

3D DIGITAL REALITY • GENES GET CULTURE • ALZHEIMER'S DISCOVERY

AMERICAN Scientist

March–April 2010

www.americanscientist.org

The Biomechanics of Whale Feeding

SIGMA XI
THE SCIENTIFIC RESEARCH SOCIETY

Congratulations

2010 Sigma Xi Award Winners



William Procter Prize for Scientific Achievement

The award includes a \$5,000 Grant-in-Aid of Research given to a young scholar selected by Dr. Spivey.

Michael Spivey, professor of cognitive science, University of California-Merced



John P. McGovern Science and Society Award

The McGovern Award was established in 1984 to honor an individual who has made an outstanding contribution to science and society.

Barbara Gastel, professor of veterinary integrative biosciences and of humanities in medicine and biotechnology, Texas A&M University



Walston Chubb Award for Innovation

The Chubb Award honors and promotes creativity among scientists and engineers.

Howard Moskowitz, an expert on sensory psychology and its commercial application, is president and CEO of Moskowitz Jacobs Inc.



Young Investigator Award

Sigma Xi members within ten years of their highest earned degree are eligible for this award.

Kevin Gurney, assistant professor of Earth and atmospheric science, Purdue University

Sigma Xi prize lectures will be highlights of the 2010 Sigma Xi Annual Meeting and International Research Conference on November 11-14 at Raleigh Convention Center, Raleigh, North Carolina

www.sigmaxi.org

SIGMA XI

THE SCIENTIFIC RESEARCH SOCIETY

AMERICAN Scientist

Volume 98 • Number 2 • March–April 2010

DEPARTMENTS

98 From the Editor

100 Letters to the Editors

102 Macroscopic

Just-as-good Medicine
David M. Kent

106 Computing Science

Avoiding a digital dark age
Kurt D. Bollacker

112 Engineering

Challenges and prizes
Henry Petroski

117 Marginalia

Two lives
Roald Hoffmann

121 Science Observer

Amplifying with acid • Sunburned ferns? • In the news

156 Sightings

Tracking the Karakoram Glaciers

SCIENTISTS' BOOKSHELF

158 Book Reviews

Empathy • Earthquake prediction •
Stephen Jay Gould

FROM SIGMA XI

175 Sigma Xi Today

2010 Sigma Xi awards • Student conference medalists

FEATURE ARTICLES



124 The Ultimate Mouthful: Lunge Feeding in Rorqual Whales

New technologies bring action at depth to light at the surface
Jeremy A. Goldbogen



132 The Race for Real-time Photorealism

Algorithms and hardware promise graphics indistinguishable from reality
Henrik Wann Jensen and Tomas Akenine-Möller

140 Gene-Culture Coevolution and Human Diet

Biology and culture have conspired to make us who we are

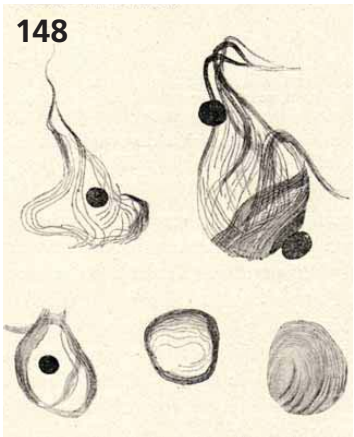
Olli Arjamaa and Timo Vuorisalo

140



148 Finding Alzheimer's Disease

Confidence in the physical basis of mental disorders led to discovery of a disease
Ralf Dahm

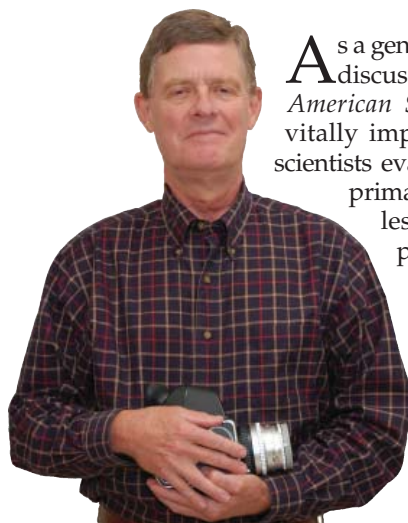


THE COVER



The accordionlike blubber on a blue whale's underside extends from mouth to bellybutton (on the cover). The structure, found only in the family of baleen whales called rorquals, is made from firm ridges (left) connected by deep furrows of delicate elastic tissue, and can stretch to more than twice its original length. Thus the whale's oral cavity can expand to enormous size and hold many tens of tonnes of water and krill; the whale then filters out the water with its baleen while retaining its tiny shrimplike prey. Exactly how rorquals engulf such quantities of water has long been obscured by ocean depths, but as Jeremy A. Goldbogen recounts in "The Ultimate Mouthful: Lunge Feeding in Rorqual Whales" (pages 124–131), electronic devices are aiding researchers in understanding the complex biomechanics behind how these enormous animals eat. (Cover image and image at left courtesy of Nick Pyenson.)

Methodology in Our Madness



As a general rule, we avoid much discussion of methodology in *American Scientist*. Although it's vitally important information for scientists evaluating other scientists' primary work, it usually has less value in a secondary publication, where the science reported has already been through the peer-review filter. Sometimes, though, it's just too interesting for us to push aside.

Jeremy Goldbogen's piece "The Ultimate Mouthful:

Lunge Feeding in Rorqual

Whales" (pp. 124–131) is a great

example. For decades, what baleen whales are up to when they dive as much as 300 meters deep in search of krill has been shrouded in mystery. Theoretical studies could speculate on the biomechanics of rorqual feeding, but it took the development of temporary tags and infrared cameras to actually ride along to dinner with the largest of marine mammals. What Goldbogen and his colleagues found is truly extraordinary, but I won't spoil it by revealing more than that we're talking school-bus scale here.

Once you've digested the biomechanics of whale feeding, you need only turn the page to find an example of technology where methodology *is* the message. Henrik Wann Jensen and Tomas Akenine-Möller are experts in making dancing pixels as convincing as possible. In "The Race for Real-time Photorealism" (pp. 132–139) they describe approaches to representing images in ways that prove convincing to the hu-

man eye yet remain computationally practical. To be honest, much of the driving force behind photorealistic rendering has been to satisfy the ravenous consumers who support the video gaming industry, and science has been the happy secondary beneficiary. We'll take it.

Sometimes, though, the methodology confounds expectations—at least mine. As a photographer, I came late to the digital revolution. And like so many reluctant adopters, I became an enthusiastic (some might add "over-" to the previous word) proponent. Which is why I was so caught by surprise when I read the caption for "Sightings" (pp. 156–157). When Fabiano Ventura set out to duplicate scenes of glaciers captured by Vittorio Sella 100 years ago, digital was not his medium. Instead, he used a 4 x 5 view camera and film—exactly the same medium Sella used. This allowed him to duplicate the geometries of the originals for exact comparisons showing glacial changes. Then he scanned the film and stitched the digital images together to produce panoramas impossible in digital alone. The result certainly throws down the gauntlet to the photorealists mentioned above.

Recently Sigma Xi hosted *ScienceOnline2010*, a conference on electronic media and the communication of science held here each year since 2008. Over that time I've watched the conference grow from a gathering of bloggers to a remarkably diverse meeting of minds on everything from podcasting to social networking to citizen science to, well, blogging. Many of this year's sessions were video recorded, most of which should be available on YouTube by the time you read this. For starters, you might look for a talk by our own Elsa Youngsteadt and her Public Radio International counterpart, Rihitu Chatterjee, on "The World Science" podcast.—David Schoonmaker

American Scientist

David Schoonmaker *Editor*
Morgan Ryan *Managing Editor*
Fenella Saunders *Senior Editor*
Catherine Clabby *Associate Editor*
Mia Smith *Editorial Associate*
Barbara J. Aulicino *Art Director*
Tom Dunne *Assistant Art Director*
Brian Hayes *Senior Writer*
Christopher Brodie *Contributing Editor*
Rosalind Reid *Consulting Editor*
Elsa Youngsteadt *Contributing Editor*

SCIENTISTS' BOOKSHELF

Flora Taylor *Editor*
Anna Lena Phillips *Assistant Editor*

AMERICAN SCIENTIST ONLINE

Greg Ross *Managing Editor*

www.americanscientist.org

Jerome F. Baker *Publisher*
Katie Lord *Associate Publisher*
Jennifer Dorff *Marketing Manager*
Eric Tolliver *Marketing Associate*

ADVERTISING SALES

Kate Miller *Advertising Manager*
advertising@amsi.org • 800-282-0444

EDITORIAL AND SUBSCRIPTION CORRESPONDENCE

American Scientist
P.O. Box 13975
Research Triangle Park, NC 27709
919-549-0097 • 919-549-0090 fax
editors@amsionline.org • subs@amsi.org

PUBLISHED BY SIGMA XI THE SCIENTIFIC RESEARCH SOCIETY

Howard Ceri *President*
Richard L. Meyer *Treasurer*
Joseph A. Whittaker *President-Elect*
James F. Baur *Immediate Past President*
Jerome F. Baker *Executive Director*

PUBLICATIONS COMMITTEE

A. F. Spilhaus, Jr. *Chair*
Jerome F. Baker, Howard Ceri, Lawrence M. Kushner, David Schoonmaker

PRINTED IN USA

SIGMA XI, THE SCIENTIFIC RESEARCH SOCIETY was founded in 1886 as an honor society for scientists and engineers. The goals of the Society are to foster interaction among science, technology and society; to encourage appreciation and support of original work in science and technology; and to honor scientific research accomplishments.



I'm Big...on Hope

***Adopt-A-Manatee®
this Easter***

Call 1-800-432-JOIN (5646)
www.savethemanatee.org

Photo © Ray Eccleston



We all love our magazines,
but when it's time to let go, it just takes
a little extra effort to recycle.
After all, helping the environment
is everyone's responsibility.

GIVE OLD MAGAZINES A NEW LIFE



For more information on how
to recycle your magazines, go to
www.Earth911.com.

MAKE IT A HABIT. RECYCLE. 

Brought to you by the Magazine Publishers of America.

LETTERS TO THE EDITORS

Understand the Material

To the Editors:

Heather Patisaul's feature article "Assessing Risks from Bisphenol A," (January–February) nicely illustrates the difficulties in trying to assess human health effects from low levels of chemicals in our environment. The author would have benefited, however, from collaboration with a materials scientist.

Early on she notes that BPA is a common ingredient in many hard plastics. BPA is a monomer for polycarbonate (66 percent) and epoxy (30 percent). Those polymers constitute only about 3 percent of U.S. plastics production of over 100 billion pounds in 2008. BPA is hardly a common ingredient.

By the end of the article, Patisaul said, "DDT undoubtedly saved lives, and likely still does. No such case can be made for BPA. It is time to develop a clear and comprehensive strategy for assessing the potential public health consequences of endocrine disruptors such as BPA that may contribute only economic value." To understand the public health consequences and develop a clear strategy, one must understand the materials involved, how they are used and the routes and levels of exposure to compounds of concern. Why are epoxy coatings used for certain cans? They reduce the likelihood of botulism. Someone behind impact- or bullet-resistant windows might value the protection they give.

Technologies exist largely because of the underlying materials. Consider CDs. Polycarbonate is a lightweight, high-impact, heat-resistant, intrinsically flame-retardant plastic. The first three qualities are why BPA was used in baby bottles. If better materials are available, great. But let's not throw CDs, electrical appliances and bullet-resistant windows out with the baby bottles.

Gordon L. Nelson
Dean, College of Science
Florida Institute of Technology

Dr. Patisaul responds:

Information about BPA production levels comes from the "NTP-CERHR

Monograph on the Potential Human Reproductive and Developmental Effects of Bisphenol A," at <http://cerhr.niehs.nih.gov/chemicals/bisphenol/bisphenol.pdf>.

As for routes of exposure, you don't need solvents to get BPA to migrate. As it turns out, you don't even need heat. A Harvard University research group reported in June that consumption of cold beverages from polycarbonate bottles containing BPA raises human urine levels of BPA by 69 percent.

Exposure to BPA is low but that does not mean it is innocuous. That type of "the dose makes the poison" thinking may not apply to endocrine disruptors because their dose responses appear to be non-monotonic in many cases. Given that, if you don't need it in food containers, why not pull it out and be on the safe side?

Another View of Hydrogen Sulfide

To the Editors:

I enjoyed Roger P. Smith's article "A Short History of Hydrogen Sulfide" (January–February). Many may not know that the gas plays productive roles in several geologic settings. First, hydrogen sulfide occurs as a minor constituent in most natural gas deposits and must be removed. It is then oxidized to elemental sulfur, a process that produces virtually the sole source of sulfur in North America. Previously, "biogenic" sulfur had to be mined by the Frasch process primarily in Gulf Coast salt domes.

Hydrogen sulfide also plays an important role in forming metallic-sulfide ores of zinc, lead and copper. The bearing fluids of these base metals must encounter a source of hydrogen sulfide (either biogenic or magmatic) along their flow path to precipitate metal-sulfide minerals. Alternatively, volcanic hydrogen sulfide helps form ores of gold and silver, as the precious metals form stable aqueous complexes with hydrogen sulfide, greatly enhancing their solubility in ore fluids. Precious metals often precipitate when the hydrogen sulfide is destroyed by a number of processes, including boiling, which puts hydrogen

sulfide into the vapor phase. Hydrogen sulfide can also be oxidized by certain bacteria to make sulfuric acid, which is thought to be important in cave formation. Volcanic hydrogen sulfide from hot springs at mid-ocean ridges becomes the basis for complex biological communities where chemosynthetic (chemical producing) bacteria use it and carbon dioxide. Finally, we use hydrogen sulfide produced by stimulating sulfate-reducing bacteria to remediate groundwater contaminated by metals, arsenic and radionuclides (US Patent 5,833,855). With hydrogen sulfide, one should consider its good, bad and smelly aspects!

Jim Saunders
Auburn University

An Apollonian Opportunity

To the Editors:

In Dana Mackenzie's interesting column "A Tisket, a Tasket, an Apollonian Gasket" (January–February), Peter Sarnak remarked on the present-day inability of mathematics to prove or explain certain conjectures, including his own. Those conjectures concern the number series of the bends in Apollonian Gaskets. "The necessary mathematics has not been invented yet," Sarnak said.

It is interesting to remember something stated more than 200 years ago by Carl Friedrich Gauss. In his one-page proof of the long-unproven Wilson's prime number theorem, first published by Edward Waring, Gauss noted that "neither of them was able to prove the theorem, and Waring confessed that the demonstration seemed more difficult because no notation can be devised to express a prime number. But in our opinion truths of this kind should be drawn from notions rather than from notations."

Sarnak seems to have ignored Gauss's advice. That, unwittingly, may dissuade those who might otherwise attempt to prove those unsolved theorems.

Bernard H. Soffer
Pacific Palisades, CA

American Scientist (ISSN 0003-0996) is published bimonthly by Sigma Xi, The Scientific Research Society, P.O. Box 13975, Research Triangle Park, NC 27709 (919-549-0097). Newsstand single copy \$4.95. Back issues \$6.95 per copy for 1st class mailing. U.S. subscriptions: one year \$28, two years \$50, three years \$70. Canadian subscriptions: one year \$36; other foreign subscriptions: one year \$43. U.S. institutional rate: \$70; Canadian \$78; other foreign \$85. Copyright © 2010 by Sigma Xi, The Scientific Research Society, Inc. All rights reserved. No part of this publication may be reproduced by any mechanical, photographic or electronic process, nor may it be stored in a retrieval system, transmitted or otherwise copied, with the exception of one-time noncommercial, personal use, without written permission of the publisher. Second-class postage paid at Durham, NC, and additional mailing office. Postmaster: Send change of address form 3579 to Sigma Xi, P.O. Box 13975, Research Triangle Park, NC 27709. Canadian publications mail agreement no. 40040263. Return undeliverable Canadian addresses to P.O. Box 503, RPO West Beaver Creek, Richmond Hill, Ontario L4B 4R6.

To the Editors:

Considering that the geometry of the circle involves irrational numbers such as pi and square roots, I was struck by the seemingly infinite array of integers in the Apollonian gaskets described by Dana Mackenzie. One view of this is that each pair of mutually tangent circles has two infinite series of tangent circles spiraling into crevices between them. There are an infinite number of these mutually tangent pairs, each with a pair of infinite series. Some series appear more than once and some are part of other series.

