and ritual specialists—used as a mechanism for economic control and political power. Archaeologists do not know the language for which the script was developed, but it was probably used for writing more than one language, as was the case in Mesopotamia. We will never know for sure until someone discovers some form of bilingual tablet (an Indus Rosetta stone) that will help scholars break the code of the writing system.

This script became widespread in all the major urban centers of the Indus Valley. Its most prominent use was on seals that also bore animal motifs and ritual objects. The unicorn image [*see illustration on page 68*] is the most common (found on more than 65 percent of known seals), but other animals appear as well, including elephants, humped zebu bulls, water buffaloes, bison, tigers and rhinoceroses. The animal symbols may have represented important clans or official social classes; the writing probably listed the name of the owner and statements of legal legitimacy. The regular appearance of the unicorn sign indicates a widespread and powerful community, possibly merchants.

Beyond seals inscribed with writing and animal signs, excavators have found large pottery vessels for holding merchandise for trade etched with what may have been names of owners or con-

THE AUTHOR

JONATHAN MARK KENOYER is professor of anthropology at the University of Wisconsin–Madison, where he teaches archaeology, ethnoarchaeology, experimental archaeology and ancient technology. His main focus is the Indus Valley civilization of Pakistan and India, where he has conducted research for the past 27 years. Kenoyer received his advanced degrees from the University of California, Berkeley. Since 1986 he has served as co-director and field director of the Harappa Archaeological Research Project, which is funded by the National Endowment for the Humanities, the National Science Foundation, the National Geographic Society, the Peabody Museum of Archaeology and Ethnology of Harvard University, the American School of Prehistoric Research, the University of Wisconsin, the Smithsonian Institution, the Kress Foundation, as well as private donors.

RANDY OLSON *Aurora Photos*

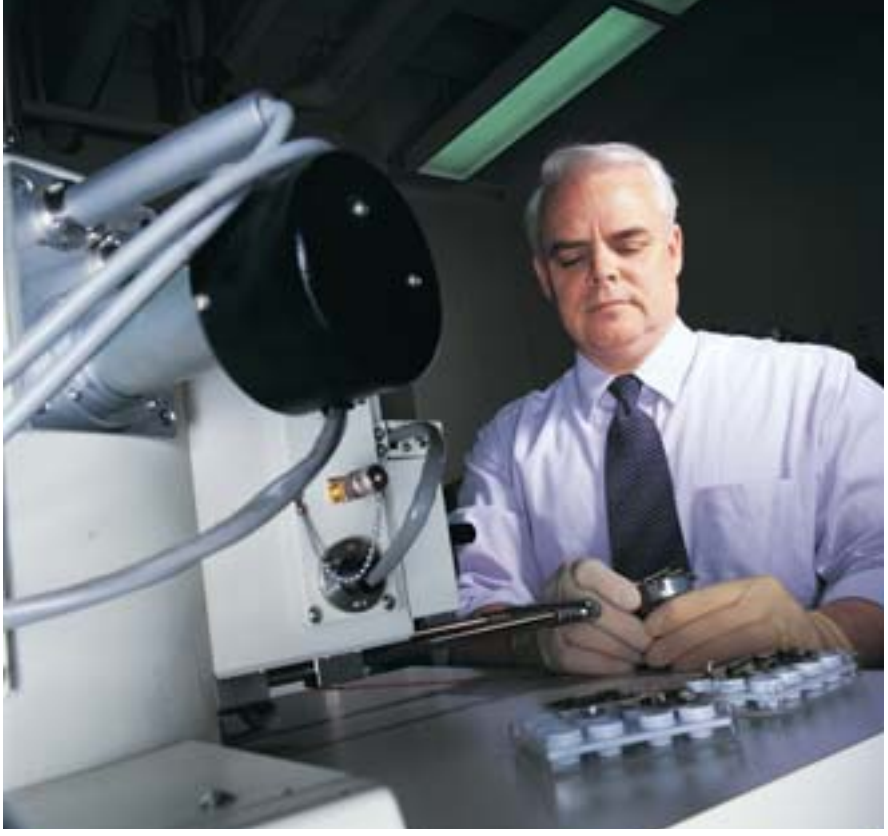
COPYRIGHT 2003 SCIENTIFIC AMERICAN, INC.

signees, together with a description of the contents. Investigators have interpreted examples of script scratched on bronze tools and gold jewelry as personal names or values.

We also see for the first time small tokens or tablets made of steatite (soapstone) incised with script and symbols. Because many of these items feature the same set of signs and what appear to be numbers, researchers think that they served as tools for accounting. The writing may indicate the owner's name, the commodity and its quantity or the creditor to whom it is being presented. Similar tablets made of clay or faience that were stamped with a seal are found quite commonly in the later levels of the site. Some of these tablets have been intentionally snapped in two, possibly to indicate a contract between two parties, each of whom retained half until the transaction or work was completed. Distinctive copper tablets with script and animal motifs and their widespread use in Mohenjo Daro and Harappa are perhaps the earliest evidence for a form of city coinage in the Indus Valley.

Our discovery in 2001 of a workshop that manufactured seals and inscribed tablets at Harappa [*see box on page 74*], combined with the past 16 years of excavation at the site, has provided a new chronology for the development of the Indus writing. In the past, scholars had lumped all the seals and tablets together, but now we can demonstrate that different types of seals and tablets emerged at various times and that the writing itself may have changed over the years. We are currently trying to reach conclusions about the dating of the script changes and expect that this chronology will revolutionize the attempts at decipherment.

Between 2300 and 1900 B.C.E. the urban populations in the Indus Valley expanded, and ornaments, tools and manufacturing technologies grew increasingly diverse. Archaeologists believe that these stylistic changes reflect two competing processes—the desire to differentiate among communities and classes within the cities and attempts to integrate these separate groups by promoting a common ideology.



TELLTALE TOOL MARKS become visible when the author examines ancient Indus Valley artifacts with a scanning electron microscope at the University of Wisconsin–Madison.

In this period, we see writing combined with narrative mythological scenes that indicate its use related to religious themes. Spiritual leaders most likely used seals and tablets to communicate the names of deities and rituals to the common people. Although researchers do not know the names of the Indus deities, we see the repeated motif of a seated male figure in a yogic position, wearing a horned headdress. In one set of narratives, a buffalo is sacrificed in front of the seated deity; in others, he is surrounded by ferocious wild animals. Female deities are also depicted on seals, sometimes with a horned headdress and fighting a tiger. A theme on many seals and impressions on clay tablets is that of a deity strangling two tigers, in some instances depicted standing on top of an elephant. A similar scene found in Mesopotamia is associated with the Gilgamesh epic, in which the hero strangles a pair of lions. The similarities between these motifs are evidence of shared concepts of power and dominance, but we do not know if the specific narrative spread from one region to the other.

As Harappa began to host more people from foreign lands, the elite classes probably felt the need to help legitimize their rule through public rituals and ceremonies that reinforced the principal religious traditions. In Mesopotamia and ancient Egypt, political and religious leaders accomplished this task by constructing massive stone or brick murals covered with religious and political propaganda depicting rulers conquering their enemies, but no similar artifacts have been unearthed in the Indus Valley.

Scribes developed a writing system that the society's elite used as a mechanism for economic control and political power.

The Indus Culture Changes

PREVIOUSLY SCHOLARS argued that the Indus cities were suddenly abandoned around 1750 B.C.E., but our recent work at Harappa has clearly demonstrated that during its late phase, from 1900 to 1300 B.C.E., Harappa was indeed

DAVID NEVALA

COPYRIGHT 2003 SCIENTIFIC AMERICAN, INC.

What the Beads Say

OUR EXCAVATIONS at Harappa have yielded stone beads from all the major occupation phases of the site, indicating that the inhabitants produced them from the earliest settlement onward. Other evidence, particularly small figurines of city dwellers adorned with copious quantities of jewelry, reveals that Harappans often wore multiple strands of beads made from colored and patterned stones. Some of the bead-making areas that we uncovered were probably sponsored by the elite urban classes and produced rare, difficult-to-make beads as symbols of wealth, status and power.

From the unfinished samples and the remains of workshops, we have been able to document how bead styles and drilling techniques evolved over 1,400 years. This continual elaboration was probably stimulated by competition between workshops, market demand, and encouragement by wealthy patrons to create more valuable and attractive ornaments.

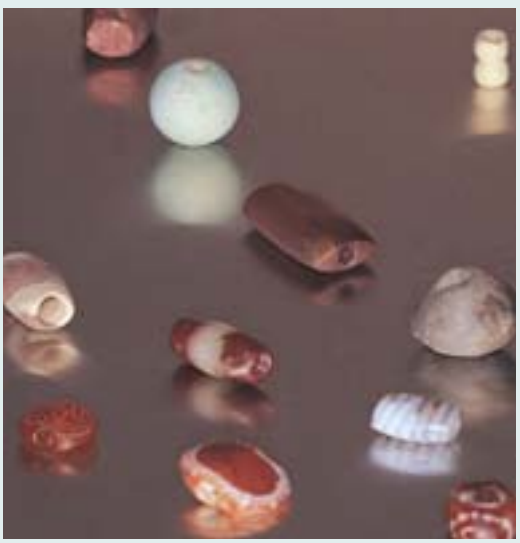
The desire for high-quality beads, and thus refined fabrication methods, was strong even in the early Ravi phase (3300 to 2800 B.C.E.) of Harappa. Although it is relatively simple to make large stone beads, by far the most prevalent types during that period were quite small, between 1.5 and three millimeters in diameter and one to two millimeters long. These were particularly difficult to fashion because, for stringing, they needed tiny holes (one half to three quarters of a millimeter across).

Some of the earliest beads were fabricated from steatite, a soft talc also known as soapstone. Archaeologists have recovered more steatite beads than any other type, and they occur in all parts of Harappa, dating from the Ravi phase through the Late Harappan (1900 to 1300 B.C.E.). Artisans easily bored holes in roughly formed beads of steatite using copper drills or even hardened acacia thorns about half a millimeter in diameter. Then they ground the unfinished beads to size and polished them on grindstones. Finally, the workers fired the beads in kilns, turning them white and converting the mineral into a harder, more durable form.

Some beads of steatite were left with a rough surface, which artisans coated with a finely ground silica frit (a glassy paste made of quartz powder) mixed with copper oxide, which yielded a blue-green glaze when heated. When fired at more than 850 degrees Celsius (1,562 degrees Fahrenheit), steatite crystals give off water and transform into much harder minerals, including cristobalite, enstatite and alumina. Many craftsmen of the Indus region practiced this kind of processing, which may have served as the foundation for later vitreous glazing procedures employed to fabricate faience ornaments, seals and tablets.