I have found a linear relation that is somewhat different than Mackenzie's by choosing two tangent circles (say with curvatures a and b) from any four mutually tangent circles. Of the remaining two circles, call the curvature

of the larger d_0 and the curvature of the smaller d_1 . Thus d_0 is the starting term in a series of curvatures and d_1 is the second term. Other curvatures are determined by the linear formula obtained by subtracting Descartes's equation written for a, b, d_{n-2}, d_{n-1} from that for a, b, d_{n-1}, d_n . The resulting equation, $d_n = 2(a + b + d_{n-1}) - d_{n-2}$, can be used to determine the successive values of d_n by a process of iteration. Because Descartes's equation is a quadratic, the difference of the differences between consecutive terms is a constant and equal to $2(a + b)$ in each series. This seems to apply to the irrational roots of Descartes's equation, also.

Ronald Csuha
New York, NY

Dr. Mackenzie responds:

I see no conflict between Sarnak's quote and Gauss's admonition. Sarnak would certainly agree that new notions, not new notations, are needed to prove his "local-to-global principle" for Apollonian packings.

I am glad to report that Elena Fuchs (Sarnak's student) and Jean Bourgain have proven a "positive density theo-

rem." This relates to the question Ron Graham asked, about whether even 1 percent of the numbers that could occur in an Apollonian gasket actually do occur. Fuchs has shown that the answer is yes, provided 1 percent is replaced by a sufficiently small (but positive) number. Interestingly, her approach was to use carefully selected subsets of the Apollonian gasket, an approach not too dissimilar from what Ronald Csuha proposes. Instead of the sequence of all circles tangent to two fixed circles, she looks at the somewhat more complicated sequence of circles tangent to a single fixed circle. The preprint will be posted at the open-access site <http://arXiv.org>.

ILLUSTRATION CREDITS

Macroscope

Page 103 Morgan Ryan
and Barbara Aulicino

Page 104 Barbara Aulicino

Computing Science

Pages 199, 200 Tom Dunne

Engineering

Page 113 Tom Dunne

Marginalia

Pages 118, 119 Barbara Aulicino

The Ultimate Mouthful:

Lunge Feeding in Rorqual Whales

Figures 3, 4 (bottom),
6, 9, 10 Tom Dunne

The Race for Real-time Photorealism

Figures 3, 5, 6 Tom Dunne

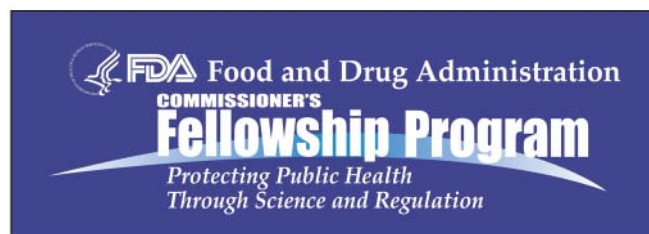
Figure 4 Morgan Ryan

Gene-Culture Coevolution and Human Diet

Figures 2, 7, 8 Barbara Aulicino

The Origins of Alzheimer's Disease

Figure 7 (left) Barbara Aulicino



FDA Commissioner's Fellowship Program

Touch the Lives of All Americans!

The FDA Commissioner's Fellowship Program is a two-year training program designed to attract top-notch health professionals, food scientists, epidemiologists, engineers, pharmacists, statisticians, physicians and veterinarians. The Fellows work minutes from the nation's capital at FDA's new state-of-the-art White Oak campus in Silver Spring, Maryland or at other FDA facilities. The FDA Commissioner's Fellowship offers competitive salaries with generous funds available for travel and supplies.

Coursework & Preceptorship

The FDA Commissioner's Fellowship program combines coursework designed to provide an in-depth understanding of science behind regulatory review with the development of a carefully designed, agency priority, regulatory science project.

Who Should Apply?

Applicants must have a Doctoral level degree to be eligible. Applicants with a Bachelor's degree in an Engineering discipline will also be considered. Candidates must be a U.S. citizen, a non-citizen national of the U.S., or have been admitted to the U.S. for permanent residence before the program start date. For more information, or to apply, please visit: www.fda.gov/commissionersfellowships/default.htm.

Applications will be accepted from December 15, 2009 – March 15, 2010

MACROSCOPE

Just-as-good Medicine

David M. Kent

THE RABBI'S EULOGY for Sheldon Kravitz solved a minor mystery for my father: what was behind the odd shape of the juice cups he had been drinking from after morning services for the last few years? Adding a bit of levity while praising his thrift and resourcefulness, the rabbi told of how Sheldon purchased, for pennies on the dollar, hundreds of urine specimen cups from Job Lot, that legendary collection of pushcarts in lower Manhattan carrying surplus goods—leftovers, overproduced or discontinued products, unclaimed cargo. At the risk of perpetuating a pernicious cultural stereotype, for men of my father's generation like Sheldon, raised during the Great Depression, bargain hunting was a contact sport and Job Lot was a beloved arena. My father, too, would respond to the extreme bargains there with ecstatic automatisms of purchasing behavior and come home with all manner of consumer refuse, including, and to my profound dismay, sneakers that bore (at best) a superficial resemblance to the suede Pumas worn and endorsed by my basketball idol, the incomparably smooth Walt "Clyde" Frazier. My father would insist that such items were "just as good" as the name brands. But we, of course, knew what "just as good" really meant.

In fairness to my father and his friends, from a utilitarian perspective (decidedly not the perspective of pre-adolescents), maximizing the overall good of the family involves economic trade-offs. Money saved from something "just as good" can be reallocated toward items that bring greater benefit

*Less expensive,
lower-quality
innovations abound
in every economic
sector—except
medicine*

than the value sacrificed. Indeed, these types of cost-versus-quality trade-offs are ubiquitous in our economy, and are especially useful when resources are tightly constrained. Those following the long march to health-care reform know that one of the few things beyond argument is that the old approach is unsustainable and threatens to bankrupt the country. Perhaps a little belt tightening and bargain hunting of this sort might make our health-care dollars stretch farther.

The Cost-effectiveness Plane

To help maximize the overall benefits in health care under a utilitarian framework and conditions of constrained resources, health economists use an analytic tool called cost-effectiveness analysis (CEA) that quantifies the added expenditure necessary to obtain a unit of health benefit (typically measured in quality-adjusted life years or QALYs, pronounced "kwallies"). The most common application of CEA is to examine the value of medical innovations compared to the standard of care routinely available, since new technologies are an important cause of the increase in health-care costs.

If the "unit cost" for a QALY of benefit (that is, the cost-effectiveness ratio) is less than some threshold (conventional-

ly \$50,000 or \$100,000 per QALY), then adoption of the innovation is deemed "incrementally cost-effective," since the benefit obtained compares favorably to that obtainable at similar cost using accepted medical technologies (such as dialysis, which has a cost-effectiveness ratio variously estimated at between \$50,000 and \$80,000 per QALY). Above the ratio, they are deemed not to be cost-effective. That is, the (relatively small) incremental benefits of the intervention do not justify the (relatively large) incremental costs.

Comparisons between alternative approaches in cost-effectiveness analyses can usefully be depicted on a cost-effectiveness plane, shown in the figure opposite. Most studied medical innovations fall into the northeast quadrant of this plane; that is, they increase both costs and health benefits. Within this quadrant, the acceptability threshold would be represented by a line of constant slope, indicating the "willingness to pay" (WTP) for a QALY, separating nominally cost-effective therapies from cost-ineffective therapies.

Of course, if all innovation in health care fell into this northeast quadrant, innovation could only increase the costs of care. That is, even so-called cost-effective health-care innovations would always cost more money than the alternatives they replaced. This is often a point of confusion, sometimes purposeful, as when our political leaders claim that "preventative medicine" is highly cost-effective and would therefore save money. In fact, while most *recommended* preventative services are cost-effective (meaning the value of their benefits in terms of QALYs gained justifies the costs in terms of dollars spent), only very rarely are preventative services actually cost-saving, even when all the "downstream" avoided medical expenses are folded into the analysis. Indeed, new "cost-effective" innovations are one of

David M. Kent is an associate professor of medicine at Tufts Medical Center and the associate director of the Clinical and Translational Science Program at the Sackler School for Graduate Biomedical Sciences of Tufts University. Address: Tufts Medical Center, 800 Washington St. #63, Boston, MA 02111. Email: dkent1@tuftsmedicalcenter.org

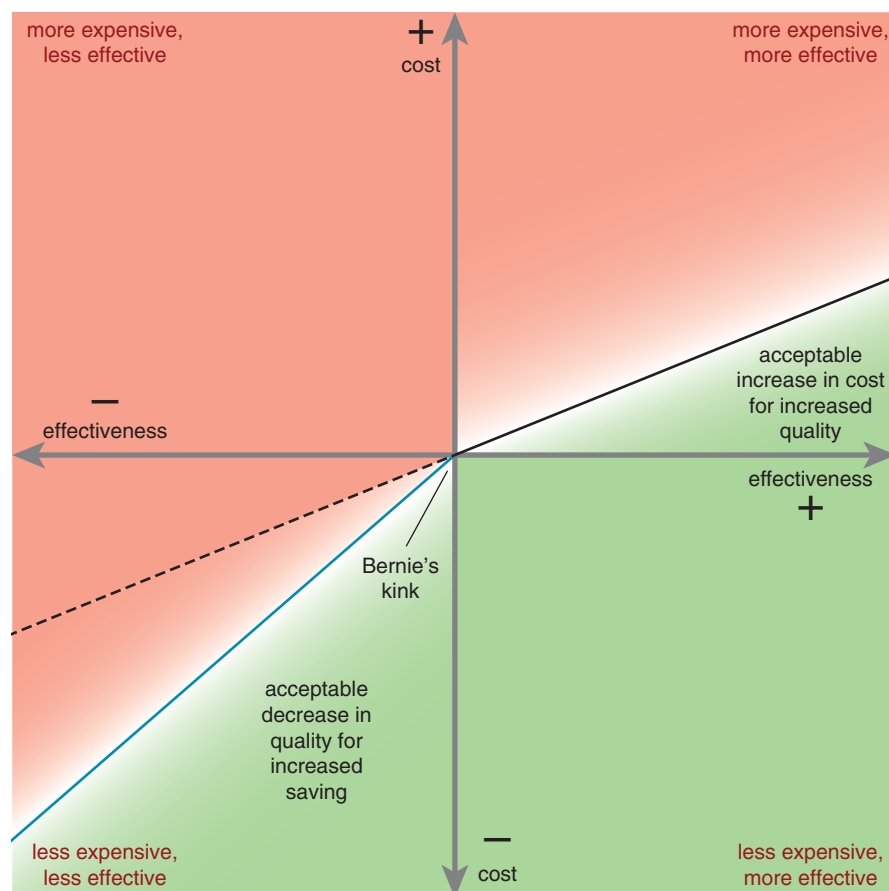
the principal reasons that health-care costs continue to soar.

In fact, only innovations that fall south of the equator in the cost-effectiveness plane are actually cost-saving. When those innovations are also superior to the alternative, or standard of care, they are considered “dominant” (that is, cost decreasing *and* quality improving); adoption of these southeast quadrant innovations should not be controversial. However, as health-care costs continue to rise, cost-saving innovations may be increasingly attractive even when they do not improve care, particularly in a weak economy. While some innovations in the southwest quadrant would clearly be unattractive because they are substantially worse than the available standard of care or offer only trivial cost savings, what about innovations that offer substantial cost saving and are genuinely almost as good as the standard? In a 2004 article in *Medical Decision Making*, fellow researchers and I described innovation that is greatly cost saving but only slightly quality reducing as “decrementally” cost-effective. In such cases, the savings could potentially increase the overall good despite the sacrificed benefit. Indeed, if “much cheaper, almost as good” products are attractive in other economic sectors because they permit the reallocation of saved resources to items of more value than the benefits sacrificed, why not in medical care as well?

Bernie's Kink

Men generally fix their affections more on what they are possessed of, than on what they never enjoyed: For this reason, it would be greater cruelty to dispossess a man of any thing than not to give it [to] him.—David Hume, *A Treatise on Human Nature*

Theoretically, perfectly rational economic agents seeking to maximize their welfare would be similarly willing to relinquish QALYs obtained from some routinely available standard-of-care for a new “much cheaper, almost as good” therapy, if the savings could be reallocated to an item of equal or higher value than what was sacrificed. Put another way, the selling price (often referred to as willingness to accept, or WTA) and the buying price (willingness to pay, WTP) of a QALY should



Medical innovations fall into one of four quadrants on the cost-effectiveness plane, based on how they compare with existing standards of care. For example, the top left quadrant represents innovative treatments that are more expensive and less effective—an off-putting combination; bottom right represents less expensive, more effective treatments—an easy decision. In between, the decision process is not so obvious. The diagonal lines represent thresholds for the acceptability of cost-effectiveness tradeoffs. Above the diagonals (in the red regions), the balance of cost and effectiveness is rejected. Of special interest is “Bernie’s kink” at the origin, which reveals how medical markets actually behave. People prove to be unwilling to surrender quality using the same formula they would use to accept increased cost.

be similar, and the societal threshold for accepting or rejecting a technology should be symmetric and pass through the origin of the cost-effectiveness plane as a straight line. However, as David Hume anticipated, a reproducible observation is that consumers’ willingness to accept monetary compensation to forgo something they have is typically greater, and often much greater, than their stated willingness to pay for the same benefit. Several explanations exist, including the so-called “endowment effect,” the psychological principle that people value items that they already have simply because they already have them.

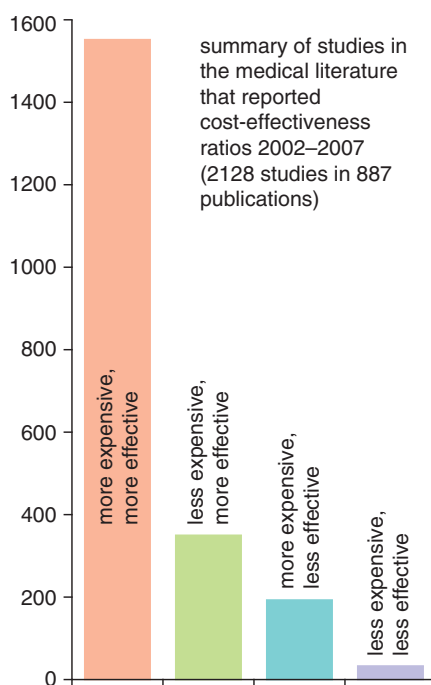
A 2002 review of 20 studies by the late Bernie O’Brien and his colleagues at McMaster University found that the ratio of individuals’ WTA to WTP was always greater than 1 and ranged from

1.9 to 6.4 for two scenarios specifically related to health care. They suggested that rather than a symmetric accept-reject threshold on the cost-effectiveness plane, societal thresholds should reflect the WTA-WTP gap seen in individual preferences, which would be captured by a downward “kink” (subsequently known as “Bernie’s kink”) in the threshold as it passed through the origin, indicating that a QALY’s selling price in the southwest would always be higher than a QALY’s buying price in the northeast.

Thus, there may be an inherent cognitive bias against relinquishing the gains of health-care interventions that have already been accepted, and the cost savings from decrementally cost-effective innovation may need to be substantially greater than conventionally used thresholds suggest.

Bargain Hunting

Whereas all this fancy theory plus a token can get you on the subway, might there be practical applications of “decrementally” cost-effective innovation? To explore this, working with colleagues at the Tufts Center for the Evaluation of Value and Risk (who maintain a comprehensive database of cost-utility studies), we enlisted Aaron Nelson, then a medical student, to help us sort through more than 2,000 cost-utility comparisons for any potential examples that might be decrementally cost-effective. We found that about three-quarters of published comparisons described new technologies or treatment strategies that increase both costs and benefits, and that most of these (about 65 to 80 percent) were cost-effective by conventional criteria (depending on which conventional threshold was used, \$50,000 or \$100,000 per QALY gained). Less often, published analyses described innovations that are either dominant or dominated (about 10 percent and 15 percent of the time,



A survey of more than 2,000 medical studies that reported cost-effectiveness ratios highlights a striking difference between medical and other consumer markets. In the hurly-burly of retail markets, producing “nearly as good” products for less money is a major competitive strategy; in the medical literature, that type of innovation (“less expensive, less effective”) is hardly represented at all (purple bar).

respectively), but only very rarely were innovations both cost- and quality-decreasing. Indeed, fewer than 2 percent of all comparisons were classified in the cost- and quality-decreasing “southwest quadrant”, and only 9 (involving 8 innovations) were found to be decrementally cost-effective (0.4 percent of the total)—that is, they saved at least \$100,000 for each QALY relinquished.

Examples of these cost-saving interventions include using the catheter-based percutaneous coronary intervention in place of bypass surgery for multivessel coronary disease, which on average saves about \$5,000 while sacrificing a half day of perfect health (for a cost-savings of more than \$3 million for every QALY lost) and using repetitive transcranial magnetic stimulation instead of electroconvulsive therapy for drug-resistant major depression, which avoids the need for general anaesthesia and saves on average over \$11,000 but sacrifices about a week of perfect health (for a ratio of more than \$500,000 for every QALY lost). Nearly all the remaining innovations involved the tailored withholding of standard therapy, including watchful waiting for selected patients with inguinal hernia, withholding mediastinoscopy for selected patients with lung cancer, and abbreviated physiotherapy or psychotherapy for patients with neck pain or deliberate self-harm, respectively. Finally, the cost-saving innovations included the sterilization and reuse of dialysate, the chemical bath used in dialysis to draw fluids and toxins out of the bloodstream—a degree of thrift even the late Sheldon Kravitz would have to admire.

That decrementally cost-effective innovations are so rarely described in the health-care literature suggests that medicine is distinct from most other markets, in which cost-decreasing, quality-reducing products are continuously being introduced—think IKEA, Walmart and the Tata car. Several reasons may explain this “medical exceptionalism.” First, there is fundamentally a lack of incentives both for physicians to control costs, especially under a fee-for-service regime, and for patients to demand less expensive treatment when insurance shields them from the direct costs of care. Second, medical “bargains” frequently come with health risks, and trading health for money strikes some as vul-

gar, regardless of ratio. The inherent ethical unease that decrementally cost-effective innovations can elicit poses a serious public relations and marketing challenge.

However, consumers have been comfortable with many decrementally cost-effective options outside of health care that pose similar health risks. For example, automobile manufacturers produce many vehicles that lack certain safety features (for example, side-impact airbags), because some consumers are willing to forgo those options to reduce the purchase price. Why not in health care?

Lowering Health Costs: Buy Less Stuff

Even by the standards of political rhetoric, it strains credulity when politicians suggest that the declared goals of health-care reform—increasing access, improving quality and controlling costs—are somehow mutually reinforcing. I’m no Peter Orszag, the über-wonk overseeing President Obama’s Office of Management and Budget, but if my father taught me anything it was that saving money rarely involves buying more and better stuff. Plain talk about ways to cut costs are buried in rhetoric about rooting out inefficiencies and various prevarications about savings from investing in (that is, spending on) more preventative medicine, health information technology, and comparative effectiveness research about what therapies work best for which patients. While these goals may all be worthwhile, and there is much of little or no value in the current system (including the immense amount of money spent to maintain our Byzantine for-profit insurance system), ultimately we simply do not have the resources to give away an expensive commodity like health care in quantities that people want, subject to no budgetary constraints.

It is beyond dispute that some mechanisms for the controlled distribution of these expensive goods and services are required. In most markets, prices play this role, and many feel that the fundamental problem in health care is that many consumers are shielded from the costs of their care. A system based largely on prices (that is, price rationing) may control costs better than our current system, but it would of course mean that those with the most money have first dibs on scarce health-care resources, and

there might be little left over for those without means. (There are other reasons too why most consumers can't be expected to comparison shop for emergency coronary angioplasty or for charged-particle radiosurgery for their glioblastoma the same way they might for gasoline, underwear and cling peaches). It is a fantasy to believe that price rationing alone can provide an acceptable mechanism for the controlled distribution of medical services, and some other means are thus also needed. Perhaps we should take it as a sign of the robustness of our democracy that this rather technical issue of the proper mix and variety of price and non-price rationing has somehow managed to plunge our national conversation about health-care reform into a Jerry Springer-style shouting match, except without the civility.

But regardless of the mix, expanding coverage to the uninsured, caring for our aging baby boomers, and accommodating new, effective technologies—while still feeding, clothing, housing, and educating ourselves, and catching an occasional movie—will require our system of distribution of health services to be more cost-

sensitive, and will almost certainly mean the adoption of some decrementally cost-effective strategies for saving money. For example, Canadian-style delays for expensive diagnostic or surgical procedures certainly pose real, albeit small, medical risks, and would fall into this southwest category. Getting insured Americans to accept such new risks may be difficult, but slightly quality-reducing (that is, risk-increasing) cost-saving strategies have already been widely adopted within the American system, even if not studied or widely acknowledged. The gradual increase in the "hassle factor" in accessing medical care is one covert way that the industry has found to limit the distribution of services. More overt examples of rationing already adopted include aggressively shortening hospital stays and limiting formulary options (which sometimes require patients to change from a medicine they have been tolerating well to another in the same class). Despite the fact that doctors regularly (although sometimes disingenuously) deploy patter informing patients that the hospital is a dangerous place to stay and that the formulary medication is "just as good" as the one they've been

taking, these strategies are certainly associated with small but real risks. Even a preadolescent quickly learns the true meaning of "just as good"; perhaps a more mature citizenry can also come to appreciate some of the upside of having "just as good" alternatives.

Bibliography

- Orszag, P. R., and P. Ellis. 2007. The challenge of rising health care costs—a view from the Congressional Budget Office. *New England Journal of Medicine* 357:1793–1795.
- Cohen, J. T., P. J. Neumann and M. C. Weinstein. 2008. Does preventive care save money? Health economics and the presidential candidates. *New England Journal of Medicine* 358:661–663.
- Kent, D. M., A. M. Fendrick, and K. M. Langa. 2004. New and dis-improved: On the evaluation and use of less effective, less expensive medical interventions. *Medical Decision Making* 24:281–286.
- O'Brien, B. J., K. Gertsen, A. R. Willan and L. A. Faulkner. 2002. Is there a kink in consumers' threshold value for cost-effectiveness in health care? *Health Economics* 11:175–180.
- Nelson, A. L., J. T. Cohen, D. Greenberg and D. M. Kent. 2009. "Much Cheaper, Almost as Good": Decrementally Cost Effective Medical Innovation. *Annals of Internal Medicine* 151:662–667.



There's someone you need to meet.

Come meet your group's newest member, the GEICO Gecko. Sigma Xi members could get an additional discount on car insurance. Call or click for your **FREE** quote.

GEICO
geico.com
1-800-368-2734

Discount amount varies in some states. Discount is not available in all states or in all GEICO companies. One group discount applicable per policy. Coverage is individual. In New York a premium reduction is available. Some discounts, coverages, payment plans and features are not available in all states or companies. Government Employees Insurance Co. • GEICO General Insurance Co. • GEICO Indemnity Co. • GEICO Casualty Co. These companies are subsidiaries of Berkshire Hathaway Inc. GEICO: Washington, DC 20076. GEICO Gecko image © 1999-2010. © 2010 GEICO

COMPUTING SCIENCE

Avoiding a Digital Dark Age

Kurt D. Bollacker

WHEN I WAS A BOY, I discovered a magnetic reel-to-reel audio tape recorder that my father had used to create “audio letters” to my mother while he was serving in the Vietnam War. To my delight (and his horror), I could listen to many of the old tapes he had made a decade before. Even better, I could make recordings myself and listen to them. However, all of my father’s tapes were decaying to some degree—flaking, stretching and breaking when played. It was clear that these tapes would not last forever, so I copied a few of them to new cassette tapes. While playing back the cassettes, I noticed that some of the sound quality was lost in the copying process. I wondered how many times I could make a copy before there was nothing left but a murky hiss.

A decade later in the 1980s I was in high school making backups of the hard drive of my PC onto 5-1/4-inch floppy disks. I thought that because digital copies were “perfect,” and I could make perfect copies of perfect copies, I couldn’t lose my data, except by accident. I continued to believe that until years later in college, when I tried to restore my backup of 70 floppy disks onto a new PC. To my dismay, I discovered that I had lost the floppy disk containing the backup program itself, and thus could not restore my data. Some investigation revealed that the company that made the software had long since gone out of business. Requests on electronic bulletin board systems and searches on Usenet turned up nothing useful. Although all of the data on them

Data longevity depends on both the storage medium and the ability to decipher the information

may have survived, my disks were useless because of the proprietary encoding scheme used by my backup program.

The Dead Sea scrolls, made out of still-readable parchment and papyrus, are believed to have been created more than 2,000 years ago. Yet my barely 10-year-old digital floppy disks were essentially lost. I was furious! How had the shiny new world of digital data, which I had been taught was so superior to the old “analog” world, failed me? I wondered: Had I had simply misplaced my faith, or was I missing something?

Over the course of the 20th century and into the 21st, an increasing proportion of the information we create and use has been in the form of digital data. Many (most?) of us have given up writing messages on paper, instead adopting electronic formats, and have exchanged film-based photographic cameras for digital ones. Will those precious family photographs and letters—that is, email messages—created today survive for future generations, or will they suffer a sad fate like my backup floppy disks? It seems unavoidable that most of the data in our future will be digital, so it behooves us to understand how to manage and preserve digital data so we can avoid what some have called the “digital dark age.” This is the idea—or fear!—that if we cannot learn to explicitly save our digital data, we will lose that data and, with it, the record that future generations might use to remember and understand us.

Save Our Bits!

The general problem of data preservation is twofold. The first matter is preservation of the data itself: The physical media on which data are written must be preserved, and this media must continue to accurately hold the data that are entrusted to it. This problem is the same for analog and digital media, but unless we are careful, digital media can be more fragile.

The second part of the equation is the comprehensibility of the data. Even if the storage medium survives perfectly, it will be of no use unless we can read and understand the data on it. With most analog technologies such as photographic prints and paper text documents, one can look directly at the medium to access the information. With all digital media, a machine and software are required to read and translate the data into a human-observable and comprehensible form. If the machine or software is lost, the data are likely to be unavailable or, effectively, lost as well.

Preservation

Unlike the many venerable institutions that have for centuries refined their techniques for preserving analog data on clay, stone, ceramic or paper, we have no corresponding reservoir of historical wisdom to teach us how to save our digital data. That does not mean there is nothing to learn from the past, only that we must work a little harder to find it. We can start by briefly looking at the historical trends and advances in data representation in human history. We can also turn to nature for a few important lessons.

The earliest known human records are millennia-old physical scrapings on whatever hard materials were available. This medium was often stone, dried clay, bone, bamboo strips or even tortoise shells. These substances were very durable—indeed, some specimens have

Over the past two decades, Kurt D. Bollacker has romped through the fields of artificial intelligence, digital libraries, linguistics, databases and electrocardiology. He currently is the digital research director of the Long Now Foundation and gets his hands dirty as a freelance data miner and builder of collaborative knowledge-creation tools. He also works on the Rosetta Project. He received his Ph.D. in computer engineering from the University of Texas at Austin in 1998. Email: kurt@longnow.org

survived for more than 5,000 years. However, stone tablets were heavy and bulky, and thus not very practical.

Possibly the first big advance in data representation was the invention of papyrus in Egypt about 5,500 years ago. Paper was lighter and easier to make, and it took up considerably less space. It worked so well that paper and its variants, such as parchment and vellum, served as the primary repositories for most of the world’s information until the advent of the technological revolution of the 20th century.

Technology brought us photographic film, analog phonographic records, magnetic tapes and disks, optical recording, and a myriad of exotic, experimental and often short-lived data media. These technologies were able to represent data for which paper cannot easily be used (video, for example). The successful ones were also usually smaller, faster, cheaper and easier to use for their intended applications. In the last half of the 20th century, a large part of this advancement included a transition from analog to digital representations of data.

Even a brief investigation into a small sampling of information-storage media technologies throughout history quickly uncovers much dispute regarding how long a single piece of each type of media might survive. Such uncertainty cannot be settled without a time machine, but we can make reasonable guesses based on several sources of varying reliability. If we look at the time of invention, the estimated lifespan of a single piece of each type of media and the encoding method (analog or digital) for each type of data storage (see the table, above right), we can see that new media types tend to have shorter lifespans than older ones, and digital types have shorter lifespans than analog ones. Why are these new media types less durable? Shouldn’t technology be getting better rather than worse? This mystery clamors for a little investigation.

To better understand the nature of and differences between analog and digital data encoding, let us use the example of magnetic tape, because it is one of the oldest media that has been used in both analog and digital domains. First, let’s look at the relationship between information density and data-loss risk. A standard 90-minute analog compact cassette is 0.00381 meters wide by about 129 meters long, and a typical digital audio tape (DAT) is 0.004 meters wide by 60 meters long. For audio encodings of sim-

type of medium	data medium	approximate year of invention	ideal expected lifetime of medium
analog	clay/stone tablet	8000 BC	>4,000 years
analog	pigment on paper	3500 BC	>2,000 years
analog	oil painting	600	centuries
analog	silver halide black and white photographic film	1820	>100 years
analog	modern color photographic film	1860	decades
analog	phonograph record	1877	>120 years
analog/digital	magnetic tape	1928	decades
analog/digital	magnetic disk	1950	3–20 years
analog/digital	polycarbonate optical WORM disk	1990	5–20 years

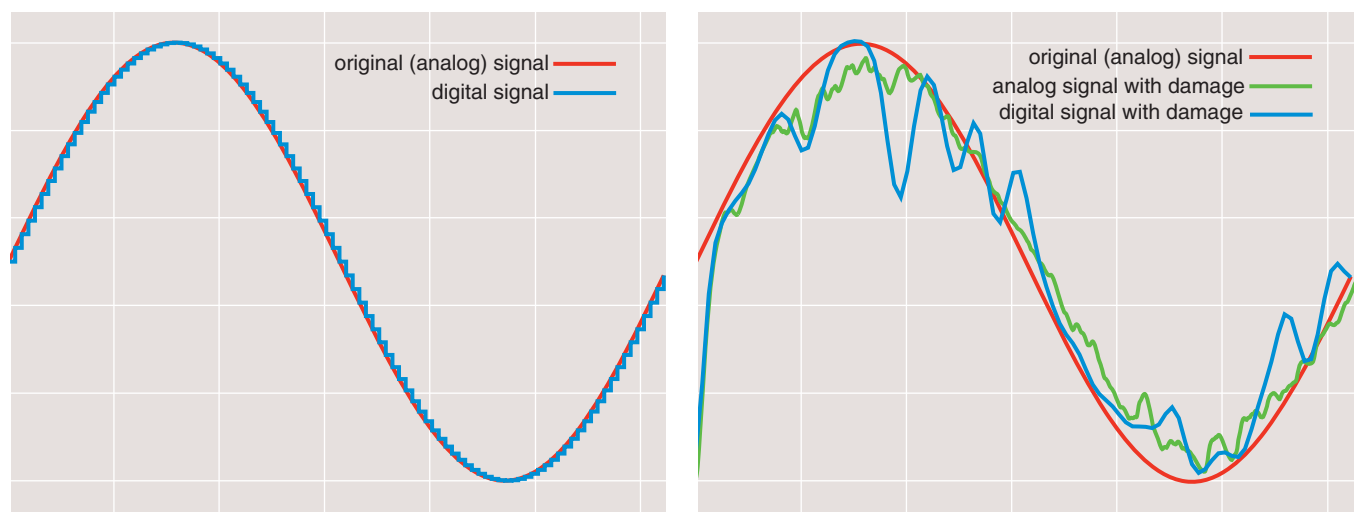
When we compare the different data-storage media that have appeared over the course of human history, a trend emerges: Digital data types are expected to have shorter lifetimes than analog ones.

ilar quality (such as 16 bit, 44.1 kilohertz for digital, or 47.6 millimeters per second for analog), the DAT can record 500 minutes of stereo audio data per square meter of recordable surface, whereas the analog cassette can record 184 minutes per square meter. This means the DAT holds data about 2.7 times more densely than the cassette. The second table (below) gives this comparison for several common consumer audio-recording media types. Furthermore, disk technologies tend to hold data more densely than tapes, so it is no surprise that magnetic tape has all but disappeared from the consumer marketplace.