From the Ravi through the fully urban phase, bead makers in Harappa and other Indus cities also worked harder stones, such as agate and jasper, which required other drilling techniques: pecking with a stone or copper tool followed by boring with a harder variety of stone drill. Indus scholars have not replicated the pecking technique, so it is not well understood. But it is known that workers perforated some beads with hard stone drills made by shaping tiny slivers of stone into long, tapered cylinders. Hafted onto a thin wooden dowel and turned with a bow, these drills penetrated the beads halfway through from each side to form a continuous passageway for the string. This process was tedious, as both the stone drills and the beads were of approximately the same hardness.

Around 2600 B.C.E., Indus craftsmen discovered a much tougher stone drill, which we have called Ernestite, in honor of English archaeologist Ernest J. H. Mackay, who first discovered it in Chanhu Daro, Pakistan. These highly effective drills remain something of mystery, and despite numerous studies and surveys, we have not yet discovered their precise composition or their source area. Ernestite seems to contain various minerals, and preliminary analysis suggests that it is a fine-grained metamorphic rock composed mostly of quartz, sillimanite, mullite, hematite and titanium oxide. Mullite is rarely found in natural form but is produced in modern high-temperature ceramic materials. Its occurrence in the drills hints that it is the by-product of intentional heating of the original rock, a process that was widely practiced by Harappan craftsmen for hardening steatite and making other rocks easier to chip.



MASTER ARTISANS of the Indus used complex grinding, drilling and decorating methods to fashion colored minerals, often imported from far afield, into exquisite beadwork.

Although we do not know where the first Ernestite drills were manufactured, so far they have been found only at sites in the Indus Valley, which implies that they were unique to that region. They gave the craftsmen a way to make exquisite carnelian beads that were worn by the Indus elites and traded to peoples in Central Asia and Mesopotamia, including Ur. People often wore the long carnelian beads as a belt of multiple strands that were held in place with polished bronze spacers.

Harappan bead makers used Ernestite drills for seven centuries, but around 1900 B.C.E. this drilling technique disappeared in the northern sites, perhaps because of the disruptions in trade networks that occurred at that time.

But the drilling of beads did not stop. Later artisans began perforating stone with hollow tubular copper drills used in combination with abrasives. The Harappans had always employed this method for hollowing out large stone rings and alabaster vessels, but then they miniaturized the technique, working with tiny tubular drills only a millimeter in diameter. Though not as efficient as Ernestite drills, the copper drills could perforate relatively hard stone beads by working halfway through from each side. —*J.M.K.*

DAVID NEVALA; HARAPPA ARCHAEOLOGICAL RESEARCH PROJECT AND DEPARTMENT OF ARCHAEOLOGY AND MUSEUMS, GOVERNMENT OF PAKISTAN

COPYRIGHT 2003 SCIENTIFIC AMERICAN, INC.

Secrets of Harappan Glazing Techniques

FAIENCE, AN ATTRACTIVE glazed ceramic or stone featuring a lustrous sheen, is one of the complex craft techniques used by artisans to create high-status goods for elite consumers throughout Harappan history. Indus Valley faience was stronger than that from Egypt or Mesopotamia because it was made with partially melted quartz that was reground into a fine paste before a second firing, which fused the glassy powder to the core of the object. The Indus artisans needed a stronger form of faience because they were making different types of objects than those used in other regions, specifically faience bangles. They also produced tiny glazed beads and delicately molded tablets inscribed with written symbols used as credit tokens to keep accounts of goods coming into or moving out of the cities.

The Harappan elite classes employed faience not only for decoration and commerce but for ritual purposes as well. Tablets bearing narrative scenes may have been manufactured for special ceremonies and handed out to participants who had offered donations or sacrifices. Control of faience production techniques and the fabrication facilities was essential for Harappan elite classes to create and maintain their symbols of status and power.

Although workers produced faience from easily obtained materials, the technical expertise needed to process them into finished goods was highly specialized. Artisans partially melted powdered rock quartz in high-temperature kilns, using flux additives (fusion aids) made of plant ash to create a glassy frit. They then reground the frit, which they refired at around 940 degrees Celsius (1,724 degrees Fahrenheit) to create dense, glazed faience. When copper oxide or azurite was added to the frit, the resulting goods resembled turquoise or lapis lazuli but were much less likely to discolor when worn against the skin in a hot and humid climate.

PRODUCTION of glazed stone and ceramics, known as faience, required Indus craftsmen to develop sophisticated coating and firing techniques. Copper oxide–bearing minerals (*center*) were added to color the silica glaze blue so that it resembled precious turquoise or lapis lazuli.

Archaeologists have found manufacturing debris from faience production throughout the habitation layers in various parts of Harappa, but for 70 years they had failed to locate faience kilns. Our discovery in 2001 of a small faience workshop revealed that researchers had been looking for the wrong type of kiln. This find was akin to discovering the Harappan mint, because the workshop was also used for making steatite (soapstone) tablets, as well as beads and other ornamental objects—all objects of wealth.

Painstaking excavation and mapping of hundreds of artifacts from the workshop allowed us to reconstruct some of the processes used to fabricate beads, bangles and tablets. To produce molded faience tablets, craftsmen first sawed raw lumps of steatite into thin slabs and then carved them in reverse images to make molds. They used these molds to form tiny tablets that were placed in pottery containers made of sand and straw-tempered clay that would not melt down at the high temperatures needed for glazing. To keep the glazed tablets and other objects from sticking to the pottery firing container, the inside surface was thickly coated with a coarse white powder consisting of burned bone and sometimes ground-up steatite.

No kiln was discovered in the small workshop area, even though we recovered large quantities of charcoal, frothy vitreous yellow-green faience slag and broken firing canisters. After carefully examining the containers and vitrified supports used to stabilize them during firing, I began to suspect that the artisans had relied on a novel firing method. With the assistance of my graduate students, I was able to successfully test this procedure through an experimental reconstruction undertaken at the University of Wisconsin–Madison during the summer of 2001.

It looks as if the Harappan faience glazers assembled two firing containers to form a mini kiln, rather than using bigger, more conventional firing structures. If heated in an open bonfire, they could reach temperatures high enough to glaze small objects. In the experiment, I replicated faience paste as well as the canisters and molds with materials and tools similar to those used by the ancient Harappans. We placed the molded prefired faience tablets and other test objects, including the steatite molds, in the container and covered it with a lid. Then we inserted conical supports to hold the lid up and to leave a small gap to allow flames to enter the miniature firing chamber. I positioned the canister on a low pile of firewood, covered it with more wood and lit the fire, adding more fuel to keep the assembly red-hot.

After about three hours of firing, the faience objects, just barely visible in the crack between the lid and the lower canister, began to glow a deep red-orange and to emit a distinctive odor. Using a thermocouple with a digital pyrometer, we were able to document the firing temperature at around 935 degrees C, the critical glazing point for Harappan faience. We maintained this temperature for about an hour by adding more wood.

After the mini kiln had cooled down, my students and I gathered around to observe the results. Though not identical to faience objects created by Harappan master craftsmen, fully glazed faience tablets and beads resulted. This first attempt of ours indicates that the canister-kiln technique would have been a highly efficient method for producing faience and fired steatite objects. Even more important was that the remains of the process—a pile of charcoal and ash, a cracked firing container, some discarded conical supports, charred bone, and a few rejected beads and tablets—closely resembled what we had found in the workshop at Harappa. —*J.M.K.*

DAVID NEVALA; HARAPPA ARCHAEOLOGICAL RESEARCH PROJECT AND DEPARTMENT OF ARCHAEOLOGY AND MUSEUMS, GOVERNMENT OF PAKISTAN

COPYRIGHT 2003 SCIENTIFIC AMERICAN, INC.

STANDARDIZED STONE WEIGHTS were employed by Harappan traders to measure out high-value merchandise for sale in urban markets.

inhabited. In fact, signs that drains and city walls were not maintained provide proof of crowding and a breakdown of civic order. The remains suggest that the ruling elites were no longer able to control the day-to-day functioning of the urban center. This loss of authority must have eventually led to a reorganization of society, not just in Harappa but throughout the entire region that the upper classes had dominated for 700 years. Similar changes were occurring at the other big cities, such as Mohenjo Daro to the south and Dholavira in western India.

The crisis led to a cessation of the hallmarks of Indus elite culture. The distinctive pottery with ritual motifs and Indus script and traditional square seals with unicorn and other animal motifs disappeared. Cubical weights for taxation and trade fell into disuse, and the international trade networks began to deteriorate. Shells from the coastal regions no longer made their way to the northern sites, and lapis lazuli from the north failed to reach the sites in the plains. In Mesopotamia the texts that had recorded ongoing trade with a region called Meluhha, which is probably the Indus Valley, no longer mentioned it.

There seems to have been no single cause of the decline and reorganization of the Indus civilization but rather an array of factors. The growth of trade and the expansion of Indus settlements onto the Ganges River plain as well as into what is now the state of Gujarat in western India led to the overextension of the Indus political and economic system. Around 1900 B.C.E., one of the major rivers of the Indus Valley, the Ghaggar-Hakra (also called the Saraswati), began to shift its course and eventually dried up, leaving many sites without a viable subsistence base. These communities would have migrated to other farming regions or to cities such as Mohenjo Daro and Harappa, resulting in overcrowding and civic disorder. Without a tradition of army-enforced social integration, leaders had no mechanism for maintaining trade networks and controlling the movement of peoples as they spread out into new regions.

The speed of change varied in different areas, but by 1300 to 1000 B.C.E. a new social order characterized by a distinctive ideology and language began to emerge in the northern Indus Valley and the Ganges River region to the east. According to ancient Indian literary records such as the Vedas and the Mahabharata and Ramayana epics, this area was populated by numerous competing polities practicing Vedic religion and speaking Indo-Aryan languages such as Sanskrit and its various dialects. Our information is hampered by the fact that most of the Indus settlements dating to this period have either been destroyed by later erosion or brick robbing or are covered by continuous inhabitation, which makes excavation impossible. Both Harappa and Mohenjo Daro supported later settlements dating to this time, but these levels have been badly disturbed.

Even though many features of elite Indus culture faded away, some aspects of its urbanism and the important craft technologies survived. Pottery, faience, and copper and bronze continued to be produced in the Indus region, although they were adapted to changing raw materials and social needs. Around 1700 B.C.E. came the first evidence for glass beads produced at Harappa, some 200 years before glass was being made in Egypt. During the subsequent Painted Gray Ware culture (1200 to 800 B.C.E.), glass bottles and bangles as well as beads were being fabricated at sites throughout northern India and Pakistan. It was then, too, that iron production—a new technology that spread throughout the northern Indus Valley and the Ganges River region—emerged. Although the peoples of Anatolia and West Asia smelted iron during this period, the iron manufacture of ancient India and Pakistan was a separate development unrelated to activities in the west.

Much of the Indus culture has yet to be investigated, but the results of recent work at sites such as Harappa are finally lifting the veil from an important civilization that has long been shrouded in mystery. SA

Ancient Indus Valley artifacts are on exhibit in "Art of the First Cities: The Third Millennium B.C. from the Mediterranean to the Indus" at the Metropolitan Museum of Art in New York City through August 17, 2003.