However, enhanced recording density is a double-edged sword. Assume that for each medium a square millimeter of surface is completely corrupted. Common sense tells us that media that hold more data in this square millimeter would experience more actual data loss; thus for a given amount of lost physical medium, more data will be lost from digital formats. There is a way to design digital encoding with a lower data density so as to avoid this problem, but it is not often used. Why? Cost and efficiency: It is usually cheaper to store data on digital media because of the increased density.

type of medium	audio data medium	recording capacity (minutes per square meter)
analog	6.35 millimeter wide 190.5 millimeters per second reel-to-reel magnetic tape	13.8
analog	33-1/3 RPM vinyl album	411
analog	90-minute audio cassette	184
digital	compact disk (CD)	8,060
digital	60-meter digital audio tape (DAT)	500
digital	2 terabyte 89-millimeter hard drive	4,680,000

As technology has advanced, the density of data storage on analog and, subsequently, digital recording media has tended to increase. The downside of packing in data, however, is that more of the information will be lost if a portion of the recording medium becomes damaged.



A simple audio tone is represented as a sine wave in an analog signal, and as a similar wave but with an approximated stepped shape in a digital signal (left). If the data receive simulated damage, the analog signal output is more resistant to damage than the digital one, which has wilder swings and higher error peaks (right). This result is largely because in a digital recording, all bits do not have the same worth, so damage causes random output error.

A possibly more important difference between digital and analog media comes from the intrinsic techniques that comprise their data representations. Analog is simply that—a physical analog of the data recorded. In the case of analog audio recordings on tape, the amplitude of the audio signal is represented as an amplitude in the magnetization of a point on the tape. If the tape is damaged, we hear a distortion, or “noise,” in the signal as it is played back. In general, the worse the damage, the worse the noise, but it is a smooth transition known as *graceful degradation*. This is a common property of a system that exhibits *fault tolerance*, so that partial failure of a system does not mean total failure.

Unlike in the analog world, digital data representations do not inherently degrade gracefully, because digital encoding methods represent data as a string of binary digits (“bits”). In all digital symbol number systems, some digits are worth more than others. A common digital encoding mechanism, pulse code modulation (PCM), represents the total amplitude value of an audio signal as a binary number, so damage to a random bit causes an unpredictable amount of actual damage to the signal.

Let’s use software to concoct a simulated experiment that demonstrates this difference. We will compare analog

and PCM encoding responses to random damage to a theoretically perfect audiotape and playback system. The first graph in the third figure (above) shows analog and PCM representations of a single audio tone, represented as a simple sine wave. In our perfect system, the original audio source signal is identical to the analog encoding. The PCM encoding has a stepped shape

showing what is known as *quantization error*, which results from turning a continuous analog signal into a discrete digital signal. This class of error is usually imperceptible in a well-designed system, so we will ignore it for now.

For our comparison, we then randomly damage one-eighth of the simulated perfect tape so that the damaged parts have a random amplitude re-

ZIP code digit value	POSTNET code	POSTNET code with missing middle digit
0		
1		
2		
3		
4		
5		
6		
7		
8		
9		

The U.S. Postal Service uses an encoding scheme for ZIP code numbers called POSTNET that uses an error-correcting code. Each decimal digit is represented as five bars. If, say, the middle bar disappears, each number is still distinguishable from all the others.



The Phaistos Disk, housed at the Heraklion Archaeological Museum in Crete, is well preserved and all its data are visible, but the information is essentially lost because the language in which it is written has been forgotten. (Photograph courtesy of Wikimedia Commons.)

sponse. The second graph in the third figure (*facing page, top*) shows the effect of the damage on the analog and digital encoding schemes. We use a common device called a *low-pass filter* to help minimize the effect of the damage on our simulated output. Comparing the original undamaged audio signal to the reconstructions of the damaged analog and digital signals shows that, although both the analog and digital recordings are distorted, the digital recording has wilder swings and higher error peaks than the analog one.

But digital media are supposed to be better, so what's wrong here? The answer is that analog data-encoding techniques are intrinsically more robust in cases of media damage than are naive digital-encoding schemes because of their inherent redundancy—there's more to them, because they're continuous signals. That does not mean digital encodings are worse; rather, it's just that we have to do more work to build a better system. Luckily, that is not too hard. A very common way to do this is to use a binary-number representation that does not mind if a few bits are missing or broken.

One important example where this technique is used is known as an error correcting code (ECC). A commonly used ECC is the U.S. Postal Service's POSTNET (Postal Numeric Encoding Technique), which represents ZIP codes on the front of posted envelopes. In this scheme, each decimal digit is represented as five binary digits, shown as long or short printed bars (*facing page, bottom*). If any single bar for any decimal digit were missing or incorrect, the representation would still not be confused with

that of any other digit. For example, in the rightmost column of the table, the middle bar for each number has been erased, yet none of the numbers is mistakable for any of the others.

Although there are limits to any specific ECC, in general, any digital-encoding scheme can be made as robust as desired against random errors by choosing an appropriate ECC. This is a basic result from the field of information theory, pioneered by Claude Shannon in the middle of the 20th century. However, whichever ECC we choose, there is an economic tradeoff: More redundancy usually means less efficiency.

Nature can also serve as a guide to the preservation of digital data. The digital data represented in the DNA of living creatures is copied into descendants, with only very rare errors when they reproduce. Bad copies (with destructive mutations) do not tend to survive. Similarly, we can copy digital data from medium to medium with very little or no error over a large number of generations. We can use easy and effective techniques to see whether a copy has errors, and if so, we can make another copy. For instance, a common error-catching program is called a *checksum function*: The algorithm breaks the data into binary numbers of arbitrary length and then adds them in some fashion to create a total, which can be compared to the total in the copied data. If the totals don't match, there was likely an accidental error in copying. Error-free copying is not possible with analog data: Each generation of copies is worse than the one before, as I learned from my father's reel-to-reel audiotapes.

Because any single piece of digital media tends to have a relatively short lifetime, we will have to make copies far more often than has been historically required of analog media. Like species in nature, a copy of data that is more easily "reproduced" before it dies makes the data more likely to survive. This notion of *data promiscuousness* is helpful in thinking about preserving our own data. As an example, compare storage on a typical PC hard drive to that of a magnetic tape. Typically, hard drives are installed in a PC and used frequently until they die or are replaced. Tapes are usually written to only a few times (often as a backup, ironically) and then placed on a shelf. If a hard drive starts to fail, the user is likely to notice and can quickly make a copy. If a tape on a shelf starts to die, there is no easy way for the user to know, so very often the data on the

tape perishes silently, likely to the future disappointment of the user.

Comprehensibility

In the 1960s, NASA launched *Lunar Orbiter 1*, which took breathtaking, famous photographs of the Earth juxtaposed with the Moon. In their rush to get astronauts to the Moon, NASA engineers created a mountain of magnetic tapes containing these important digital images and other space-mission-related data. However, only a specific, rare model of tape drive made for the U.S. military could read these tapes, and at the time (the 1970s to 1980s), NASA had no interest in keeping even one compatible drive in good repair. A heroic NASA archivist kept several donated broken tape drives in her garage for two decades until she was able to gain enough public interest to find experts to repair the drives and help her recover these images.

Contrast this with the opposite problem of the analog Phaistos Disk (*above left*), which was created some 3,500 years ago and is still in excellent physical condition. All of the data it stores (about 1,300 bits) have been preserved and are easily visible to the human eye. However, this disk shares one unfortunate characteristic with my set of 20-year-old floppy disks: No one can decipher the data on either one. The language in which the Phaistos disk was written has long since been forgotten, just like the software to read my floppies is equally irretrievable.

These two examples demonstrate digital data preservation's other challenge—comprehensibility. In order to survive, digital data must be understandable by both the machine reading them and the software interpreting them. Luckily, the short lifetime of digital media has forced us to gain some experience in solving this problem—the silver lining of the dark clouds of a looming potential digital dark age. There are at least two effective approaches: choosing data representation technologies wisely and creating mechanisms to reach backward in time from the future.

Make Good Choices ...

In order to make sure digital data can be understood in the future, ideally we should choose representations for our data for which compatible hardware and software are likely to survive as well. Like species in nature, digital formats that are able to adapt to new environments and threats will tend to

survive. Nature cannot predict the future, but the mechanism of mutation creates different species with different traits, and the fittest prevail.

Because we also can't predict the future to know the best data-representation choices, we try to do as nature does. We can copy our digital data into as many different media, formats and encodings as possible and hope that some survive.

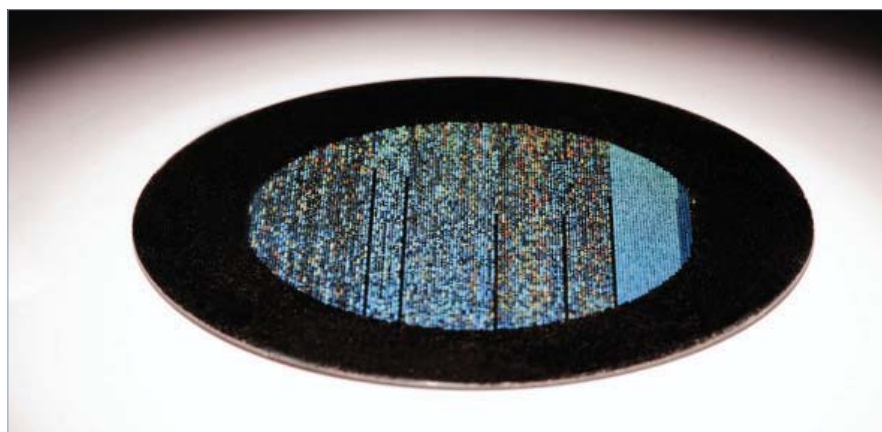
Another way to make good choices is to simply follow the pack. A famous example comes from the 1970s, when two competing standards for home video recording existed: Betamax and VHS. Although Betamax, by many technical measures, was a superior standard and was introduced first, the companies supporting VHS had better business and marketing strategies and eventually won the standards war. Betamax mostly fell into disuse by the late 1980s; VHS survived until the mid-2000s. Thus if a format or media standard is in more common use, it may be a better choice than one that is rare.

... Or Fake It!

Once we've thrown the dice on our data-representation choices, is there anything else we can do? We can hope we will not be stuck for decades, like our NASA archivist, or left with a perfectly readable but incomprehensible Phaistos disk. But what if our scattershot strategy of data representation fails, and we can't read or understand our data with modern hardware and software? A very common approach is to fake it!

If we have old digital media for which no compatible hardware still exists, modern devices sometimes can be substituted. For example, cheap and ubiquitous optical scanners have been commonly used to read old 80-column IBM punchcards. This output solves half of the problem, leaving us with the task of finding hardware to run the software and interpret the data that we are again able to read.

In the late 1950s IBM introduced the IBM 709 computer as a replacement for the older model IBM 704. The many technical improvements in the 709 made it unable to directly run software written for the 704. Because customers did not want either to lose their investment in the old software or to forgo new technological advances, IBM sold what they called an *emulator* module for the 709, which allowed it to pretend to be a 704 for the purposes of running the old software. Emulation is now a common



The Rosetta Project aims to preserve all of the world's written languages with a metal disk that could last up to 2,000 years. The disk records miniaturized versions of more than 13,000 pages of text and images, etched onto the surface using techniques similar to computer-chip lithography. (Photograph by Spencer Lowell, courtesy of the Long Now Foundation, <http://www.longnow.org>.)

technique used to run old software on new hardware. It does, however, have a problem of recursion—what happens when there is no longer compatible hardware to run the emulator itself? Emulators can be layered like Matryoshka dolls, one running inside another running inside another.

Being Practical

Given all of this varied advice, what can we do to save our personal digital data? First and foremost, make regular backup copies onto easily copied media (such as hard drives) and place these copies in different locations. Try reading documents, photos and other media whenever upgrading software or hardware, and convert them to new formats as needed. Lastly, if possible, print out highly important items and store them safely—there seems to be no getting away from occasionally reverting to this “outdated” media type. None of these steps will guarantee the data's survival, but not taking them almost guarantees that the data will be lost, sooner or later. This process does seem to involve a lot more effort than my grandparents went to when shoving photos into a shoebox in the attic decades ago, but perhaps this is one of the costs for the miracles of our digital age.

If all this seems like too much work, there is one last possibility. We could revert our digital data back to an analog form and use traditional media-preservation techniques. An extreme example of this is demonstrated by the Rosetta Project, a scholarly endeavor to preserve parallel texts of all of the world's written languages. The project has created a metal disk (*above*) on which miniatur-

ized versions of more than 13,000 pages of text and images have been etched using techniques similar to computer-chip lithography. It is expected that this disk could last up to 2,000 years because, physically, the disk has more in common with a stone tablet than a modern hard drive. Although this approach should work for some important data, it is much more expensive to use in the short term than almost any practical digital solution and is less capable in some cases (for example, it's not good for audio or video). Perhaps it is better thought of as a cautionary example of what our future might look like if we are not able to make the digital world in which we find ourselves remain successful over time.

Bibliography

- Balistier, Thomas. 2000. *The Phaistos Disc: An Account of Its Unsolved Mystery*. New York: Springer-Verlag.
- Besen, Stanley M., and Joseph Farrell. 1994. Choosing how to compete: Strategies and tactics in standardization. *Journal of Economic Perspectives* 8:117–131.
- Camras, Marvin. 1988. *Magnetic Recording Handbook*. New York: Van Nostrand Reinhold Co.
- The IBM 709 Data-Processing System. http://www-03.ibm.com/ibm/history/exhibits/mainframe/mainframe_PP709.html
- Koops, Matthias. 1800. *Historical Account of the Substances Which Have Been Used to Describe Events, and to Convey Ideas, from the Earliest Date, to the Invention of Paper*. London: T. Burton.
- Pohlmann, Ken C. 1985. *Principles of Digital Audio*, 2nd ed. Carmel, Indiana: Sams/Prentice-Hall Computer Publishing.
- The Rosetta Project. <http://www.rosetta-project.org>
- United States Postal Service, Domestic Mail Manual 708.4—Special Standards, Technical Specifications, Barcoding Standards for Letters and Flats.

Enjoy the rewards.

WorldPoints  Now everyday purchases can add up to rewards. The WorldPoints program lets you choose from among great rewards like cash, travel, brand-name merchandise, and gift cards for top retailers.* Use your Sigma Xi, The Scientific Research Society Platinum Plus® Visa® card with WorldPoints® rewards, and you'll enjoy around-the-clock fraud protection, free additional cards for others you trust, and quick, secure online access to your account.



NO ANNUAL FEE | SECURITY PROTECTION | ONLINE ACCOUNT MANAGEMENT

To apply, call toll-free **1.866.438.6262**

Mention Priority Code UAA6PC. You can also visit www.newcardonline.com and enter Priority Code UAA6PC.

Bank of America 

For information about the rates, fees, other costs and benefits associated with the use of this Rewards Card, or to apply, call the toll free number above, visit the Web site listed above or write to P.O. Box 15020, Wilmington, DE 19850.

* Terms apply to program features and Credit Card account benefits. For more information about the program, visit bankofamerica.com/worldpoints. Details accompany new account materials.

This credit card program is issued and administered by FIA Card Services, N.A. The WorldPoints program is managed in part by independent third parties, including a travel agency registered to do business in California (Reg. No.2036509-50); Ohio (Reg. No. 87890286); Washington (6011237430) and other states, as required. Visa is a registered trademark of Visa International Service Association, and is used by the issuer pursuant to license from Visa U.S.A. Inc. WorldPoints, the WorldPoints design and Platinum Plus are registered trademarks of FIA Card Services, N.A. Bank of America and the Bank of America logo are registered trademarks of Bank of America Corporation. All other company product names and logos are the property of others and their use does not imply endorsement of, or an association with, the WorldPoints program.

© 2010 Bank of America Corporation

AR96896-110909

AD-01-09-0012.C.WP.NT.0109



✓Yes



✓Yes



✓Yes



✓Yes



✓Yes



xNo



✓Yes



✓Yes

✓ Everyday Low Prices
✓ Easy To Use Website
✓ Huge Selection
✓ Fast Shipping



ALL THE PARTS YOUR CAR WILL EVER NEED

GO TO WWW.ROCKAUTO.COM ROCKAUTO, LLC (EST. 1999) 

ENGINEERING

Challenges and Prizes

Henry Petroski

THERE IS NO SHORTAGE of challenging engineering problems, and many of the more pressing of them these days have to do with the development of new and renewable energy sources or with the more efficient and environmentally friendly use of old ones. But solving these and related problems can be maddeningly difficult, for the world of energy production, transmission, storage and consumption is complex, and its component parts are systemically interrelated. This is also the case with non-energy related problems ranging from providing clean drinking water worldwide to carrying out efficient space exploration. In order to encourage research and development work on tough problems of all kinds, an increasing variety of incentives has come to be employed, including formal challenges and lucrative prizes. But meeting a challenge and walking away with a prize can be only the beginning of what might prove to be a long and convoluted engineering development project.