MORE TO EXPLORE

Deciphering the Indus Script. Asko Parpola. Cambridge University Press, 1994.

Excavations at Harappa 1994–1995: New Perspectives on the Indus Script, Craft Activities, and City Organization. Richard H. Meadow and Jonathan Mark Kenoyer in *South Asian Archaeology 1995.* Edited by Bridget Allchin and F. Raymond Allchin. Oxford & IBH, New Delhi, 1997.

Origins of a Civilization: The Prehistory and Early Archaeology of South Asia. F. Raymond Allchin and Bridget Allchin. Viking Penguin, Delhi, 1997.

Trade and Technology of the Indus Valley: New Insights from Harappa, Pakistan. Jonathan Mark Kenoyer in *World Archaeology,* Vol. 29, No. 2, pages 262–280; October 1, 1997.

Ancient Cities of the Indus Valley Civilization. Jonathan Mark Kenoyer. Oxford University Press, Karachi, 1998.

The Indus Valley Mystery. Richard H. Meadow and Jonathan Mark Kenoyer in *Discovering Archaeology,* Vol. 2, No. 2, pages 38–43; April 2000.

For more information on Harappa, visit **www.harappa.com**

RANDY OLSON *Aurora Photos;* HARAPPA ARCHAEOLOGICAL RESEARCH PROJECT AND DEPARTMENT OF ARCHAEOLOGY AND MUSEUMS, GOVERNMENT OF PAKISTAN

COPYRIGHT 2003 SCIENTIFIC AMERICAN, INC.

COPYRIGHT 2003 SCIENTIFIC AMERICAN, INC.

By Bruce Stutz

PUMPHEAD

In what has become almost routine, the heart-lung machine "breathes" for patients during coronary-bypass operations. But could this lifesaving device have a dark side?

The last thing I remembered was the cold room with a stainless-steel ceiling. I was about to undergo open-heart surgery, an experience shared by about 500,000 people in the U.S. every year. After the anesthesia took effect, surgeons made an incision in my groin to reach my femoral vein and artery. Through the vein they threaded a tube called a cannula into the right atrium,

COPYRIGHT 2003 SCIENTIFIC AMERICAN, INC.

an upper chamber inside my heart. This tube, and another attached to the artery, was connected to a cardiopulmonary-bypass pump, also known as a pump oxygenator or a heart-lung machine. A dose of heparin kept my blood from clotting as it traversed the machine's innards. As the venous blood passed through the oxygenator, it was cooled to prevent tissue damage. My body temperature lowered to 25 degrees Celsius (77 degrees Fahrenheit)—deep hypothermia. Surgeons inserted an inflatable clamp into my aorta to seal it off. Two liters of cold potassium solution stopped my heart, and for the next two hours the machine took over. An eight-inch incision below my right breast allowed the doctors to pass cameras and instruments between my ribs and then to repair my congenitally defective heart valve.

I left the hospital a week later. The incision healed quickly and painlessly. In a couple weeks I was out and about on slow but successively longer walks. Within a month I was back in the gym. Mentally, however, I felt a bit hazy, a little disconnected and sometimes even lost. I soon learned that the physician's warning, "You may be a little depressed for a time afterward," would not do justice to the long, dumbfounding struggle against what seemed to be the sudden onset, at 51 years of age, of attention-deficit disorder or incipient senility. Adrift in a clueless no-man's-land, I felt my moods range from querulous to despondent. I couldn't muster the concentration to deal with the problem. I just wanted to be able to think. Think anything.

I didn't know it at the time, but I was suffering from what surgeons among themselves call "pumphead," an all-too-apt appellative that refers to the dimwitted state in which patients seem to linger after being hooked up to a heart-lung machine for open-heart or, in my case, valve surgery. Other symptoms include patchy recall, social difficulties and personality changes. Pumphead was long suspected from anecdotal and journal reports of patients tested soon after their operations. But only recently did a five-year study of bypass patients indicate that after an initial recovery of mental capabilities in the first few months, the condition often worsens later and persists for years. Could the familiar heart-lung machine—which annually provides life-giving oxygen to blood during 900,000 coronary-bypass operations around the world—be at fault?

Adrift in a clueless no-man's-land, I felt my moods range from querulous to despondent.

Making of the Machine

FIFTY YEARS AGO John Heysham Gibbon, Jr., of Jefferson Medical College in Philadelphia performed the first successful human surgery using a heart-lung machine. He had begun developing the device in the 1930s, as a research fellow in surgery at Harvard Medical School. After years of animal trials, and with engineering advice and financing from IBM chairman Thomas J. Watson and others, Gibbon solved what was then the major impediment to heart surgery—giving doctors enough time to operate.

Before Gibbon's pump came into use, induction of, or cooling, the patient's body to slow metabolism and blood circulation gave surgeons a 15-minute window to complete their work. Any longer and the body and brain suffered from a lack of oxygen. Gibbon's machine took blood that would normally go to the heart and lungs, oxygenated it and pumped it back to the arterial system through the aorta. Today the pump's use has become commonplace. Surgeons perform so many coronary-artery-bypass graftings, or CABGs, that they commonly refer to them as "cabbages." But for decades the long-term effects of those operations received little serious study.

That has changed in recent years. In 2001 a *New England Journal of Medicine* article by Mark F. Newman, chair of Duke University Medical Center's department of anesthesiology, and his colleagues revealed that even after five years many coronary-bypass patients still struggled with severe mental impairments. Previous studies had followed subjects for only up to six months. "Adjectives such as 'subtle,' 'transient,' and 'subclinical' have been used to describe the cognitive decline that occurs after CABG, but such descriptions minimize the importance of these changes to clinicians, patients, and their families," the article concludes. Adds Newman: "The month after the study came out, we received 4,000 e-mails from bypass patients. A lot of people were happy just to know they weren't crazy."

In the study, Newman and his group administered five cognitive tests—short-story recall, repetition of number series,

Overview/*Cognitive Decline*

- The heart-lung machine, first used on humans in 1953, revolutionized coronary surgery by giving doctors an hour or more to operate on a still heart. Previous techniques allowed only 15 minutes.
- Since the inception of the machine, medicine has recorded cognitive decline in patients hooked up to it. The condition, later nicknamed "pumphead," was thought to be short-lived and was often attributed to the general trauma of surgery.
- Recently, long-term studies of patients have shown that pumphead may worsen over time and persist for years. Many factors could be involved, but microscopic cell debris and bubbles generated by the machine are under suspicion.

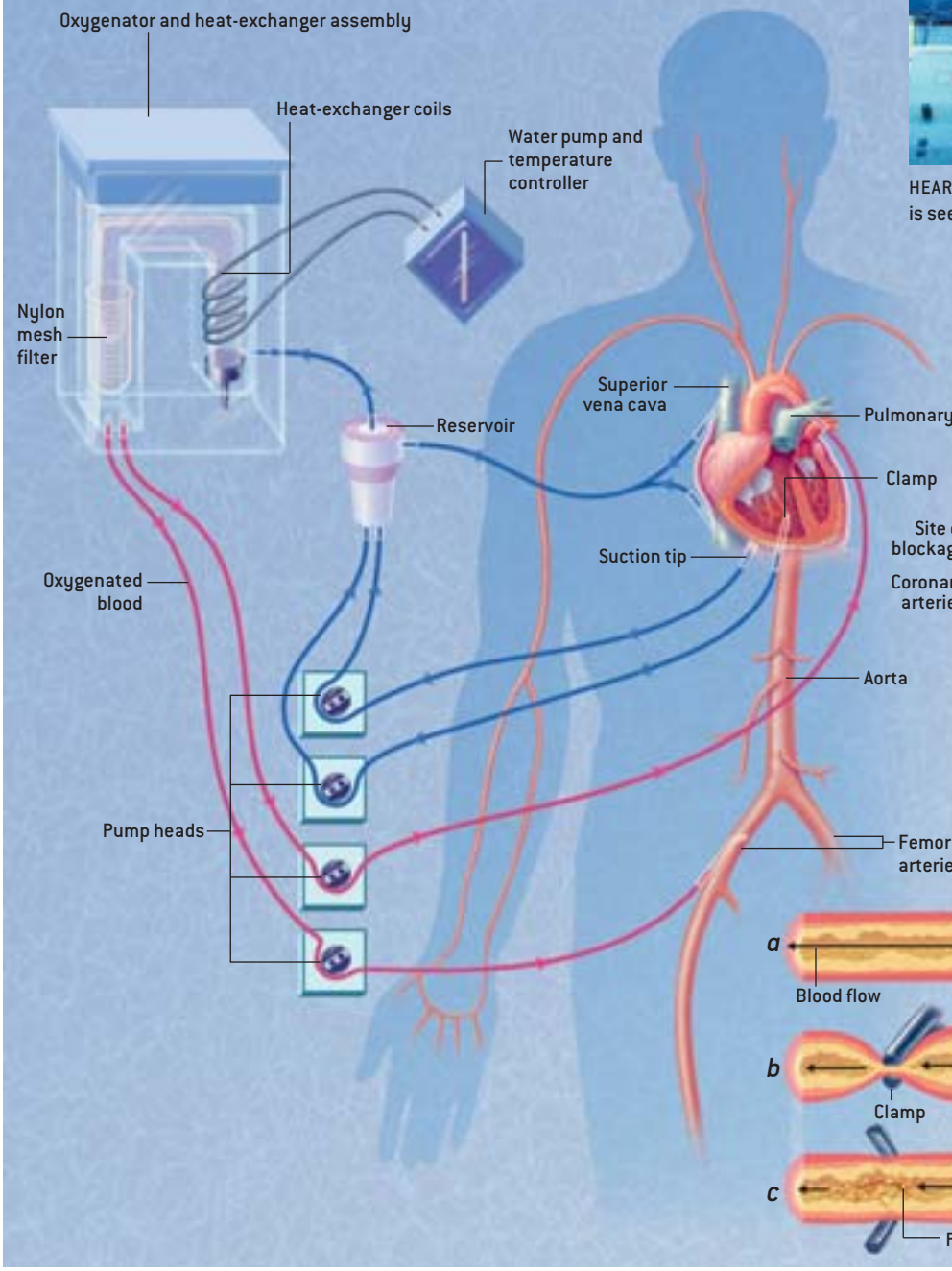
GRANT DELIN (*preceding pages*)

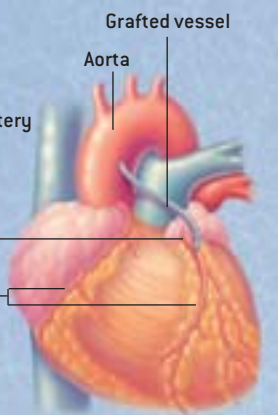
COPYRIGHT 2003 SCIENTIFIC AMERICAN, INC.

MECHANICAL HEART AND LUNGS

HEART-LUNG MACHINE circulates oxygenated blood during heart or valve surgery, typically using roller-type pumps. Blood from veins travels to a reservoir and then on to an oxygenator, where it takes up oxygen and releases carbon dioxide; a filter removes debris and bubbles. Initially a heat exchanger cools blood to protect tissue from damage. Doctors then cross-clamp the aorta and stop the heart with a cold potassium solution; the machine performs the functions of the heart and lungs during the procedure. After completion, the heat exchanger rewarms the blood, and surgeons remove the clamps and restart the heart.

HEART-LUNG MACHINE (*foreground*) is seen in action.





BYPASS OPERATION uses grafted vessels from elsewhere in the body to circumvent one (*shown*) or more blocked arteries, restoring normal blood flow.

CLAMPING OF blood vessels (*a* and *b*) or passage of blood through a heart-lung machine may create tiny particles (*c*) or bubbles that could, in turn, clog vessels and starve nearby tissue of oxygen, causing mini strokes that can contribute to temporary mental impairment in patients.

PHOTO RESEARCHERS, INC./LECA (*photograph*); ALICE Y. CHEN (*illustration*); SOURCE: ST. JUDE MEDICAL, INC.

COPYRIGHT 2003 SCIENTIFIC AMERICAN, INC.

visual retention, pairing digits with symbols, and connecting a series of numbers and letters—to 261 heart-bypass patients the week before hospitalization, during the week after the operation, after six weeks and again six months later. After surgery, 53 percent of the subjects were unable to match their earlier cognitive performance. Six weeks later, 36 percent continued to be affected, and that number dropped to 24 percent after six months.

Those results surprised no one. Short-term declines in mental function after operations involving the heart-lung machine had been reported in the literature since the beginning, and they were frequently chalked up to the general trauma of surgery. Five years after the first round of tests, however, Newman's team checked in with their subjects once more. Some of them performed about the same as they did originally, but 42 percent fared so poorly that they were again declared cognitively impaired—even after controlling for increased age.

What's to Blame?

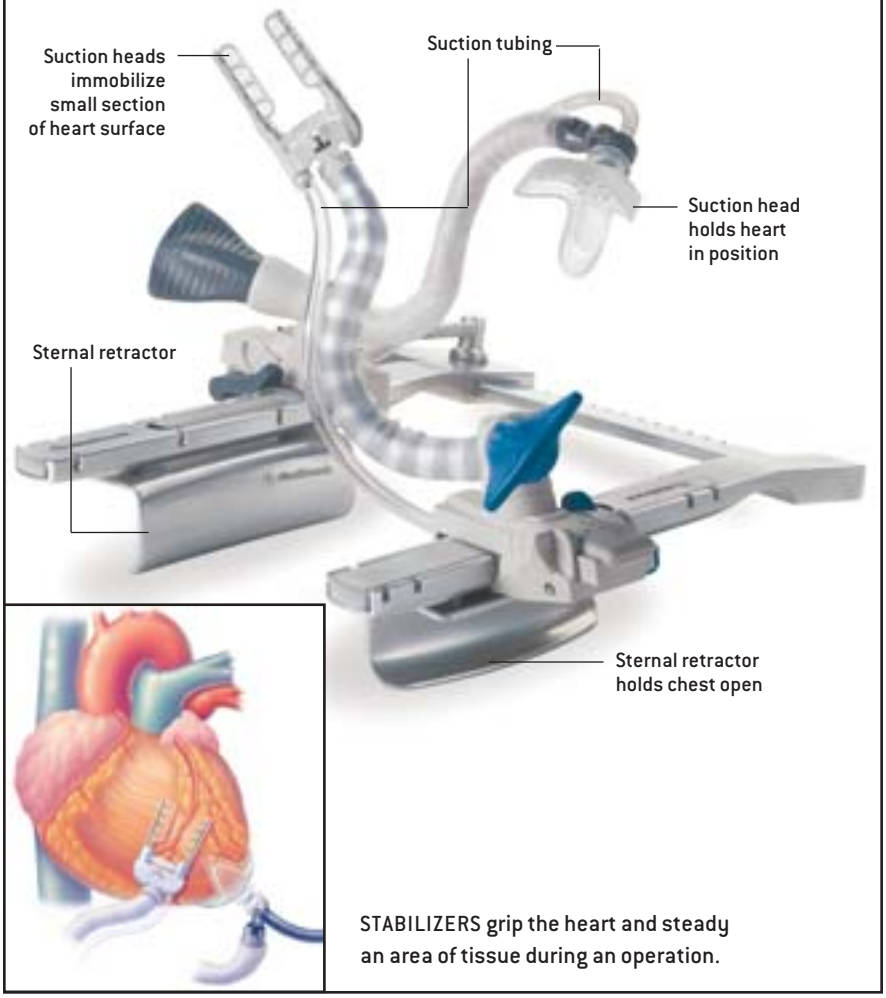
ALTHOUGH THE IDEA is not proved, the heart-lung machine is a suspect in cognitive decline for several reasons. Physicians speculate that the pump may cause damage by altering blood flow or by releasing minute debris—fat particles, blood clots, bubbles—into the patient's bloodstream. Or perhaps red blood cells can be damaged as they journey through the machine, losing their capacity to carry sufficient oxygen to the brain and the rest of the body. In Gibbon's original design—which, with refinements, is often still used—the machine pumped blood through tubing that curved around rollers attached to rotating arms. As the arms turned in eggbeater fashion, the rollers pushed the blood through the tubing. Even in today's machines, contact with tubing can damage cells, or they may be sheared or crushed by the roller pumps.

THE AUTHOR

BRUCE STUTZ is the former editor in chief of *Natural History* magazine and writes most often on science, nature and travel. Stutz is the author of *Natural Lives, Modern Times,* a natural and cultural history of the Delaware River Valley.

SURGEON'S HELPING "HAND"

OFF-PUMP SURGERY avoids the use of the heart-lung machine. Using suction, a local-tissue stabilizer (*inset*) immobilizes a section of the heart surface, permitting surgeons to work on the beating organ. Whether this procedure prevents pumphead symptoms is still unclear.



STABILIZERS grip the heart and steady an area of tissue during an operation.

In Gibbon's models, blood dripped over a wire mesh to expose it to oxygen. But this direct contact often led to too much oxygen being absorbed and resulted in oxygen toxicity. Air bubbles were also common, and they could course through the machine and cause blockages in arteries. To minimize both problems, researchers eventually developed a closer approximation of the lung: a gas-permeable synthetic membrane.

Despite these and other improvements—such as polyvinyl tubing that prevents blood cells from adhering to it, centrifugal pumps that handle cells more gently, gas exchangers that reduce bubble size, and better temperature controls—an intractable problem remains. The entire system, and the surgery itself, can still generate a variety of debris. In addition to bubbles, clotted shards of blood cells, particles of corroded tubing and arterial plaque—all collectively called emboli—can make their way through the pump and cannulae and back into the body. The workings of the pump may loosen debris; some of these materials may also be released when surgeons clamp the aorta to connect the tubing. If the resulting emboli become trapped in small vessels, they can

MEDTRONIC (*devices*); ALICE Y. CHEN (*illustration*)

COPYRIGHT 2003 SCIENTIFIC AMERICAN, INC.

block blood flow in a manner akin to a mini stroke, starving or even killing nearby tissue.

Technology has all but eliminated the largest of these emboli. Screens of woven polymer thread placed in the machine filter out particles of 0.2 to 5.0 microns in size from the blood. Sutureless connectors reduce manipulation of the aorta, curtailing specks that might otherwise enter the bloodstream. Doppler ultrasound detectors search for errant microbubbles. If they do appear, the specialist who controls the machine, called the perfusionist, can adjust the flow of blood through the gas exchanger. But microemboli—one tenth the size of the detectable ones and numbering 200 to 300 an hour—may still escape discovery and potentially damage body or brain tissue.

Newman and others believe that these types of anomalies during surgery could cause cognitive problems in patients, but some researchers remain skeptical. For instance, a five-year study of 52 bypass patients at the University of Würzburg in Germany led by Wolfgang Müllges, published last fall in *Neurology,* found essentially no "global decline" in cognitive abilities compared with the baseline (presurgery) test performances of subjects. Nevertheless, a growing concern among surgeons about the machine has led to a demand for improvements in its technology and operation as well as a search for an alternative.

The Off-Pump Option

A "BEATING HEART," or "off pump," coronary-artery bypass may seem, as one surgeon puts it, "like cutting a gemstone on horseback," but new instruments can steady the steed while the rider works [*see box on opposite page*]. The stabilizers are akin to spidery legs with padded feet that grip the heart with vacuum suction. Although off-pump surgery may not be suitable for every situation (it was not in my case), it may be an option for some patients who cannot tolerate the stress of the heart-lung machine. Off-pump surgery now accounts for anywhere from 20 to 80 percent of heart surgeries, depending on the surgeon and the hospital. [For more information about this alternative procedure, see "Operating on a Beating Heart," by Cornelius Borst; SCIENTIFIC AMERICAN, October 2000.]

Off-pump surgery appears to reduce the number of larger emboli. In a 2001 study of 40 patients by B. Jason Bowles and his colleagues at St. Francis Hospital in Honolulu, Doppler ultrasound detected an average of only 27 emboli, compared with an average of 1,766 during on-pump operations. Any cognitive benefits, however, were unclear. Other studies have also produced less than definite—or even conflicting—results.

Experts say that all previous attempts to decide which operation is better have been hindered by studies with too few participants, differing designs or goals or with lingering complexities in assessing the quality of cognition—factors that make valid conclusions problematic. In 2007 the U.S. Department of Veterans Affairs will conclude an extensive $10-million investigation comparing on-pump surgery with off-pump procedures. Led by Frederick L. Grover, chair of the department of surgery at the University of Colorado Health Sciences Center in Denver, the study is monitoring the cognitive outcomes of 2,200 heart patients in 17 V.A. hospitals.

Bearing in mind that off-pump surgery is cheaper and that patients may be ready for discharge faster, is it time to retire the heart-lung machine? No one is ready to say that. Too many other factors that occur during and immediately after heart surgery could also be participants in cognitive decline: inflammation, hypoxia (insufficient oxygen delivery to tissues), lowered blood pressure, irregular heart rhythms, or body temperature that is too warm or too cold. Hilary P. Grocott, associate professor of anesthesiology at Duke, has discovered that patients in whom fever developed within the first 24 hours following coronary-bypass surgery tended to suffer an increased probability of cognitive decline after six weeks; no one is certain whether the fever causes the dysfunction or is a symptom of it.

The overarching question is, What part of what we do is significant with regard to perioperative effects?

The overarching question, Newman says, is "What part of what we do is significant with regard to perioperative effects? If different interventions and improvements can change the early results, we think we can [positively] change the slope of [cognitive] decline." Clearly, medicine will continue to search for better answers.