Energy Innovations

Large-scale wind power may appear to be promising, but it presents some persistent difficulties. Perhaps first and foremost is the need to provide transmission lines from remote areas where the wind blows strongly—and where neighbors do not object to the location of massive wind farms—to urban areas where vast amounts of electricity are consumed. But long transmission lines equal energy losses that in effect lower the efficiency of the wind turbines. Also, since the wind is notoriously fickle, it cannot be counted on to produce a

*Incentives help
motivate solutions
to humanity's needs
and wants*

steady output of power. Thus, a means of storing excess electricity when the wind howls and of releasing it during calms is very desirable. Batteries are a familiar technology for performing this task, but installing dedicated banks of batteries would be an expensive and far-from-elegant solution. Thus, some more creative proposals exploit the growing interest in plug-in electric vehicles.

Although the battery packs of such vehicles draw power from the grid when they are being recharged, great numbers of them plugged in can also provide a means of stabilizing the grid, by drawing from it when power production surges and releasing to it when production dips. Millions of plug-ins connected to the grid could thus play the role of a flywheel, alternately storing and releasing energy as needed. A fully effective symbiotic relationship between power production, consumption and storage is still years in the future, however, because electric vehicles still have developmental problems of their own. In addition, car owners might have to adapt to a charging regimen compatible with other demands on the electric grid, which at the moment at least may not be smart enough to cope with millions of electric cars being plugged into it simultaneously.

Even though battery technology is centuries old—the word “battery” for a series of electrical storage devices was coined in the 18th century by Benjamin Franklin, who saw the analogy with an

assemblage of artillery pieces—truly perfecting the technology has proved elusive. The heavy and bulky lead-acid batteries in conventional automobiles need replacing every five years or so, when they can no longer hold a charge. The lithium-ion batteries that power so many laptop computers are more compact, but they are still relatively heavy and expensive, and some have been known to burst into flame.

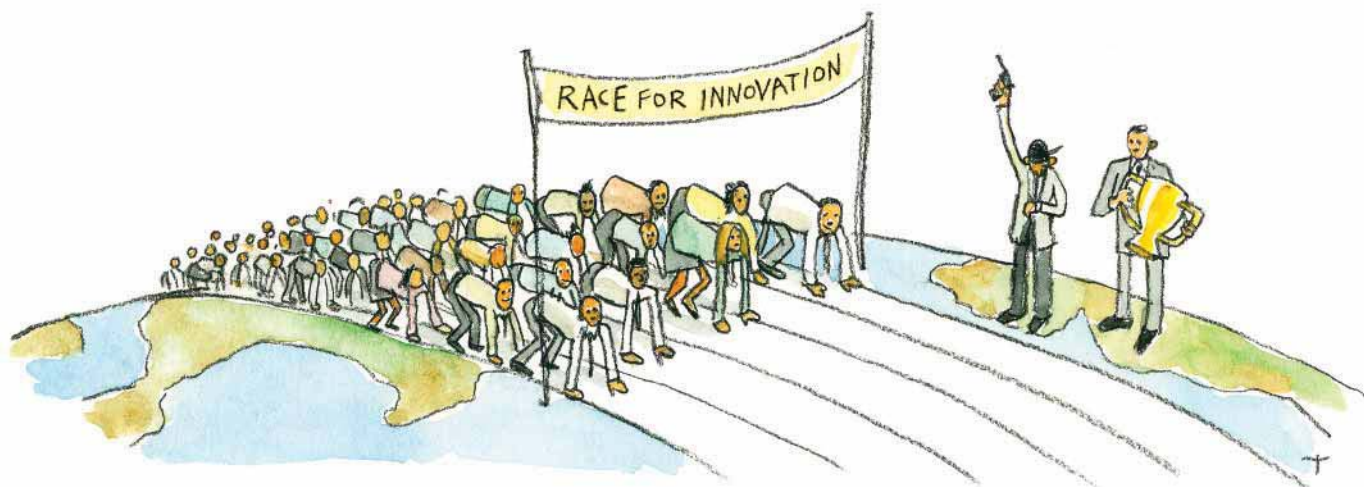
Still, lithium-ion batteries seem to be the devices of choice for powering all-electric cars. Unfortunately, it takes a goodly number of cells to pack sufficient energy to drive an automobile a reasonable distance before recharging. (Gasoline has a much higher energy density than a storage battery.) Tesla Motors has been making and selling in limited quantities its all-electric Roadster, but the \$109,000 price tag keeps it out of reach of virtually all but the rich and famous. Tesla's forthcoming Model S luxury sedan is expected to be priced at about \$60,000, which is still quite a premium to pay for a silent ride. Among the principal reasons for such prices is that the electric car's one-ton battery pack itself will cost about \$10,000.

The fuel cell, the basic scientific principles of which have been known for well over a century and which has been used in space vehicles for some time, promises to be an alternative to the automobile battery—when its price drops and when problems relating to the production, availability and distribution of its fuel (typically, hydrogen) are resolved. It has become somewhat of a standard joke among technology reporters that each year a practical fuel-cell technology is still only 10 years away.

Seeing the Light

The incandescent light bulb, which has been around for well over a century, is not significantly different from Thom-

Henry Petroski is the Aleksandar S. Vesic Professor of Civil Engineering and a professor of history at Duke University. His latest book, The Essential Engineer: Why Science Alone Will Not Solve Our Global Problems, was published in February. Address: Box 90287, Durham, NC 27708-0287.



as Edison's invention. But this virtual symbol of inspiration and creativity is notoriously inefficient when it comes to converting electricity to illumination. In 2007, Australia instituted a ban on the so-called filament bulbs, and the ban goes into effect this year. The European Union, also seeking to conserve energy and thereby reduce the amount of carbon dioxide released to the atmosphere, has established a similar ban. The United States will begin to phase in a ban on filament bulbs in 2012.

A ban on incandescent light bulbs would not be practical without alternative lighting technology, of course, and it was the compact fluorescent bulb that was expected to be the standard replacement. The compact fluorescent was developed at General Electric during the energy crisis of the 1970s, but manufacturing difficulties kept GE from pursuing commercialization at the time. Other companies did eventually pursue the new bulb, however, and before long it was commonly encountered in hotel rooms. At first, the unfamiliar bulb that did not light up immediately the way its predecessor did was confusing to use. Also, the color of the light it threw off was cool, unconventional and unflattering, and therefore was criticized. However, improvements in the technology and the promise of energy savings enabled the compact fluorescent gradually to gain a foothold in the marketplace, especially as its price began to drop.

Many of the cheaper compact fluorescents have been manufactured abroad, and their quality control can be poor. Bulbs burned out well before consumers could recoup in lower electricity costs the higher prices they paid

for the corkscrew-shaped fluorescents. Furthermore, it became widely known that the bulbs presented an environmental hazard, as they contained mercury in vapor form that escaped when the glass envelope was broken. This made disposing of the bulbs problematic, and to ameliorate the negative publicity retail outlets that sold them instituted special programs to collect spent bulbs for safe disposal.

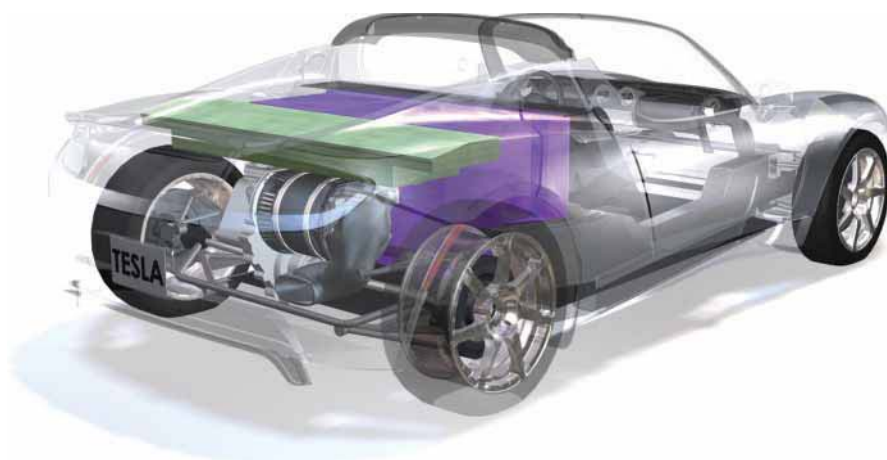
In the meantime, the light-emitting diode (LED), hailed as "the most efficient lighting source available" began to be employed increasingly in commercial lighting applications, where the high capital investment could be most easily justified. With incandescent bulbs being outlawed and compact fluorescents posing a hazard, the stage was set for the LED to become the replacement technology of choice. Philips Lighting, one of the leading manufacturers of compact fluorescent bulbs, redirected its research and development programs from them to LEDs, which had some problems of their own to overcome. In particular, they produced more concentrated heat than compact fluorescents, and so the newer bulbs had to be designed with fins to radiate heat away from their base. This led to bulb designs that presented problems for interior decoration. Finally, the bulbs are even more costly than compact fluorescents; this will no doubt eventually lead to cheaper imitators—and to inferior products.

Grand Challenges

If such familiar and seemingly simple technologies as storage batteries and light bulbs can present such convoluted engineering challenges, then how much more difficult must we expect

it to be for engineers to solve more complex problems relating to energy production, storage, distribution, conservation and use. And what of problems not related directly to energy? In order to identify the most challenging and consequential of those problems, a few years ago the National Academy of Engineering appointed a committee of engineers, scientists and inventors to compile a list of "opportunities that were both achievable and sustainable to help people and the planet thrive." These opportunities have come to be known as engineering's grand challenges of the 21st century, and meeting the challenges is not expected to be either quick or easy. The 14 challenges identified fall into "four themes that are essential for humanity to flourish—sustainability, health, reducing vulnerability, and joy of living."

Though the list is unranked, the first challenge mentioned is to "make solar energy affordable." The sun is being harnessed to generate electricity today, but generally at a high cost per kilowatt-hour. Solar cells are expensive to manufacture, and mirror configurations that focus the sun's rays to concentrate their heat on pipes or boilers generally require a lot of land and water resources for their effective operation. One solar farm proposed for Amargosa Valley, Nevada, reportedly would consume 20 percent of the available water in that desert location. Where water is not used to generate steam, it is used to wash dust and dirt off solar panels and mirrors in order to maintain efficiency. Furthermore, solar, like wind energy, also needs to be paired with a backup or an energy storage system, such as batteries,



A large number of lithium-ion cells are undoubtedly part of the reason the Tesla Roadster electric car has a sticker price of \$109,000. At the same time, these cells are in part responsible for its supercar level of performance. The upcoming Model S luxury sedan will be somewhat less expensive. (Image courtesy of Tesla Motors Inc.)

which we have seen present their own developmental challenges.

Other sustainability grand challenges are to “provide energy from fusion,” to “develop carbon sequestration methods” and to “manage the nitrogen cycle.” Fusion energy has been a holy grail for the past half century or so, and it promises to remain elusive for the foreseeable future. The capture and sequestration of carbon produced by coal-burning power plants is possible in scientific principle but undemonstrated in full-scale engineering reality. Even if the unproven technology can store carbon in deep underground rock formations, it may pollute groundwater supplies as a byproduct. Some scientists deem nitrogen to be as threatening, if not more so, as carbon dioxide to the atmosphere and the planet’s climate, but controlling nitrogen gases without having adverse impacts on the food supply involves many unknowns. These are clearly complex problems.

The grand challenges relating to human health include to “provide access to clean water,” to “restore and improve urban infrastructure,” to “advance health informatics,” to “engineer better medicines” and to “reverse-engineer the brain.” Clean drinking water is essential to good health, but worldwide there are problems with aquifers contaminated with arsenic and other naturally occurring poisons, as well as by manmade pollution. Urban areas may not depend on water wells, but the lead pipes and aging cast-iron mains that constitute the distribution network can be the source of contaminants and failures. The problem of modernizing

an aging infrastructure alone is enormous in magnitude. But keeping water supplies clean and sewers flowing does not remove the need for efficient and effective health care. Informatics of all kinds, from medical monitoring devices to record keeping, must be kept up to the task. The engineering of better medicines will be necessary to combat persistent diseases and conditions. It is a little-acknowledged fact that a lot of engineering lies behind the targeted delivery of effective drugs, such as those that maintain their potency while they zero in on a tumor. The use of engineered materials and nanotechnology is making such achievements possible, but like most problems in medicine and health care, their full realization will take time. Likewise, what may be the supreme challenge—reverse engineering the brain—cannot be expected to be accomplished easily or quickly, but its achievement will provide tremendous insight into learning processes and artificial intelligence and how to treat conditions ranging from the psychiatric to the neurological.

Advanced health informatics are also essential for responding to world-population threatening conditions. The grand challenges relating to reducing vulnerability are to “prevent nuclear terrorism” and “secure cyberspace.” Nuclear terror is especially world threatening, of course, and seeking ways to reduce the planet’s vulnerability is clearly important. Solving this problem naturally must involve a good deal of political effort, but engineering better detection, monitoring and verification devices must also play a part.

An electromagnetic pulse remains a threat to information networks worldwide, but there are increasingly also the threats from less dramatic means, such as destructive worms and viruses in the World Wide Web, the Internet and cyberspace generally, upon which we have come to rely so much.

Food and shelter, health, and security constitute basic human needs, but humans also need and desire more out of life. This more the grand challenges committee categorized under the joy of living, which includes playfulness, lifelong learning, and other intellectual and creative pursuits. Engineers have a role to play in this area, too, and they have been challenged to “enhance virtual reality,” to “advance personalized learning” and to “engineer the tools for scientific discovery.” As in other theme areas, there can be a considerable overlap in the challenges, with enhanced virtual reality providing the technology to advance personalized learning. It has long been recognized that the tools for scientific discovery are technological as much as psychological. From Galileo’s telescope to the Hubble Space Telescope, it is advancements in technology that enable new scientific discoveries and hence new theories about everything in and including the universe. Engineering is not simply applied science, and science can benefit greatly from applied engineering.