In the meantime, I have found ways to cope. I've planned shorter work periods. I've walked frequently and gotten off at a subway station many blocks from my destination in order to navigate the long way back. I've played the piano to focus my mind. More than a year and a half after my valve repair, I seem to be returning to life and intelligence as I knew it. Pumpheads, take heart: I wrote this article, didn't I? SA

MORE TO EXPLORE

Longitudinal Assessment of Neurocognitive Function after Coronary-Artery Bypass Surgery. Mark F. Newman et al. in *New England Journal of Medicine,* Vol. 344, No. 6, pages 395–402; February 8, 2001.

Cognitive Outcome after Off-Pump and On-Pump Coronary Artery Bypass Graft Surgery: A Randomized Trial. Diederik Van Dijk, Erik W. L. Jansen, Ron Hijman et al. in *Journal of the American Medical Association,* Vol. 287, No. 11, pages 1405–1412; March 20, 2002.

Protecting the Brain in Coronary Artery Bypass Graft Surgery. Daniel B. Mark and Mark F. Newman in *Journal of the American Medical Association,* Vol. 287, No. 11, pages 1448–1450; March 20, 2002.

COPYRIGHT 2003 SCIENTIFIC AMERICAN, INC.

Terms of Engagement

Irving Weissman directs a new institute dedicated to the cloning of human embryonic stem cells. Just don't call it cloning By SALLY LEHRMAN

In a human biology classroom this past March, stem cell biologist Irving Weissman described his Jewish grandparents' flight to the U.S. for religious freedom. He turned to his host, William Hurlbut, a fellow Stanford University professor and member of President George W. Bush's Council on Bioethics, and accused him of attempting to tie American public policy to his own religious beliefs. A student raised his hand and asked, "But once you get rid of religion, do you have any guide other than popular opinion?"

OLIVIER LAUDE

IRVING WEISSMAN: PUSHING STEM CELLS

- **A prodigy who had published two research papers by the time he finished high school in Montana; maintains a home there, where he fly-fishes.**
- **Co-founded three companies that develop stem cells to battle such conditions as cancer, neurodegenerative disorders and diabetes.**
- **Headed the National Academy of Sciences panel on human reproductive cloning; testified before Congress to support embryonic stem cell research.**

The Hippocratic oath, answered Weissman, who graduated from Stanford's medical school. "You shall not as a doctor allow any of your personal ethical, religious, even moral concerns stand between you and care of the patient," he paraphrased. "I interpolate this to mean not only the patient you might treat but future patients that might be helped by your research."

Weissman, director of Stanford's new Institute for Cancer/Stem Cell Biology and Medicine, has relied on Hippocrates and a careful definition of terms as he navigates the controversy surrounding human embryonic stem cell research. Already known on Capitol Hill for his defense of free scientific inquiry, Weissman found himself a target for attack when he declared that the privately funded Stanford institute would explore human nuclear transplantation—commonly known as cloning.

Weissman's institute, the first of its kind at a university, will not create human clones; rather it will generate new stem cell lines for research. At first, Stanford biologists would most likely derive these from existing lines and use mouse embryos to study ways to "reprogram" transplanted nuclei so that they behave like embryonic nuclei. But a $12-million anonymous donation frees the institute to move into cloning as quickly as it likes. Private funds are also helping sizable programs at the Burnham Institute in La Jolla, Calif., and at the University of California at San Francisco to sidestep the federal funding ban on work with stem cell lines created after August 9, 2001. California has legislation that will provide additional funding.

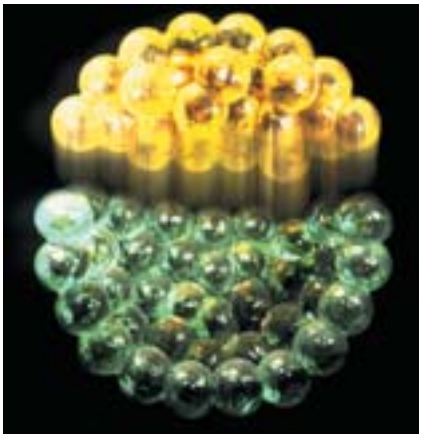
Despite such backing, scientists here and across the country sit uncomfortably at the focal point for religious and moral qualms about the direction and power of bioscience. This past February the House passed a bill that would impose a $1-million penalty and 10 years in jail for any human embryonic cloning. Several states, in-

COPYRIGHT 2003 SCIENTIFIC AMERICAN, INC.

cluding Iowa and Michigan, have enacted bans. Fearing criminalization of their work, some biologists devote a third of their time to lobbying and education. Bioethicist-physicians and the American Association for the Advancement of Science (AAAS) have begun compiling suggestions for self-regulation.

Weissman places his faith in scientists' ability to one day clearly demonstrate the technology's promise. He plans to forge ahead by hiring an expert in nuclear reprogramming. He also wants to explore the possibility of retrieving viable oocytes from oophorectomies instead of using donated eggs. Eventually the institute would combine these two initiatives. But do not consider the transfer of cell nuclei to be "cloning," Weissman insists. Moreover, he states, an "embryo" is not the result.

Weissman hopes to prod the public toward a language that steers clear of dire imagery such as cloned fetuses grown for spare parts. "Whenever we asked people, even scientists, to draw an embryo, they'd usually draw a fetus with legs, head, and so on," he explains. Referring to a National Academy of Sciences cloning panel he chaired, he adds: "Nobody drew a ball of 150 cells." Redefining the terms of the debate won't win Americans over, he thinks. But vague language contributes to a merging of stem cell research with concerns about manipulating human life. "As soon as they start using catchphrases that don't describe what's going on, it's easier for people to say we're cloning human beings," Weissman says. "You're always going to pay if you accept language that is incorrect."



"EMBRYO" does not usually evoke a cell ball. The artwork shows stem cells (*yellow*) and trophoblast cells (*green*), which attach to the uterus.

Although two NAS reports press the same point, some researchers worry that Weissman's parsing may be counterproductive. Investigators have toyed with terms over the years, even floating the short-lived "clonote." But "to change the vocabulary now at this point for public consumption—it would be taken badly," remarks John Gearhart, who led the Johns Hopkins University team that isolated human pluripotent stem cells. Scientists are losing the debate, he fears: " 'Cloning' is used; 'embryo' is used. Let's try to deal with that and not avoid it."

Indeed, terminology has become a touchstone for both sides. When the AAAS brought together 33 experts in early March to begin discussing a regulatory scheme, the gulf between advocates and opponents included an inability to agree on the definitions of "blastocyst," "embryo" and "viability." When Weissman announced the institute's plans last December, Leon Kass, chair of the president's bioethics council, attacked his words as "artful redefinition." He reminded Weissman that a majority of his committee (including Stanford's Hurlbut) had called for a four-year moratorium on biomedical cloning. Kass went on to chastise the university for attempting to obfuscate the nature of its work and trying to race ahead without public scrutiny.

Nigel Cameron, director of the Council for Biotechnology Policy, which is affiliated with the Christian think tank Wilberforce Forum, suggests that Weissman and others are digging in their heels because they fear a trend of regulation and resistance. Biology's rebellion against public unease, Cameron predicts, will sow distrust for decades to come. "We now have a major university willing to give its name to something many people regard as unethical science," Cameron says. "There seems to be no regard for public conscience."

The 63-year-old Weissman says it would be unethical *not* to proceed. He tells those who would halt embryonic stem cell research that they are responsible for the lives that could have been saved by future therapies. Based on his 20-plus years with stem cells—he was the first to isolate mouse blood-forming stem cells, then those of humans—he is adamant that stem cells from adults cannot evolve into a variety of tissues, as those from embryos can. Further, he points out, the existing embryonic lines came from white, well-to-do, infertile couples and so have limited genetic diversity and utility.

The institute will not try to create stem cell lines immediately; it will initially explore the machinery that tumor cells share with stem cells. "I believe nuclear transfer for human stem cells is five years away, minimum," Weissman says. In fact, the first American university to transfer the nuclei of human cells may be U.C.S.F., which is building a $10-million graduate training program in stem cell research.

Although Weissman tells students that the potential for scientific advances must sometimes outweigh individual moral, ethical or religious beliefs, he doesn't brush aside social responsibility. At critical junctures he consults with Stanford's bioethicists—most recently, before trying to grow human neurons in mice. He hopes to study learning, memory and diseases such as schizophrenia but wants to stop short of creating a mouse with human characteristics in its brain.

Weissman thinks that reasoned debate will never be enough to settle the stem cell dispute. The general public doesn't have a desire to delve into the issue, he argues, and the Bush administration has religious reasons to remain unconvinced. Words such as "cloning" and "embryo" bolster a hidden quest to redefine human life as beginning at the moment of nuclear transfer or of fertilization, according to Weissman. This is a position that neither a scientist nor a religious person can argue, he says: "You'd have to have an assay for a human soul." SA

Sally Lehrman is based in San Francisco.

LAGUNA DESIGN/SCIENCE PHOTO LIBRARY

COPYRIGHT 2003 SCIENTIFIC AMERICAN, INC.

SCANNING ELECTRON MICROSCOPES

Fine Focus

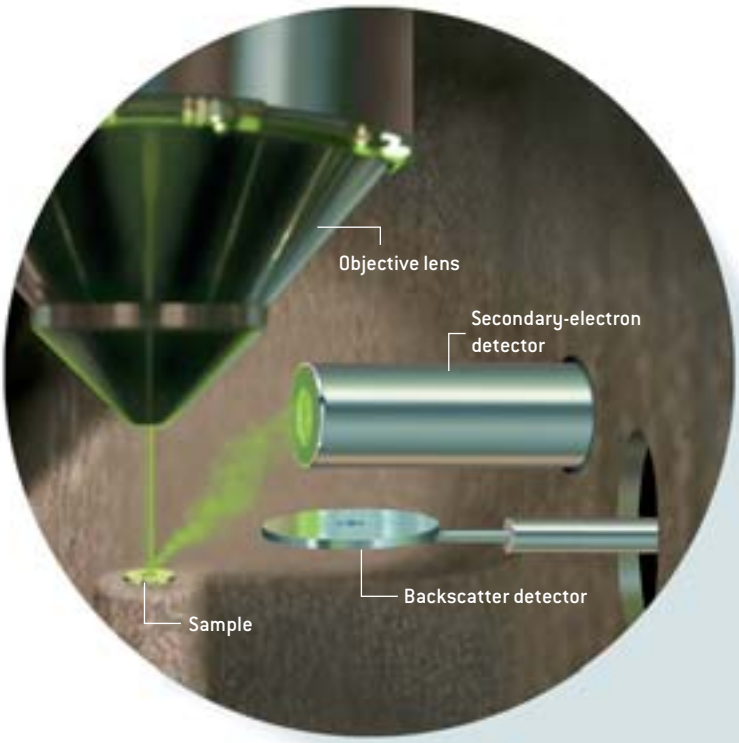
The image is striking: a tiny insect, magnified 10,000-fold, looms like a giant alien from a science-fiction movie, with terrifying talons and mandibles ready to devour the world. Thank the scanning electron microscope for the fascinating impression.

Scanning electron microscopes, or SEMs, enlarge objects up to one million times. But what makes their images so captivating is that they accurately portray the object's surface with resolutions finer than 100 angstroms. They can also indicate which atomic elements make up the sample.

Although the pictures of bugs and other organisms are most familiar, "SEMs are incredibly versatile," says Scott Chumbley, professor of materials science and engineering at Iowa State University. Industrial engineers, he notes, use SEMs to learn why materials fail—say, whether a metal shears because of fatigue, corrosion or tensile stresses. Jet-engine makers analyze particles trapped in oil filters to determine which engine parts show excessive wear. By examining cells, medical researchers can tell how a bone is degrading or if a bacterium has attacked tissue. And forensic scientists can ascertain whether hair, clothing fibers or gunshot residue found in different locations came from the same source.

A standard scanning electron microscope that can perform such functions costs $150,000 to $300,000. The sample is dried and coated with a layer of conductive atoms, so negative charge does not build up as a high-energy electron beam scans the item. Advanced instruments, such as low-pressure SEMs or environmental SEMs, can image samples that are uncoated or wet, but they cost in the neighborhood of $1 million when fully equipped. One such product, Quanta, from FEI in Hillsboro, Ore., allows the operator to moisten, heat, stretch or compress the sample.

SEMs don't just provide great utility. Artists are using them to depict common and not so common objects, colorizing the naturally black-and-white images to enhance their allure. Frames of skin cells, molds or minerals may soon appear in books and on museum walls beside the photographs and paintings of traditional artists. *—Mark Fischetti*



KENT SNODGRASS *Precision Graphics*; MICROANGELA/TINA CARVALHO *Biological EM Facility, University of Hawaii (spider hair, black ant)*

AN OBJECT TO BE IMAGED is dried and coated with conductive atoms, typically gold, that attract an electron beam. The beam interacts with the electron cloud of the sample, giving off low-energy electrons. A charged, secondary-electron detector immediately attracts all the low-energy electrons inside its tube, millions of times a second. Each capture corresponds to a pixel on the object's surface. A cascade multiplier amplifies the electrons, which a computer reads as current. Alternatively, a sample's elements can be determined by moving the backscatter detector under the lens. The beam passes through a hole and through the sample. Different atomic nuclei bend the electrons' paths to various degrees, and some hit the detector's underside, creating a telltale signature.

COPYRIGHT 2003 SCIENTIFIC AMERICAN, INC.

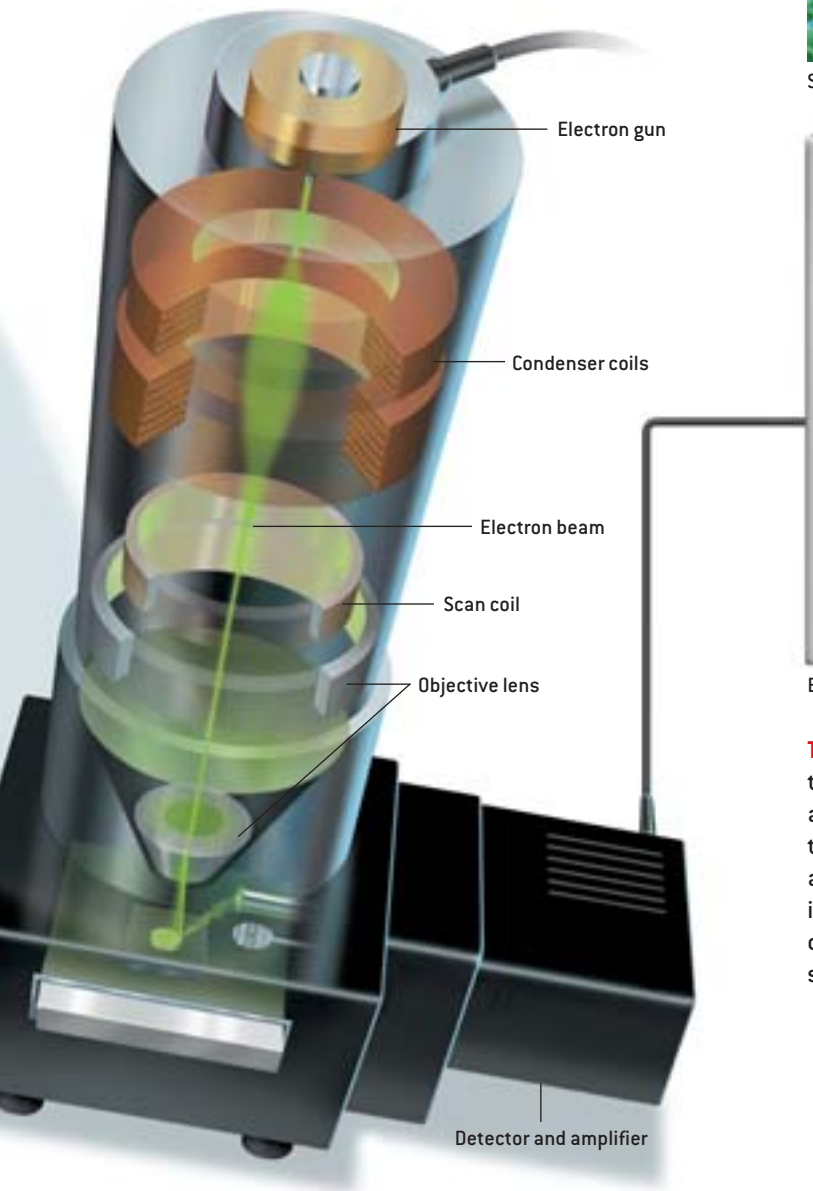
DID YOU KNOW . . .

➤ IN, NOT ON: Rather than strike the surface of an object with electrons to show its topography, a transmission electron microscope sends electrons through it, to show its innards. TEMs can resolve features as small as one angstrom. The catch is that the sample must be prepared as a slice of material no more than 1,000 angstroms thick. So the instrument cannot show a magnified image of a mosquito, but it can reveal a virus hiding inside one of the insect's cells.

➤ K–12: Elementary and high school classes can apply to use Iowa State University's SEM online. Under its Excel program (www.mse.iastate.edu/excel), funded by the National Science Foundation, the materials science and engineering department will prepare a sample sent by the class and place it in an SEM. At a prearranged time, the class connects to the Web server, controls the instrument and sees the image. Students can move the sample; change the magnification, focus, contrast and brightness; and obtain chemical-composition data. Favorite objects include paper, pencil lead, sugar, leaves, bugs, dirt and candy.

➤ MICRO SANDBLASTER: Focused-ion-beam machines operate similarly to SEMs, but they direct a beam of ions, such as gallium, onto a sample. They can create a surface image but can also etch away material like a sandblaster. FIBs are often used to find and repair flaws on integrated circuits and to mill other microstructures.

AN ELECTRON GUN EMITS the electron beam. Wire condenser coils collimate and tighten the beam, and an objective lens sharply focuses it. Scan coils create a varying magnetic field that sweeps the beam across the sample in a raster-scan pattern.




Spider hair


Black ant


Blood cells

THE COMPUTER TRANSLATES the varying signal from pixel to pixel of a surface scan into a corresponding image on a cathode-ray tube. A high spot on the sample is designated as white on the screen; a low spot appears dark. If the surface spot is angled toward the detector, it appears slightly brighter; if it is angled away, it appears slightly duller. The human brain interprets these values as shine and shading, just as it interprets glare or shadow amid sunshine.

This month's topic was suggested by reader Keith McVicker. Have an idea for a future column? Send it to workingknowledge@sciam.com

COPYRIGHT 2003 SCIENTIFIC AMERICAN, INC.

Through a Glass Deeply

WALKING BENEATH THE WAVES AT THE MONTEREY BAY AQUARIUM **BY MARGUERITE HOLLOWAY**

The southern lip of Monterey Bay—below which opens the deepest submarine canyon on the West Coast—was for three decades home to the world's largest sardine fishery, a lucrative industry that petered out in the early 1950s when the fish disappeared. Scientists reported recently that overfishing most likely did not run the foot-long fish out of the bay, cooler waters did, but no matter. The cannery operations are long gone, and the California town of Monterey has an entirely new industry, one also based on the ocean's wealth.

The Monterey Bay Aquarium is housed in a converted cannery, just on the water's rocky edge. Its design, which won an architectural award in 1988, makes visitors feel as though they are in a liminal state: poised between land and water, able to live equally well in both. In the darkness of the Drifters Gallery, for example, undulating, crenulated, luminous jellies are inches away, hundreds of them: from large sea nettles that glow orange against a royal blue background to spinning, tiny sea gooseberries that seem to have internal electric filaments lighting their transparent bodies. It is mesmerizing to sit as if underwater and watch dozens of jellies slowly pulse and rise and fall.

The 17 species of jellies are among the aquarium's principal attractions, in part because researchers there have pioneered the cultivation of certain species to keep the displays well stocked. (One of management's goals, says spokesperson Karen Jeffries, is to raise as many varieties as possible so that collectors don't have to harvest much from the bay itself; thus far 51 kinds of fish, invertebrates, corals and jellies are husbanded in house—about 10 percent of the ones on display.) In collaboration with a Japanese company, Monterey Bay Aquarium and other researchers have developed a special tank—called a kreisel—that is round so the jellies don't get stuck in the corners, where they would die. Jellies have become more popular in aquariums generally because of these two advances.

The Drifters Gallery is part of the Outer Bay Wing, which opened in 1996. The section displays the life of the bay dozens of miles offshore. Pelagic creatures, including ocean sunfish, oceanic whitetip shark, stingray, sea turtle, soupfin shark, California barracuda and yellowfin tuna, swim in a two-story

HOVDEN CANNERY, now home of the Monterey Bay Aquarium (*top*), was one of many sardine processing plants along Cannery Row—the life of which was described in John Steinbeck's 1945 novel. Visitors can see what a local wharf looks like from below the surface (*left*).

ROB LEWINE (*bubble window*) AND RICK BROWN (*inset*) *Monterey Bay Aquarium*

COPYRIGHT 2003 SCIENTIFIC AMERICAN, INC.

SEA NETTLES and other jellies are often hard to maintain, but the aquarium has developed tanks that keep the fragile creatures moving (*above*). The Kelp Forest (*left*) is replenished by bay water every day and is open to the air and sunlight.

tank—contained by a 78,000-pound, 13-inch-thick window. On the lower level of the wing, a Mysteries of the Deep exhibit shows what is happening farther down, along the walls, midwaters and floor of the two-mile-deep, 60-mile-long Monterey Bay Canyon. Some of the creatures, such as the vampire squid, arrive here by video only—they are not adapted for life at less pressurized climes. Others are here in their full, weird glory. A thin tunicate worm opens a hood three times its width to catch diatoms and additional tiny prey. It looks like a baby bird: all mouth.