Incentives

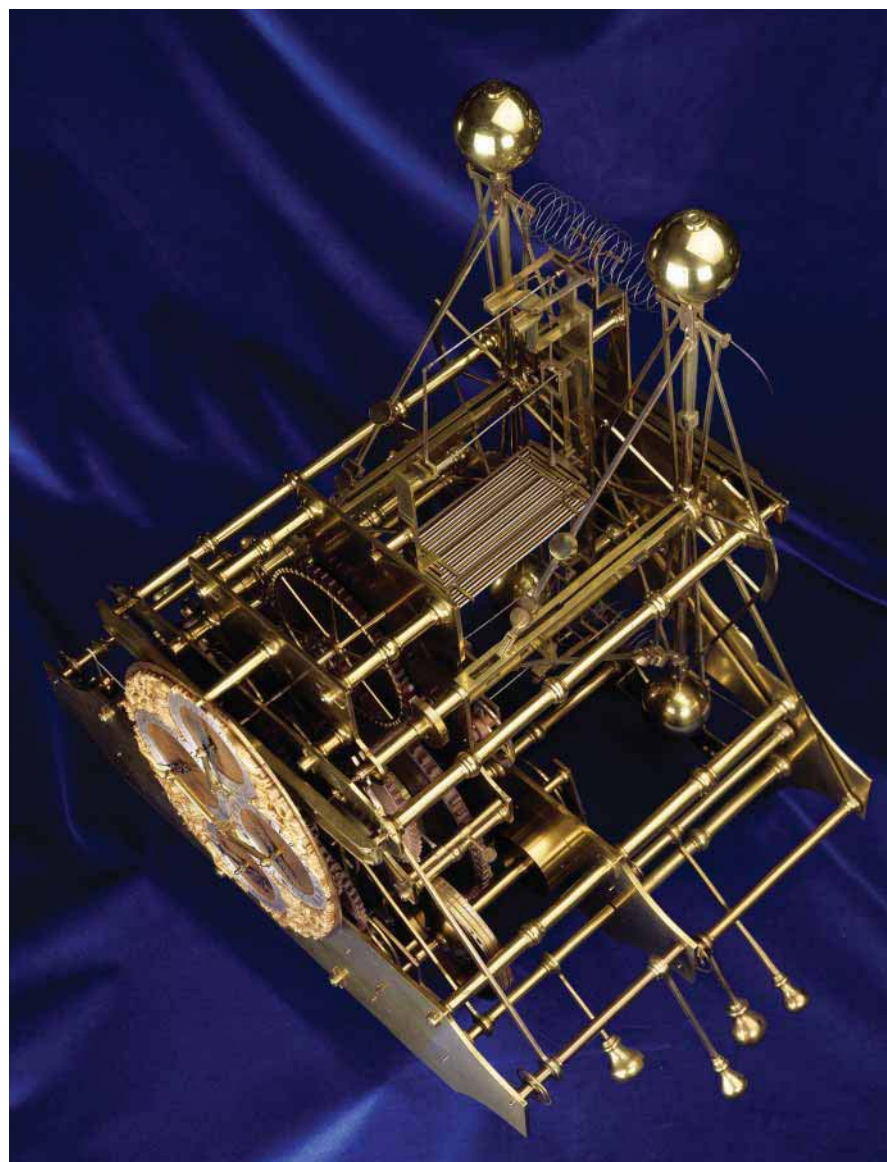
The list of grand challenges alone may or may not motivate a given engineer, team or company to work on any one of them, but there is a kind of challenge that does motivate engineers and groups alike. This is the competition or prize. Generally speaking, a design competition has a specific structure or device as its objective. This might be a bridge for a specific location, and the competition guidelines would typically spell out specific requirements that a successful design must meet. Thus, the winning bridge design may have to have a minimum clearance above mean high water, and it might have to open to let ships pass. Competition announcements and guidelines also can require that engineers work with architects or artists, and the teams may have to prequalify by establishing their credentials in bridge design. The winner and runner-up in a design competition may receive a cash award, but generally it will not be

nearly enough to cover the team's time or expenses incurred in preparing its entry. Engineers and architects enter competitions because they provide creative challenges and practice in solving real-world problems, and can result in beneficial exposure to the design and client community.

Prizes, on the other hand, tend to be for invention and innovation, and can be associated with very lucrative cash awards. Indeed, offering prizes to promote change has recently been described as "one of the most intriguing trends in philanthropy." The Nobels are, of course, the most well-known science prizes and ones that come with a sizeable honorarium. The will of chemical engineer Alfred Nobel that established the eponymous awards stated that they should go "to those who, during the preceding year, shall have conferred the greatest benefit on mankind." However, as instituted they have favored scientific advances in the specified fields of physics, chemistry and physiology or medicine, often for achievements that are decades old. It can be argued that Nobel had much more immediate recognition of engineering achievements in mind. Today, the National Academy of Engineering's annual Draper Prize does recognize Nobel-class engineering achievements, and often in advance of their recognition by the Nobel Foundation. The first Draper Prize, which was awarded in 1989, went to Jack S. Kilby and Robert N. Noyce for independently inventing and developing the monolithic integrated circuit in the late 1950s. Kilby was awarded the Nobel Prize in physics in 2000 for the same accomplishment; Noyce would no doubt have shared in the prize had he not died 10 years earlier. Charles K. Kao, who last year was awarded a Nobel prize in physics for his work in fiber optics, in 1999 shared the Draper Prize with Robert D. Maurer and John B. MacChesney for the same achievement.

Find the Longitude

The more common form of scientific or engineering prize is one that states a technological goal in prospect, rather than singling one out in retrospect. Among the most widely known of historical prizes with a specific practical goal is the Longitude Prize, which was established in 1714 by the British government. The £20,000 prize, which would amount to the equivalent of about \$5 million in today's money,



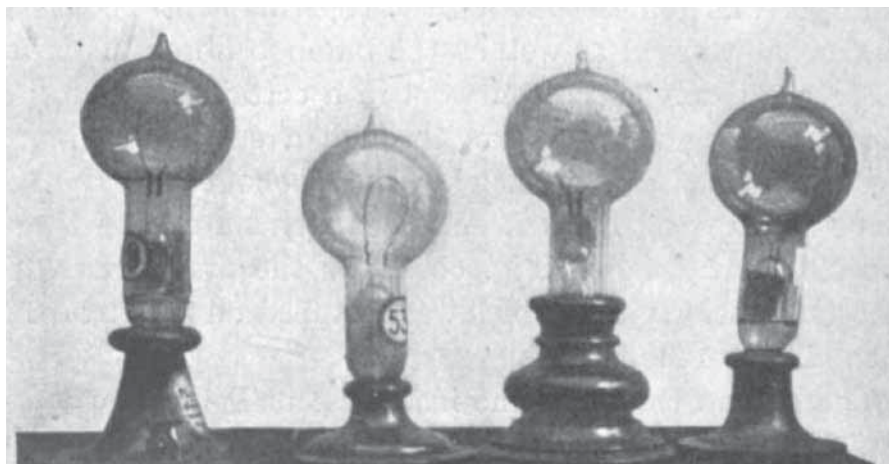
National Maritime Museum, Greenwich, London

Among the best known of prizes was the Longitude Prize, awarded to John Harrison for the development of a chronometer that remained accurate in the roughest of seas. The difference between local solar noon and the time in London correlates to longitude.

was to encourage the development of a means of determining accurately a ship's longitude at sea. A Board of Longitude was set up to administer the prize and judge entries, and it received "more than a few weird and wonderful suggestions," some of which involved squaring the circle and perpetual motion machines. In fact, because so many people doubted that there was a solution to the problem, the phrase "finding the longitude" came to be identified with "the pursuits of fools and lunatics." Neither fool nor lunatic, the clockmaker John Harrison devoted much of his professional life to pursuing the prize. His product was a chronometer that remained accurate in the roughest of seas, thus telling a ship's captain the

exact time in London when the sun was directly overhead at sea. This correlated with longitude, which is essentially the distance from London. However, the Board awarded the full prize money to Harrison only after he had petitioned Parliament.

Another historic prize for achievement was the Orteig Prize, with a purse of \$25,000 offered by a French-born hotel owner "for the first nonstop aircraft flight between New York and Paris." Many a life was lost in pursuit of the Orteig before Charles Lindbergh achieved the feat in 1927 in his single-engined *Spirit of St. Louis*. Today, there is a growing number of prizes—with growing purses—for meeting stated challenges ranging from more efficient



The current incandescent light bulb differs little from those developed by Thomas Edison (left). One candidate for an energy-efficient replacement comes from Philips Lighting (right). If accepted by the U.S. Department of Energy, it could win the \$10 million L Prize. (Photograph at right courtesy of Rick Friedman, copyright 2009.)

batteries and light bulbs to moon landings and reusable space ships.

Modern Prizes

When John McCain was running for president in 2008, gasoline was approaching \$5 a gallon in California, so he proposed to “inspire the ingenuity and resolve of the American people by offering a \$300 million prize for the development of a battery package that has the size, capacity, cost and power to leapfrog the commercially available plug-in hybrids or electric cars.” The amount of the award might appear to be enormous, but McCain reminded his audience that it represented only one dollar per capita, which he considered to be “a small price to pay for helping to break the back of our oil dependency.” The candidate’s idea may have been suggested by the Wearable Battery Prize sponsored by the Pentagon for a device that would weigh less than nine pounds and provide at least 96 hours of uninterrupted power for soldiers in the field, who had to carry around as much as 20 pounds of conventional batteries to run such things as their radios, computers and night-vision goggles. The Pentagon was offering a \$1 million first prize, with lesser amounts for second and third prize, but these amounts are not likely to cover anywhere near the cost of a successful research and development program.

The energy-inefficient incandescent light bulb is the motivation behind the L Prize, which is sponsored by the Department of Energy. Among the criteria a challenger must meet to win up to \$10 million is to come up with a new kind of bulb that must match the color

and amount of light given off by a 60-watt conventional incandescent. Furthermore, in doing so the winning bulb must consume only 10 watts of power and last more than 25,000 hours. In addition, at least three-quarters of the bulb must be manufactured in America. Last September, the first bulb to be entered into the contest was one made by the Philips Lighting company, which is headquartered in the Netherlands. The Department of Energy was expected to take as much as a year to fully test the entrant bulb (in spite of the fact that a year contains only about a third of the 25,000 hours that the bulb is supposed to work). Even though \$10 million in prize money cannot be expected to cover R&D expenses, winning a contest like the L Prize can be a boon to a manufacturer because the achievement can be expected to give the winner a marked advantage in receiving government contracts for vast quantities of the product. In addition, the bulb would have a distinct advantage in the retail market.

Among the most publicized of recent prizes have been a series of X PRIZES. The original X competition was the Ansari X PRIZE, which was for launching the first privately financed reusable spacecraft that would carry “three adults to an altitude of 100 kilometers, twice within two weeks.” The prize was won in 2004 by SpaceShipOne, which was designed by a team headed by the aerospace engineer Burt Rutan. The second-generation “reusable spaceliner” SpaceShipTwo is part of the business plan of the space tourism firm Virgin Galactic, which has proposed to offer to carry civilians into space for \$200,000 a ticket.

Other X PRIZES include the Google Lunar X PRIZE, for landing “a rover on the moon that will be able to travel at least 500 meters and send high resolution video, still images and other data back home.” The winner of the prize will receive \$20 million, and there is the possibility of a \$5 million bonus for traveling 10 times as far or for transmitting images of an artifact left behind from the Apollo program. Perhaps more down to earth, the Progressive Automotive X PRIZE is defined as an “international competition designed to inspire a new generation of viable, super fuel-efficient vehicles.” The \$10 million purse is held out to promote “revolution through competition.”

Recent years have seen a proliferation of challenges and prizes designed to address global problems and improve the quality of life, and these challenges and prizes promise to inspire, prod and reward those engineers, inventors and entrepreneurs who choose to pursue them.

Bibliography

- Belfiore, Michael. 2009. *The Department of Mad Scientists: How DARPA Is Remaking Our World, from the Internet to Artificial Limbs*. New York: HarperCollins.
- National Academy of Engineering. 2008. *Grand Challenges for Engineering*. Washington, D.C.: National Academy of Sciences.
- Petroski, Henry. 2010. *The Essential Engineer: Why Science Alone Will Not Solve Our Global Problems*. New York: Alfred A. Knopf.
- Sobel, Dava. 1996. *Longitude: The True Story of a Lone Genius Who Solved the Greatest Scientific Problem of His Time*. New York: Penguin Books.
- Taub, Eric A. 2009. A bright idea: build a better bulb, win \$10 million. *New York Times*, September 25, pp. B1, B4.

MARGINALIA

Two Lives

Roald Hoffmann

I MEET ANSGAR BACH first through his writing. He sends me an endearing account, in German (which I can read, because I was once a refugee in post-World War II Germany), of a trip he took to New York. In the middle of the 47th Street Diamond District, Ansgar spots an oasis (sadly gone today): The Gotham Bookmart. In it he finds my second poetry collection, *Gaps and Verges*, signed. The bookstore sign says “Wise men fish here.” Ansgar fishes, writes about it, sends me what he writes and catches a new friend.

Books mean much to Ansgar. I discover that when I do what can only be done today, an Internet search on his name. From the many sons of Bach and still more Bach festivals, I excavate an interesting fact: He is an unusual chemist. For although he does at this time have an institutional affiliation—the Department of Chemistry at the Free University of Berlin—he also runs a small business: *Literarisch Reisen*. The phrase is resonant; in one sense it means “Travel in Literary Fashion.” And indeed Bach’s firm organizes excursions in his region of Germany that follow the footsteps of Thomas Mann or visit the sites of a series of stories by E.T.A. Hoffmann or Heinrich Heine. Very German. Very literary. And most unlike what chemists do.

Ansgar is on a path quite different from that of most academic scientists. Like him, like many other scientists, I have wide-ranging interests that include literature and art. But can one professionally embrace those multiple interests? Can one successfully pursue scientific research only on a part-time basis, with the rest of one’s attention focused elsewhere, no matter how gripping or worthwhile the other pursuits might be?

Roald Hoffmann is Frank H. T. Rhodes Professor of Humane Letters, Emeritus, at Cornell University. Address: Baker Laboratory, Cornell University, Ithaca, NY 14853-1301

Scientists do any number of things, besides science

Crossovers

Ansgar and I continue to correspond. He sends me a detective story he has published, *Ukrainische Verbindung* (*The Ukrainian Connection*). In German such stories are called *Krimis*, for *Kriminalromane*. The story, quite an exciting one, begins in a German paint factory, and part of the action takes place in the Ukrainian city of L’viv. Another connection, for I was born in Złoczów, some 60 kilometers from L’viv! My father had gone to the Lwów Polytechnic University, as it was called in Polish days. Earlier, during the Austro-Hungarian days, L’viv was Lemberg. A crossroads of the world it was, our historic region of Galicia. And it was also a place for waves of ethnic cleansing.

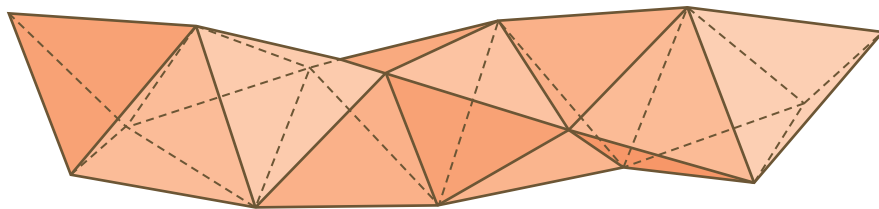
One day in 2002 Ansgar comes to visit. He is spending a month in a lab at the State University of New York at Buffalo, practically next door. Only snow divides us. He is a young man in his mid-30s, with what I might call a Ringo Starr haircut. He has an easy smile, a gentle voice, and is unnecessarily timid about his more-than-adequate English. True, he is in a minority of German scientists who have not

The logo of *Literarisch Reisen*, a German company that organizes tours to places associated with literary figures, shows Friedrich Schiller, a German writer and philosopher, riding a donkey. The engraving is by Johan Christian Reinhardt from about 1787. The portrayal is not without chemical interest, because in some drawings the figure is mirrored. Which one is the real writer, and which is the reflection? The same question also arises often for molecules.

done a postdoctoral year in the United States. The center shifts; there was once a time when every U.S. chemist went to Germany; now they come here. Twenty-two German postdoctoral associates—or postdocs—have spent a year or more in my group.

Ansgar gives a lecture about his work. The talk is not about what he did his doctoral research on, but what he does now, crystallography. I know about crystallography by “osmosis” because I got my Ph.D. in the lab of a great crystallographer, William N. Lipscomb. I did not do crystallography; I was simply in daily contact with clever people who were learning and practicing the technique. I also knew of it out of necessity, because in my work as a theoretician explaining molecular shape, I have had to make judgments





A chain of face-sharing tetrahedra forms a gently curving triple helix. This pleasing structure was given the name “tetrahelix” by R. Buckminster Fuller. The structure appeared in his book *Synergetics*, and has inspired several artistic sculptural pieces around the world. Chemical compounds that take on this structure can also be made.

as to which seeming structural anomalies are worth pursuing, and which are to be disbelieved.

What I will say next about crystallography is not anything a crystallographer would say. When the experimental technique was difficult and a crystal structure took a year to do, there was no problem; everyone wanted to have a crystallographer friend.

As crystallography became easier, an almost routine technique, the field went in search of a *raison d'être*. Some practitioners took on complexity—proteins, for instance. Others looked at large groups of molecules for trends and regularities (these researchers I especially value; I have written previously about them (“Crystal Cloudy, Crystal Clear,” *American Scientist* 86[1]:15–18, January–February 1998). Such crystallographers do just what I did in my theoretical work. Still others got the technique better and better, so that they could see not just where are the atomic nuclei and electrons near them, but also where in space reside the chemically important electrons involved in bonding and reactivity. Such a direction brings this subset of crystallographers in contact with theoreticians, who compute such things.

This is exactly the research being done both in Ansgar’s group in Berlin and in the one he is visiting in Buffalo. For my own prejudiced reasons I’m not too crazy about the work, but I won’t bore you with my prejudices. Suffice to say that when Ansgar visits Cornell, I give my guest a politely hard time during his seminar. He handles it well, as he does a pretty disconnected set of questions from the only two professors who find time to come to his talk, one of two that day, six that week. The rest of the audience is students who, as usual, sit quietly.