Opposite the Outer Bay Wing—with a two-story otter tank in between—is the other side of the aquarium, the Nearshore Wing. This three-floor section opened in 1984 and is centered on a 28-foot-high, 335,000-gallon Kelp Forest, the first display of its kind. Twice a day a diver descends into the tank to feed the wolf-eel, leopard shark, sheephead, rockfish, surfperch and famous Pacific sardine, among many other creatures. The feedings are raucous, docent-moderated events: children throng near the tall windows, looking up at the sunlight on the water above them, and the diver answers questions from inside the tank. On a recent morning, one of the most pressing concerns seemed to be why the sharks don't eat everything else. "Room service," quips the diver through her underwater microphone as she shakes the rest of the chopped squid from her bag.

The Nearshore Wing is organized along habitat type—deep reefs, shale reefs, rocky shore, sandy shore, sandy seafloor, coastal stream and so forth. A 326,000-gallon tank houses the nearshore denizens of Monterey Bay: the bat rays, halibut, salmon, white sturgeon and sevengill sharks that live among the pilings and piers of area harbors. The connection of this wing to the adjacent living ocean is very marked. Several touch tanks are scattered throughout, with starfish and anemones and kelp. On the top floor, visitors can see the machine that generates waves for the Kelp Forest. Water from the bay is piped in and pumped throughout the displays every day—and at night the Kelp Forest water is not filtered, so drifting larvae can settle inside if they so choose.

This wing contains one of the aquarium's loveliest exhibits, the aviary. Just a few feet away from viewers—and with no glass partition—stand shorebirds collected from wildlife rescue centers, some of which are released when they become healthy. Every detail of these usually skittish birds is clear and close: the colors and markings and bills of black oystercatchers, ruddy turnstones, willets, sanderlings, long-billed curlews, avocets, and red and red-necked phalaropes.

The otters in their display tank were also rescued, either as stranded pups or ill adults. Because the population of California sea otters is so low—there are only about 2,100, down from hundreds of thousands before the fur trade of the mid-1700s—the aquarium runs a sea otter research and conservation program. Indeed, Monterey Bay is thick with research and conservation outfits: the Monterey Bay National Marine Sanctuary (www.mbnms.nos.noaa.gov), Moss Landing Marine Laboratories (www.mlml.calstate.edu), the Monterey Bay Aquarium Research Institute (www.mbari.org) and Stanford University's Hopkins Marine Station (http://marine.stanford.edu).

The Monterey Bay Aquarium is located at 886 Cannery Row and is open every day except Christmas. Winter hours are 10 A.M. to 6 P.M., and in the summer (the end of May until early September) it opens at 9:30 A.M. Some 33.5 million people have visited since the aquarium opened—an average of 1.7 million a year—and lines, even in the morning, are often long. If you can, order tickets ($17.95 for adults) in advance: 831-648-4937. For further information, including driving directions from San Francisco or Los Angeles, look at the Web site (www.mbayaq.org) or call 831-648-4888. SA

RANDY WILDER *Monterey Bay Aquarium* (*top*); MONTEREY BAY AQUARIUM (*bottom*)

COPYRIGHT 2003 SCIENTIFIC AMERICAN, INC.

Drenched in Symbolism

A DAZZLING RECORD OF PREHISTORIC CARVINGS AND PAINTINGS TESTIFIES TO THE COGNITIVE COMPLEXITY OF OUR SPECIES **BY IAN TATTERSALL**

PREHISTORIC ART: THE SYMBOLIC JOURNEY OF HUMANKIND
by Randall White
Harry N. Abrams, New York, 2003 ($45)

TWO HEADLESS IBEXES embracing or fighting. Spear-thrower sculpted of reindeer antler, 9 by 7 centimeters, Grotte d'Enlène, France, roughly 14,000 years old.

COURTESY OF HARRY N. ABRAMS, INC.

About 40,000 years ago the first *Homo sapiens*—the Cro-Magnons—began to trickle into Europe, displacing the resident Neanderthals in the process. The contrast between the records of their lives that these very different hominids left behind could hardly be more striking. For no extinct human species, not even the large-brained *Homo neanderthalensis*, has bequeathed us evidence of a complex symbolic existence, based on the extraordinary cognitive capacities that distinguish us from all other living species today. In contrast, the lives of the Cro-Magnons were drenched in symbolism.

Well over 30,000 years ago these early people were creating astonishing art on the walls of caves. They crafted subtle and beautiful carvings and engravings and kept records by incising intricate notations on bone plaques. They made music on bone flutes, and if they did this, they surely sang and danced as well. They ornamented their bodies and buried their dead with elaborate grave goods, presumably to serve them in an afterlife. Technologically, a cascade of innovations included nets, textiles and ropes, even the first ceramics. In short, those Cro-Magnons were *us:* members of a species whose relationship with the rest of the world was totally unprecedented in the entire history of life.

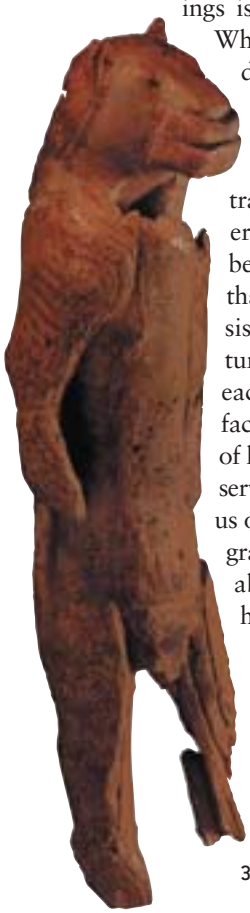
For a couple of decades now, New York University archaeologist Randy White has been a leading investigator of how the expression of the unique human capacity unfolded in Europe during the two dozen millennia that followed the arrival of the Cro-Magnons. In this thoughtful and very beautiful book, White concentrates on the most dazzling part of this record, the part that embraces what we would call art—and that includes some of the most powerful ever made. But he is careful to point out that "art" is very much a Western concept and that for its creators, what looks to us like art probably had implications vastly different from those we impute in our own society to art and decoration. For while nobody could doubt that Cro-Magnon symbolic production somehow reflected these people's conceptions of their place in the natural world, the Cro-Magnons were hunters and gatherers, with a perceived relationship to nature that must have been radically different from our own.

For this reason, White eschews the elaborate explanations that so many authors feel somehow obliged to bring to the interpretation of prehistoric art and hews to the facts. He begins with a brief history of the discovery and interpreta-

COPYRIGHT 2003 SCIENTIFIC AMERICAN, INC.

tion of Cro-Magnon art, as prelude to a largely chronological account of the evidence for symbolic expression in Europe and parts of northern Asia between about 40,000 and 10,000 years ago. In these sections, White mostly avoids stylistic analysis in favor of a focus on techniques, but he manages to address, if usually briefly, most of the major questions that Cro-Magnon art elicits.

As perhaps befits a work that grew out of a university survey course, this volume extends beyond mainly European Ice Age art to consider prehistoric symbolic and representational traditions (some earlier, others quite recent) in Africa, southern and western Asia, Australia, and the Americas. Each of these regional groupings is treated separately, and White wisely refrains from drawing close parallels between different regional traditions. Of course, including all these diverse traditions between the covers of a single book might be taken to imply a unity that contradicts White's insistence on the unique cultural roots and referents of each one of them. But the fact that all are the products of hunting-gathering peoples serves very usefully to remind us of the vast range of iconographies and aesthetics available even to noncomplex human societies.



HALF HUMAN, half lion. Mammoth ivory, 28.1 centimeters, Hohlenstein-Stadel, Germany, about 32,000 years old.

What White's spectacularly illustrated book does most clearly, then, is to bring home the astonishing diversity and intricacy of the representational traditions that the extraordinary human symbolic spirit has from the beginning produced worldwide, even in the absence of complex social and economic structures. The remarkable human cognitive capacity that early art reflects appeared quite recently, perhaps less than 100,000 years ago. And that appearance set our species on a course of accelerating technological change and elaboration that may yet run out of our control. But White shows that although our economic lives have changed out of recognition in that time, the potential that underwrites our modern lifestyles and achievements was there from the very start. Deep down, human beings haven't changed one whit since prehistoric times. SA

Ian Tattersall is a curator of physical anthropology at the American Museum of Natural History in New York City. His most recent book is The Monkey in the Mirror *(Harcourt, 2002).*

THE EDITORS **RECOMMEND**

IMAGINING NUMBERS (PARTICULARLY THE SQUARE ROOT OF MINUS FIFTEEN)

by Barry Mazur. Farrar, Straus and Giroux, New York, 2003 ($22)

Mazur, a mathematician and university professor at Harvard University, writes "for people who have no training in mathematics and who may not have actively thought about mathematics since high school, or even during it, but who may wish to experience an act of mathematical imagining and to consider how such an experience compares with the imaginative work involved in reading and understanding a phrase in a poem." It is a stimulating and challenging journey, one likely to lead the reader to share Mazur's view: "The great glory of mathematics is its durative nature; that it is one of humankind's longest conversations; that it never finishes by answering some questions and taking a bow. Rather, mathematics views its most cherished answers only as springboards to deeper questions."

NATURE VIA NURTURE: GENES, EXPERIENCE, AND WHAT MAKES US HUMAN

by Matt Ridley. HarperCollins, New York, 2003 ($25.95)

Ridley presents a history of the long debate over genes versus the environment as the dominant influence on human behavior. He asserts that "versus" is wrong. His point of departure is the recent identification of the full sequence of the human genome. "The discovery of how genes actually influence human behaviour, and how human behaviour influences genes, is about to recast the debate entirely. No longer is it nature-versus-nurture, but nature-via-nurture. Genes are designed to take their cues from nurture.... My argument in a nutshell is this: the more we lift the lid on the genome, the more vulnerable to experience genes appear to be." Ridley is a skillful writer who holds the reader's attention all the way.

All the books reviewed are available for purchase through www.sciam.com

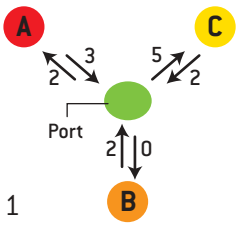
COPYRIGHT 2003 SCIENTIFIC AMERICAN, INC.

High Spies

BY DENNIS E. SHASHA

You have been approached by a spy agency to determine the amount of contraband goods that are being traded among several nefarious countries. After being shipped from its country of origin, each container of goods is routed through at least one neutral port. At the port, the containers are mixed up in a warehouse before being sent on their way, so you cannot trace individual containers from their country of origin to their final destination. But satellite cameras can tell you the number of containers traveling in each direction on each leg of the journey. You also know that every container takes the shortest possible route to its destination.