A Molecular Tetrahelix

At dinner that evening, over a bottle of Corbières, Ansgar tells me part of his story. He always loved chemistry. And he always read; German literature was close to him. Originally from Cologne, Ansgar did his Ph.D. at the Free University of Berlin in the group of Hans Hartl. Now that was a name I knew well. Hartl and his students had made some copper compounds whose shapes had two, three or four tetrahedra that shared faces. I saw these once

and thought, hey, why not an infinite chain of face-sharing tetrahedra? David Nelson, a physicist at Harvard who once had been a student of mine, came up with the same structure in a different context, and Chong Zheng, a brilliant student fresh out of China, set to work figuring out for which elements might such a structure be stable.

We thought we were original—until one of us saw a sculpture by Ted Bieler in front of the Marathon Realty Building in Toronto, with three such helices passing by each other. And then we saw Arata Isozaki’s 100-meter tower in Mito, Japan.

It looked like we weren’t that original. Actually, neither were the architects and sculptors (but they didn’t need to write footnotes, as we did), because this “tetrahelix” was the centerpiece of a chapter in Buckminster Fuller’s *Synergetics*!

In time, Hartl and his coworkers made the molecule, a copper iodide compound. It was as we had predicted; 12 orders of magnitude smaller than Isozaki’s tower, there it was.

Goethe and Zinc Iodide

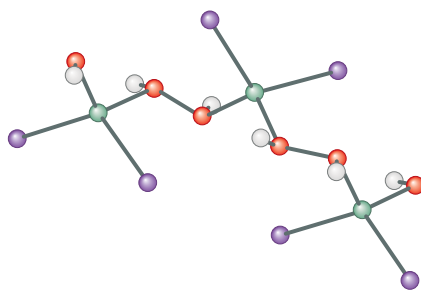
But I have been mixing my story with Ansgar’s. On finishing his Ph.D. on zinc and cadmium halides (making them and using crystallography to determine their structures), Ansgar went to work for a couple of years in a paint and coatings company, which he didn’t like. At least the company, which will remain nameless, provided the setting for one of his detective stories. I hope there were no bodies in industrial paint mixers in real German companies.

Meanwhile, Ansgar’s literary-tour business grew into a moderately successful enterprise. For a time after I met him, he was a part-time researcher—a different path indeed in a profession addicted to a permanent search for the new, and that is a time-consuming, addictive search. To remain part-time in research Ansgar had to find a sympathetic group leader—a professor who would recognize that another obsession shared the mind of the talented young scientist and give Ansgar half of his time for his literary activities. Peter Luger in Berlin did it. However, today (in 2010) Ansgar devotes all his time to *Literarisch Reisen*.

But back to our first meeting, at Cornell in 2002. At dinner, we talk of why he doesn’t “do” Goethe (as everyone else does), and we talk of Caro, Kleist, Borodin and von Arnim. In the mid-



This tetrahelix structure was designed by Arata Isozaki and built in 1990 in Mito, Japan. (Photograph courtesy of Art Tower Mito, Japan.)



Hans Hartl and Ansgar Bach made a mysterious compound, whose crystal structure is shown above. In this segment of a continuing polymeric chain, oxygen is red, hydrogen is light gray, zinc is green and iodine is purple. Note the HOOH hydrogen-peroxide groups.

dle of such literary talk, Ansgar says: "You know, what I'd really like to do is to go back to something strange we found for zinc iodide." He tells me a reaction he once ran in Hartl's laboratory, with zinc iodide (ZnI_2) and water. Out of the yellow solution came one crystal, a long, colorless needle. They "stuck it in a diffractometer," and what emerged was a structure of an inorganic polymer (above).

The ZnI_2 in the structure is unexceptional; it came into the mixture as a reagent. But where did the HOOH, hydrogen peroxide, a potent bleaching agent, come from? From the water, to be sure. But what oxidizing agent, puller of electrons, could be there, to take electrons out of the OH^- part of water and make the OH in HOOH? We were both chemists; the same question occurred to us, as it would have to Primo Levi, or as it will to every future chemist.

Ansgar doesn't know. He has also run a similar reaction in acetone, the common solvent we see as nail-polish remover. Acetone is CH_3COCH_3 . They got a linkage of the acetone units through the oxygens, and a polymeric structure through bridging with Zn_2I_4 (right).

This result is still more remarkable. "I've never seen anything like that coupling," I say. The central coupled acetone unit, $(\text{CH}_3)_2\text{C}-\text{O}-\text{O}-\text{C}(\text{CH}_3)_2$, should be a very reactive species. I begin to write mechanisms and orbitals on the paper tablecloth conveniently supplied at this favorite restaurant. Chemists cover napkins with drawings of molecules; you can always tell where they've sat.

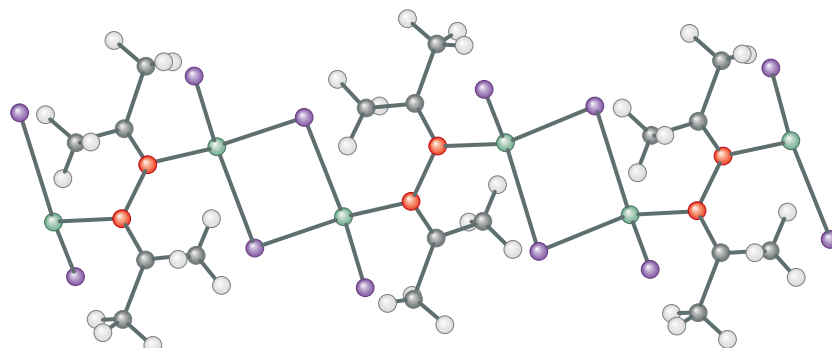
Now comes the tragedy. Ansgar says one student was able to repeat the synthesis. "But then it shut down," he says plaintively. I am not an experimentalist, but I know exactly what he meant. I know the feeling in another context—the

words falling in place after the seventh draft of a poem; an inkling of an orbital explanation. It's not a gift, it's a portal we open ourselves. It opens to the one poet in us all, as Reiner Maria Rilke wrote to Marina Tsvetaeva. Or to the one scientist who understands. And then the universe takes a jog, our attention snaps, we're out of the flow. It shuts down.

Hans Hartl, Ansgar's mentor, gives us more detail:

Waiting for ZnI_2 that we had ordered, we used some available ZnI_2 from our own chemical inventory. The compound had been bottled in a small glass flask and stored many years before. This ZnI_2 sufficed only for the first experiments, which produced some crystals of the mysterious compounds. However, it was not possible to synthesize these compounds with ZnI_2 bought or produced in our lab. We tried all sorts of experiments, for example storing the ZnI_2 in the dark, or placing it on a window with sunshine, UV irradiation and so on. For many years we regularly repeated our attempts, but all were in vain. Thus we cannot publish these compounds.

Has Ansgar made a species that immediately went extinct? An instant fossil? Maybe. Days later I ask a talented German postdoctoral associate in my group, Beate Flemmig, to do a calculation (that's our *métier*) on Ansgar's molecule. No matter what Beate does, the strangeness of the acetone coupling does not go away. She then has the bright idea of trying, instead of a $\text{C}-\text{O}-\text{O}-\text{C}$ linkage in the polymer, a $\text{C}=\text{N}=\text{N}=\text{C}$. The bonding is now normal, and the geometrical parameters fit Ansgar's compound.



An unusual compound, whose apparent crystal structure is shown here, arises in the reaction of acetone (CH_3COCH_3) with zinc iodide (ZnI_2). Carbon is shown in dark gray, oxygen is red, hydrogen is light gray, zinc is green and iodine is purple. Note the oxygen-linked acetone $(\text{CH}_3)_2\text{CO}$ units.

Ansgar thought that the plausible $\text{C}=\text{N}=\text{N}=\text{C}$ linkage could come from the reaction of acetone with hydrazine ($\text{H}_2\text{N}-\text{NH}_2$). But where did the hydrazine come from? That remains a mystery. One day someone will reinvestigate the reaction, and we will learn.

Angels in Germany

We go back to my office after dinner. There is an angel to be found, a sick student to worry about, and a lecture to write.

The angel: Carl Djerassi and I have written a play, *Oxygen*, which had its German premiere the previous year in Würzburg. Ansgar describes his visit to the play:

I arrived there just after you left, on Tuesday. The officials told me tickets were gone—no chance to get one. So I went to the box office in the theatre and got the same answer: no chance—sold out. But I stayed there, maybe about ten minutes, and a blond-haired angel appeared. She asked, "Do you want a ticket?" I said, "Thank you, you're an angel." She then said with a smile that I would meet her that night—and she was a quite attractive person. At night, shortly before the performance started, I looked around in the auditorium for the angel. I saw Carl Djerassi (my first impression of him: a Hemingway of science) who was some seats behind me. The performance began and I recognized my angel: She was playing Madame Lavoisier.

I must say that when I first came across the word angel in his email, I thought of another angel, one in Berlin: The angel who becomes a man in

Everett Collection



Bruno Ganz plays Damiel the angel, who becomes human, in Wim Wenders's classic dramatic film *Wings of Desire* (titled *Der Himmel über Berlin* in Germany). In the film, angels only see in black and white, so only scenes shot from a human's perspective are in color.

Wim Wenders's classic dramatic film *Wings of Desire*, in order to experience human love and emotion.

Ansgar's angel was female. I knew her, but I did not know her address or contact information. But I was sure that the director of the Würzburg produc-

tion would be able to put the young people in touch. How could an angel not be in favor of love?

When Ansgar comes into my office before dinner, a Cornell police officer is there talking to me. A former student of mine has had psychological problems

for a while. Troubled communications from him had crossed a line the day before; a threat was made. The police officer is there to talk about it. This also is a part of my life. Maybe it will find its way into Ansgar's next *Krimi*.

We talk into the night, of Hans Bethe and his father-in-law P. P. Ewald, one of the great figures in crystallography. Also of Hans Hellmann, perhaps the first German theoretical chemist, who was executed in the Soviet Union in 1938 (that's another story). And of that strange structure again.

But I have to send Ansgar home to his hotel. For the next morning, there is a class to be taught—a "World of Chemistry" general-education course, science for the citizen. The class tells young people about chemistry, its connections to culture, how chemists think. I should have told them Ansgar's story. But I didn't have the courage; I spoke of water instead.

Acknowledgement

I am grateful to Ansgar Bach for telling me his story, to him and Hans Hartl for allowing me to quote from their correspondence, and to Beate Flemmig for the calculations mentioned in the text.

Bibliography

Literarisch Reisen. www.literarisch-reisen.de



"...WITH SOLAR FLARES, METEOR SHOWERS AND COSMIC DUST... OOPS... YOU WANT THE WEATHER ON EARTH..."

Science Observer

Amplifying with Acid

More carbon dioxide in the atmosphere means a noisier ocean

Carbon dioxide has gained notoriety as a “greenhouse gas”; it’s one of the major waste products from human industrial activities that contribute to climate change. However, the gas that we release into the atmosphere is also absorbed into the oceans at a rate of about a million tons per hour. Seawater reacts with carbon dioxide to form carbonic acid, decreasing the pH of the oceans. This outcome has its own environmental impacts, such as damage to coral reefs and aquatic-animal respiration, but it also has a secondary consequence: It decreases the ocean’s ability to absorb low-frequency sound.

Oceanographers Tatiana Ilyina and Richard Zeebe of the University of Hawaii, along with geochemist Peter Brewer of the Monterey Bay Aquarium Research Institute in California, report in the December 20 issue of *Nature Geoscience* that lowering the pH of the ocean by 0.6 units could decrease underwater low-frequency sound absorption by more than 60 percent. “Ocean acidification is not only affecting the chemistry of the ocean, but it also affects the basic physical properties,” says Ilyina.

The ocean surface’s average pH is currently estimated to be around 8.1, and to have dropped from about 8.2 since around 1800, before the industrial revolution took off, says Zeebe. A reduction of 0.1 units does not sound like much, but pH units are on a logarithmic scale, so a drop of one unit corresponds to a tenfold increase in acidity.

Using projections of fossil-fuel CO₂ emissions over the next century from the Intergovernmental Panel on Climate Change (IPCC), the researchers calculated changes for seawater pH at the surface and at a depth of 1 kilometer, along with the corresponding changes in sound absorption at several frequencies below 10 kilohertz. In the IPCC’s “mod-

erate” scenario, in which CO₂ emissions remain at a constant level, pH drops by about 0.6 units at the ocean’s surface, and by about 0.2 to 0.4 units at depth.

The corresponding lowering of sound absorption depends on location and frequency. At a frequency of about 200 hertz, the drop ranges from about 10 to about 50 percent. Across all frequencies, the change is largest in the polar regions, because the colder water absorbs more CO₂ and thus has a greater pH change.

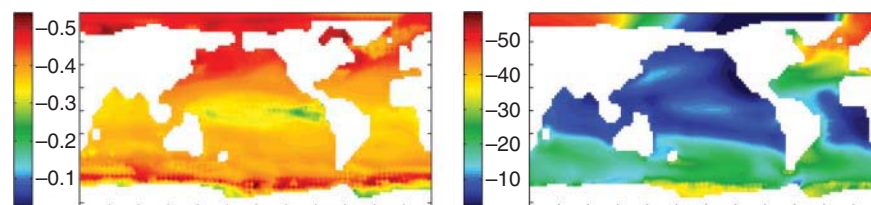
Changes in pH can impact the deep ocean because at about 1 kilometer down, the properties of temperature and pressure combine to produce a “channel” of water in which sound can propagate for many thousands of kilometers. Whales and other marine life make use of this channel for long-range communication. Most human-made ocean noise forms at the surface, but it can reflect and refract down into this channel as well.

Although the vast majority of sound loss in the ocean is due to distance, reflections and turbulence, the pH-dependent component of the ocean’s sound absorbance comes from resonance reactions in natural salts, namely boric-acid compounds and magnesium sulfate. The reaction is similar for both, but it’s more straightforward in magnesium sulfate, says Brewer: “The magne-

sium ion and the sulfate ion are attracted to each other—in human terms it’s like they’re dating—and in their normal state they exist with a single water molecule between them, like a courting couple would have a chaperone. When a sound wave comes through, it tends to squeeze that group together and the water molecule pops out, so our attracted couple just touches, ever so briefly. When the sound wave passes by, the water molecule jumps back in and separates the pair. And the work done to do that robs the sound wave of some energy.” The problem is that as the ocean becomes more acidic, the ionized form of borate decreases, so there is less of the salt form to resonate and absorb sound.

Brewer emphasizes that this decreased-absorption effect is confined to a relatively small range of frequencies, between about 100 hertz and 10 kilohertz. He estimates that the effect will be most strongly felt around 200 to 600 hertz, over distances of roughly 100 miles. “We’re talking 40 percent of a small effect, so it isn’t a lot,” he says. “On the other hand, 40 percent is a big number in itself; if any species is sensitive in that range, they would notice the change in that scale.”

The affected range includes a large proportion of the frequencies used by marine organisms. Also, most human-generated ocean noise is in the range of 10 hertz to 1 kilohertz, and the volume is rising: The biggest component is shipping, and the number of ships worldwide has approximately doubled



Carbon dioxide in the atmosphere is absorbed into the ocean, where it reacts with seawater to form carbonic acid and lower pH levels. The projected difference in ocean surface pH between 1800 and 2100, based on static carbon-dioxide emission levels, ranges up to a drop by 0.6 units (left). The corresponding decrease in the deep ocean’s sound absorption at a frequency of 200 hertz ranges up to 60 percent, depending on latitude (right). (Images courtesy of Tatiana Ilyina, Richard Zeebe, Peter Brewer and *Nature Geoscience*.)

over the past 40 years. The researchers calculate that there could be “acoustic hotspots” that are most sensitive to changes in sound propagation, such as areas at the more extreme latitudes that also experience a lot of shipping.

“This effect has been off the radar screen, so to come along and say ‘hey, what about this’ is important,” says Brewer. “It means there are new ways of looking at the Earth, it means we are nowhere near to running out of

things that are going to change. We’re fairly far down this greenhouse-gas road, and we’re nowhere near to knowing what’s going to happen to us. It’s a strange new world we’re getting into.”—*Fenella Saunders*

Sunburned Ferns?

Optical physics provides the antidote to a gardening myth

The following question appeared on an exam for Hungarian students in 2006:

In summer, at midday sunshine, it is inadvisable to water in the garden, because the plant leaves burn. Which is the only correct explanation for this?