As a warm-up problem, consider illustration 1 (*right*), which shows the number of containers traveling to and from countries A, B and C and a neutral, centrally located port. Because no containers are observed moving from the neutral port to country B, we surmise that the two containers from country C must have moved on to country A (they cannot turn back to their country of origin). And because only two containers are seen traveling from the port to country A, the two containers from country B must have gone to country C, where they joined the three containers from country A.



1

Now consider illustration 2 (*below*), which shows five countries and three neutral ports. Can you find the minimum and maximum numbers of containers that travel from country A to country D? How about from country A to country E? The answer changes if we revise the diagram as shown in illustration 3. What are the minimum and maximum numbers now? SA

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

Answer to Last Month's Puzzle

An omnidextrous prime 3-square that uses only three different digits:

3 1 1
1 8 1
1 1 3

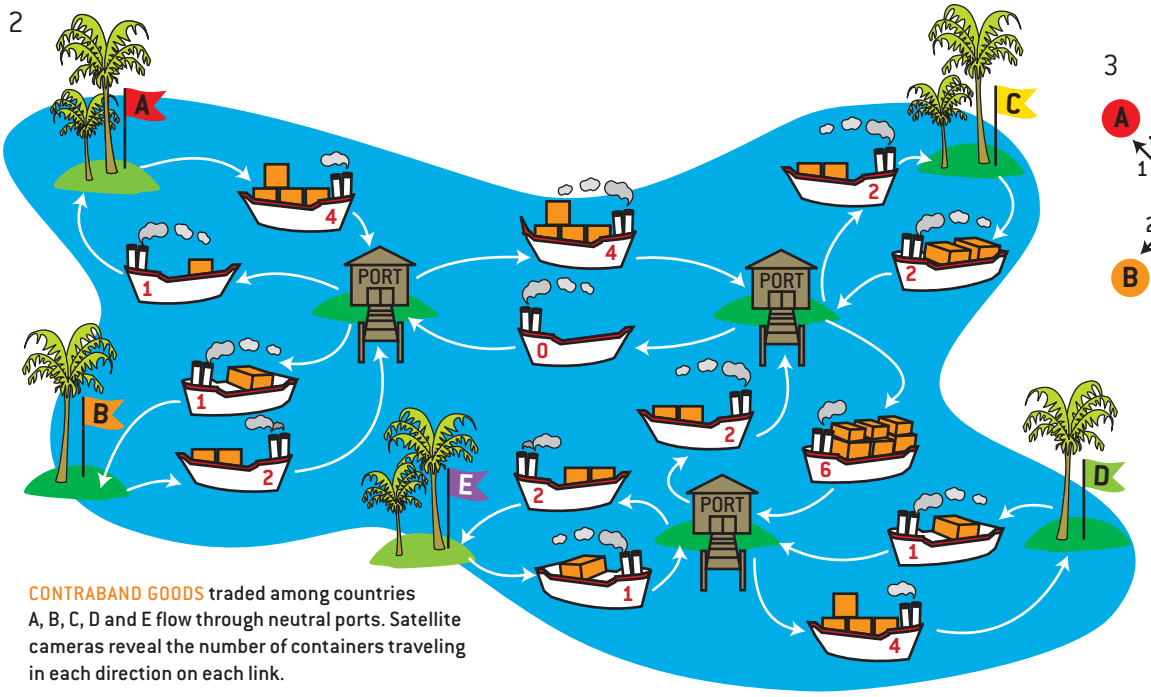
A prime 5-square that uses all 10 digits:

1 6 4 5 1
4 5 3 8 9
9 2 8 9 3
2 8 0 6 9
9 7 1 7 1

Web Solution

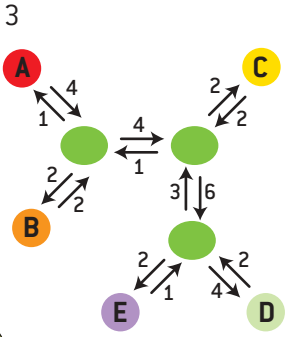
For a peek at the answer to this month's problem, visit www.sciam.com

2



CONTRABAND GOODS traded among countries A, B, C, D and E flow through neutral ports. Satellite cameras reveal the number of containers traveling in each direction on each link.

3



CORNELIA BLIK

COPYRIGHT 2003 SCIENTIFIC AMERICAN, INC.

The Yanked Clippers

HAVE SOME SECURITY MEASURES BECOME MORONIC, OR IS IT JUST ME? BY STEVE MIRSKY

It was my nagging fear of inadvertently carrying nail clippers while trying to pass through airport security, as well as my long-standing interest in stupidity, that prompted me to go hear Simon Davies present the first Stupid Security Awards at the Conference on Computers, Freedom & Privacy in April here in New York City. Davies, director of Privacy International, a nonprofit organization based in London and Washington, D.C., went through almost 5,000 nominations from people telling tales of unnecessary intrusion and harassment in the service of the *illusion* of security—I mean, seriously, can't sharp fingernails be as lethal as a set of nail clippers?

I also went because overzealous and counterproductive security is a science issue. In December 2002 the presidents of the National Academy of Sciences, the National Academy of Engineering and the Institute of Medicine issued a joint statement complaining that "outstanding young scientists, engineers, and health researchers have been prevented from or delayed in entering this country."

In February an online article by Peg Brickley in the *Scientist Daily News* pointed out that visa delays had prevented the return of 15 Cornell University students who had gone home for winter break, resulting in "research left half-done." Like it isn't already tough enough to get people to go to Ithaca in the middle of winter. (By mid-May most stranded students were back, but Cornell expected fresh problems in the next semester.)

In addition, the article outlined the case of two Bangladeshi physicians, one Hindu and one Muslim, who were being brought to Ithaca by Cornell researcher Kathleen Rasmussen through a training grant from the National Institutes of Health's Fogarty International Center, which "promotes and supports scientific research and training internationally to reduce disparities in global health," according to its mission statement. The Hindu physician had no difficulty. The Muslim's visa was repeatedly denied, finally approved but then ultimately left unprocessed. "From the point of view of someone trying to run a government grant to meet government purposes with government money," Rasmussen said, "to have the left and right hands working against each other is very disturbing." Her statement is quite true and also happens to explain the average person's aversion to accordion music.

The Stupid Security Awards, on the other other hand, dealt with more mundane examples of people being annoyed and inconvenienced by dubious security measures. (The kind, for instance, that we at Scientific American deal with: after September 11, our building installed high-tech turnstiles in the lobby. These turnstiles can be negotiated only by someone brandishing an electronic ID card or with the ability to jump over a metal bar about three feet off the floor.)

Oddly, the award citations that Davies read at the conference do not match those listed on the Privacy International Web site (www.privacyinternational.org). Maybe some kind of Orwellian historical alteration took place. Nevertheless, I recommend visiting the site for one version of the list of winners and for links to dozens of nominations, two of which I'll recount.

My favorite tale of "whoa" came from an airline pilot with clipper issues of his own. "I am searched as rigorously as any passenger," he wrote. "I am forbidden to have any nail clippers." The pilot then noted that "once on board the aircraft, securely locked away on the flight deck, I have an ax behind me."

A close second was submitted by a guy whose story starts as he is about to board a plane in San Francisco. "The polite inspector informed me that he had to check my shoes for explosives. I dutifully removed them and handed them to him. He picked them up one by one and slammed them down on the floor with full force. Apparently, as they hadn't exploded, they were not dangerous, and he handed them back to me." Perhaps it's best to look on the bright side and simply applaud any public display of the scientific method. SA

ILLUSTRATION BY MATT COLLINS; PHOTOGRAPH BY FRANK VERONSKY

COPYRIGHT 2003 SCIENTIFIC AMERICAN, INC.

Why does reading in a moving car cause motion sickness?

Timothy C. Hain, professor of neurology, otolaryngology and physical therapy/human movement science at Northwestern University Medical School, and Charles M. Oman, director of the Man Vehicle Laboratory at the M.I.T. Center for Space Research and leader of the neurovestibular research program at the NASA National Space Biomedical Research Institute in Houston, explain:

Motion sickness—whether caused by traveling in a car or a boat or being in outer space—is the unpleasant consequence of disagreement between the brain's expectations in a given situation and the information it receives from the senses.

To retain balance, the brain synthesizes data from many sources, including sight, touch and the inner ear. The last is particularly important because it detects angular and linear motion. Most of the time, all the inputs agree. When they do not jibe with what the brain expects in that situation, however, motion sickness—with its spatial disorientation, nausea or vomiting—can occur.

Imagine that you are reading in a car's backseat. Your eyes, fixed on the book with the peripheral vision seeing the interior of the car, tell the brain that you are still. But as the car changes speed or turns, the sensors in your inner ear contradict that information. This is why motion sickness is common in this case. It helps to look out the window. (The driver suffers least, because he not only has compatible sensory information but is also controlling the car—so he is prepared for variations in motion.)

Likewise, you can combat seasickness by staying on deck, where you can see the horizon. Once your balance system learns to handle the boat's motion—when you get your "sea legs"—susceptibility to illness fades. Of course, when you go ashore, your body may still anticipate the boat's movement for a few hours or even days, which can make you feel unwell again.

Spaceflight also causes motion sickness, suffered by 70 percent of rookie astronauts. In "weightless," or microgravity, conditions the inner ear cannot determine "down." Some crew members have said they felt as if they were upside down continuously, no matter what their actual orientation.

How long do stars usually live?

—A. Tate, Willard, Mo.

John Graham, an astronomer in the department of terrestrial magnetism at the Carnegie Institution of Washington, answers:

Stars' lifetimes vary from a few million years to billions of years. It depends on how fast a star uses up its nuclear fuel. Almost all stars shine as a result of the nuclear fusion of hydrogen into helium. This process takes place within their hot, dense cores, where temperatures may reach 20 million degrees Celsius. The star's rate of energy generation depends on both temperature and the gravitational compression from its outer layers. More massive stars burn their fuel much faster and shine more brightly than less massive ones. Some large stars will exhaust their available hydrogen within a few million years. On the other hand, the least massive ones that we know are so parsimonious that they can continue to burn longer than the current age of the universe itself—about 15 billion years.

Our sun has been around for nearly five billion years and has enough fuel for another five billion. Almost all the stars we can see in the night sky are intrinsically more massive and brighter than our sun. (Most longer-lasting stars that are fainter than the sun are too dim to view without a telescope.) At the end of a star's life, when the supply of available hydrogen is nearly exhausted, it swells and brightens. Stars that are visible to the naked eye are often in this stage. They are, on average, a few hundred million years old. A supergiant star, such as the 10-million-year-old Betelgeuse in Orion, in contrast, will meet its demise much more quickly. It has been spending its fuel so extravagantly that it is expected to collapse within a million years before probably exploding as a supernova. SA

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

MATT COLLINS

COPYRIGHT 2003 SCIENTIFIC AMERICAN, INC.

© 2003 ROZ CHAST, FROM CARTOONBANK.COM

COPYRIGHT 2003 SCIENTIFIC AMERICAN, INC.