It would be hard to provide a correct answer—the question itself is incorrect. But it reflects a widely held horticultural belief: that watering plants in the middle of the day causes sunburn. “This is an old environmental optical problem,” says Gábor Horváth of the environmental optics laboratory at Eötvös Loránd University in Budapest. And no one had tried to solve it.

Horváth and his colleagues decided to do so. They describe their results in a paper published online in *New Phytologist* on January 8. To test what happens when rays of sunlight pass through droplets on leaves, they covered maple (*Acer platanoides*) leaves with small clear glass beads and exposed them to direct sun. When they scanned the leaves, sunburned spots were clearly visible, their severity increasing with the amount of time the leaves had been exposed.

As is often the case with persistent myths, the idea of plants getting sunburn “made sense.” The glass beads supported that. To find whether water would do the same, the researchers placed water drops on maple leaves, which have a smooth, non-water-repellant surface, and ginkgo (*Ginkgo biloba*) leaves, whose smooth surfaces repel water. The sets of leaves were exposed to sun at varying times and left until the drops had evaporated. Scans revealed no visible leaf burn.

Several factors account for this. Water has a smaller refractive index (a measure of the decrease in the speed of a wave when it passes into a new

medium) than glass. Water droplets are not perfectly spherical in shape; an ellipsoid shape has less refractive power and, therefore, a longer focal length, than does a sphere. And water drops come into contact with the leaf and cool it as they evaporate.

The team also created a computer simulation of how water drops focus sunlight. Using measurements of drop shape and the elevation angle of the sun, they determined the light-collecting efficiency of the drops. This allowed them to estimate the focal region of a given drop and thus determine whether refracted sunlight would focus on the leaf surface and heat it. As it happens, for water drops resting on a horizontal leaf surface, the focal region falls on the leaf surface only at about 23 degrees solar elevation—in the early morning and late afternoon, when the sun is not intense enough to cause burn.

In a final experiment, they tested whether water drops can cause sunburn on leaves whose surfaces are covered with small hairs. The group used floating fern (*Salvinia natans*) leaves for this experiment; they placed water drops on the leaves and exposed them to two hours of midday sunlight. Many of these leaves were clearly burned.

“Leaf hairs can hold a water droplet at an appropriate height above the leaf so that the droplet’s focal region can fall just onto the leaf surface” and at the same time prevent the drop from providing evaporative cooling, Horváth explains. Fortunately for the floating fern, its leaf hairs are water repellent; it is likely that drops would roll off of the leaf before they could cause much burn.

Still, Horváth advises, it’s probably best not to water hairy-leaved plants in the middle of the day—and it’s not a bad idea to avoiding watering all plants



A leaf with a smooth, water-repellant surface (*Ginkgo biloba*, left) and a leaf with a smooth but not water-repellant surface (*Acer platanoides*, right) are covered with small water droplets to determine whether exposure to the sun will burn them. The drops on the ginkgo leaves are more spherical because the leaf surfaces are more water repellent. (Photographs courtesy of Gábor Horváth and *New Phytologist*.)

at noon; doing so can introduce other kinds of physiological stress.

Humans may have something in common with hairy-leaved plants: Water droplets held by the tiny hairs on our skin might focus sunlight to cause sunburn. Horváth hopes someone will investigate. In the meantime, he is happy to have thrown some light on the subject of leaf burn. “Misbeliefs and myths rule the online literature,” he says. Raymond Lee, a meteorologist at the United States Naval Academy, agrees: “Atmospheric optical phenomena such as this present many opportunities for confusion and myth-making among generalist readers—and even a surprising number of scientists.”—*Anna Lena Phillips*

In the News

This roundup summarizes some notable recent items about scientific research, selected from news reports compiled in Sigma Xi's free electronic newsletters Science in the News Daily and Science in the News Weekly. Online: <http://sitn.sigmaxi.org> and <http://www.americanscientist.org/sitnweekly>

An Unlikely Pollinator

Normally, crickets would just as soon chew on plants as pollinate them. But on a small island in the Indian Ocean, researchers have found a plain-looking orchid (*Angraecum cadetii*) that depends entirely on a wingless cricket to help it mate. Most of the orchid's mainland relatives are pollinated by hawk moths, but no such moths live on the island. During 48 days and 14 nights of observation, researchers saw birds, cockroaches, and even a gecko visit the flowers—but only raspy crickets (*Glomeremus* sp.) removed the pollen. Whether the phenomenon is a quirk of island ecology or whether it's just been overlooked on the mainland remains to be seen.

Micheneau, C., et al. Orthoptera, a new order of pollinator. Annals of Botany (published online January 11)

No Secret Ingredient

Stradivarius violins, legendary for their rich and expressive tones, remain the standard by which newer instruments are judged. Their uniformly dense wood, or the chemical treatments it received, might contribute to the violins' unique acoustics—but their varnish most likely does not. A new chemical analysis of minute samples from five Stradivarius instruments, built between 1692 and 1720, reveals a very

common finish. The violins bear a simple base coat of oil, perhaps linseed. Atop that is an oil-resin blend, tinted with ordinary red pigments of the day: iron oxide and cochineal. If Stradivari used a rare key ingredient, his instruments have kept the secret.

Echard, J.-P., et al. The nature of the extraordinary finish of Stradivari's instruments. Angewandte Chemie International Edition 49:197–201 (January 4)

The Sudden Sea

The Mediterranean basin was practically a desert 5.6 million years ago. Then, abruptly, it became a sea. It filled with water in less than two years, when the Atlantic Ocean gushed through the Strait of Gibraltar with 1,000 times the flow of the Amazon River. The Mediterranean sea level rose some 30 feet per day. Although this deluge was preceded by thousands of years of relatively slow trickling, 90 percent of the filling happened during those last several months. Geologists knew that the Mediterranean had gone from desert to sea, but until now, they weren't sure how fast. New samples drilled from the seafloor at Gibraltar revealed the size and shape of the old flood channel, informing a more vivid reconstruction of the event.

Garcia-Castellanos, D., et al. Catastrophic flood of the Mediterranean after the Messinian salinity crisis. Nature 462:778–781 (December 10)

Pyro-Chimps

Wielding fire is a quintessentially human pursuit. But chimpanzees are pretty fire-savvy too, a discovery that hints at how our ancestors may have first come to tinker with flames. An anthropologist followed a troop of

chimps during two savannah fires in Senegal and found that the apes didn't flee as other animals did. Rather, the chimps waited until the flames drew near, sometimes within 15 meters, then casually moved on. One male displayed toward the blaze and uttered what might be a unique fire-related bark. The apes' ability to predict and avoid the bushfire is probably a prerequisite to controlling and building fires—steps that eventually happened in the human lineage.

*Pruetz, J. D. and T. C. LaDuke. Reaction to fire by savanna chimpanzees (*Pan troglodytes verus*) at Fongoli, Senegal: Conceptualization of "fire behavior" and the case for a chimpanzee model. American Journal of Physical Anthropology (published online December 21)*

Stop That Ringing!

Personalized music therapy could soothe millions of people with chronic tinnitus, or ringing in the ears. Researchers custom-edited musical recordings so that eight volunteers could listen to their favorite songs—minus the notes with the same pitch as their tinnitus. After one year of listening to the modified tunes, participants' ringing ears were quieter, and overactive regions of their brains were more normal. It appears that, when deprived of real sounds at the problem pitch, the brain learns not to "hear" the tinnitus either. Control participants listened to music that lacked randomly selected placebo frequencies, and did not benefit.

Okamoto, H., et al. Listening to tailor-made notched music reduces tinnitus loudness and tinnitus-related auditory cortex activity. Proceedings of the National Academy of Sciences 107:1207–1210 (January 19)

Diagnosing Devils

A deadly contagious cancer could drive wild Tasmanian devils (*Sarcophilus harrissi*) to extinction within decades. The cancer cells spread from one individual to another during physical contact. But where did the original contagious tumor come from? To find out, researchers compared gene expression in tumors and in several healthy tissues. The closest match was in the Schwann cells—cells that normally protect the peripheral nervous system. Biologists hope that knowing which genes are active in the tumors will help them develop tests and vaccines to protect the ailing marsupials.

Murchison, E. P., et al. The Tasmanian devil transcriptome reveals Schwann cell origins of a clonally transmissible cancer. Science 327:84–87 (January 1)

Like-Minded (Planetary) Neighbors

The hunt for Earth-like planets is heating up. Astronomers have spotted a watery planet that is 2.7 times larger than Earth and only 42 light-years away. It even orbits its star at a nearly-habitable distance. Nearly, but probably not quite: The newfound planet heats up to 400 degrees Fahrenheit during the day. It also doesn't have any land. Astronomers found this steamy world by monitoring 2,000 nearby stars for recurring faint eclipses caused by orbiting planets. This and other recently discovered small planets show that the technique is working and may soon reveal even more familiar-looking worlds.

Charbonneau, D., et al. A super-Earth transiting a nearby low-mass star. Nature 462:891–894 (December 17)

FEATURE ARTICLES

The Ultimate Mouthful: Lunge Feeding in Rorqual Whales

The ocean's depths have long shrouded the biomechanics behind the largest marine mammals' eating methods, but new devices have brought them to light

Jeremy A. Goldbogen

A hungry fin whale dives deep into the ocean to perform a series of rapid accelerations with mouth agape into a dense prey patch. On each of these bouts, or lunges, the whale engulfs about ten kilograms of krill contained within some 70,000 liters of water—a volume heavier than its own weight—in a few seconds. During a lunge, the whale oscillates its tail and fluke to accelerate the body to high speed and opens its mouth to about 90 degrees. The drag that is generated forces the water into its oral cavity, which has pleats that expand up to four times their resting size. After the whale's jaws close, the sheer size of the engulfed water mass is evident as the body takes on a “bloated tadpole” shape. In less than a minute, all of the engulfed water is filtered out of the distended throat pouch as it slowly deflates, leaving the prey inside the mouth. Over several hours of continuous foraging, a whale can ingest more than a ton of krill, enough to give it sufficient energy for an entire day.

Years ago, Paul Brodie of the Bedford Institute of Oceanography described the feeding method of fin whales as the “greatest biomechanical action in the animal kingdom.” This extreme lunge-feeding strategy is exhibited ex-

clusively by rorquals, a family of baleen whales that includes species such as humpback, fin and blue whales. Like all baleen whales, rorquals are suspension filter feeders that separate small crustaceans and fish from engulfed water using plates of keratin—the same protein that forms hair, fingernails and turtle shells—that hang down from the top of their mouths. By feeding in bulk on dense aggregations of prey, baleen whales can support huge body sizes—they count among their numbers some of the largest animals that have ever lived. Rorqual lunge feeding is especially unusual not only with respect to the tremendous size of the engulfed water mass, but also in the underlying morphological and physical mechanisms that make this extraordinary behavior possible.

Because of the logistical difficulties in studying rorqual lunge feeding deep in the ocean, our knowledge of this ingestion process, until recently, has been limited to observations made at the sea surface. Over the past several years, my colleagues and I have made significant advances in understanding how lunge feeding works. Our collective effort has been motivated by unique data generated by digital tags attached to the backs of lunge-feeding rorquals. These tags have enabled us to quantify the particular body movements that rorquals undergo during a lunge-feeding event. With these data we have been able to determine the physical forces at play during engulfment and also to estimate the magnitude of the water mass taken in. In doing so, we have confirmed many predictions previously

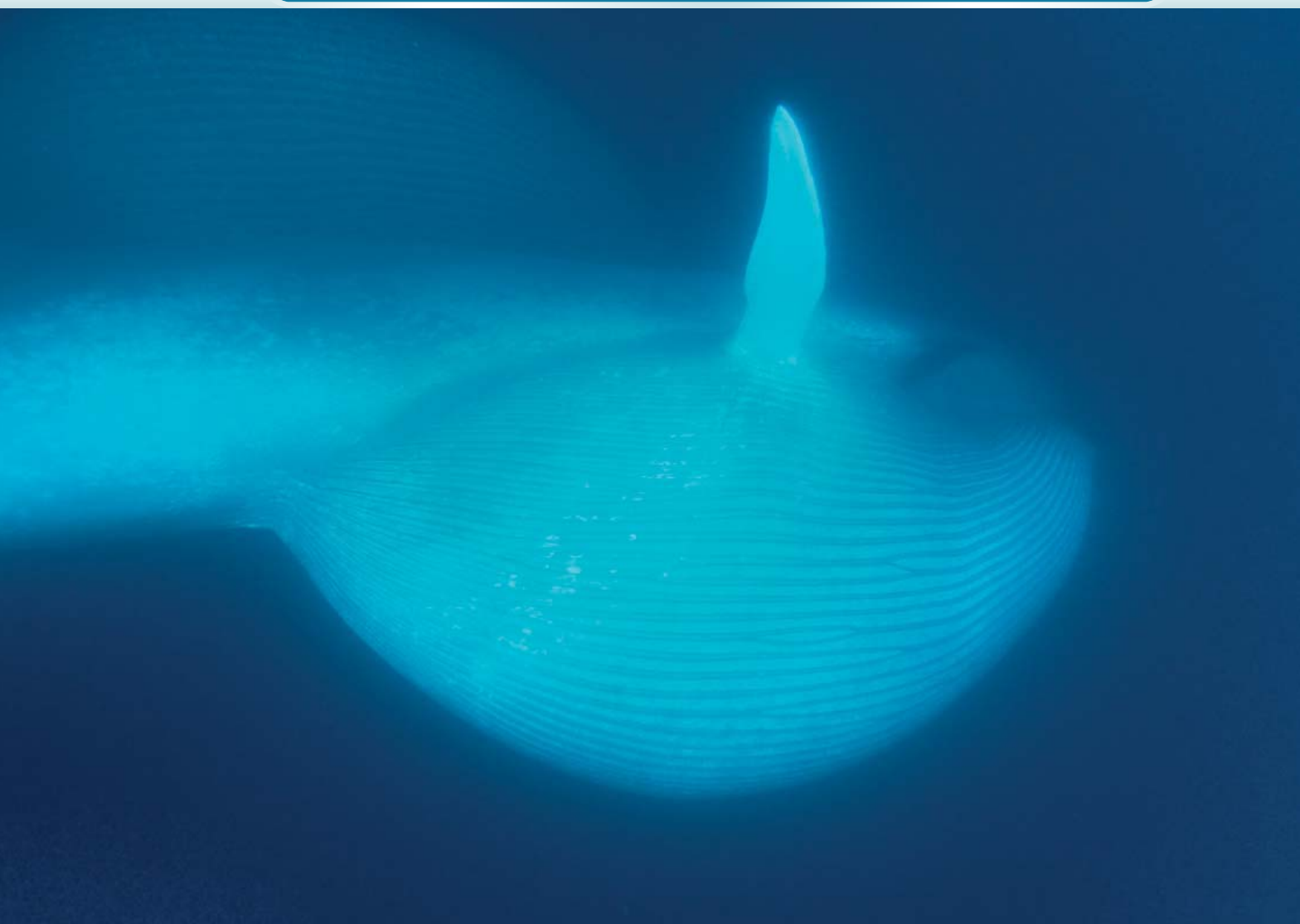
made by early investigators that were based only on anatomical knowledge and sea-surface observations. Moreover, our analyses have uncovered new engulfment mechanisms, which, in turn, have led us back to studying the remarkable morphological adaptations that drive the lunge-feeding process.

Big Heads and Inverting Tongues

Our first insight into how lunge feeding works came in large part from the pioneering studies of August Pivorunas, Richard Lambertsen and Paul Brodie over the past several decades. Their investigations focused on the anatomical machinery that makes lunge feeding possible. Rorquals exhibit a complex suite of bizarre morphological adaptations in the head, mouth and throat. The head looks more reptilian than mammalian; its shape is a key characteristic required to meet the conflicting demands of engulfment and locomotion. A rorqual must have a large, distensible mouth in order to engulf a large volume of water—but it also has to be able to contract and tighten back into the body to maintain a streamlined shape for low drag and efficient steady swimming, particularly during long dives or long-distance migration.

In the larger rorqual species, the skull and mandibles are truly massive, making up nearly 25 percent of the body. The mandibles are connected to the base of the skull through giant pads of a dense, elastic matrix of fibers and cartilage that are infused with oil. This type of jaw joint is unique to rorquals, and possibly also to the closely related gray whale. These specialized jaw joints are flexible

Jeremy A. Goldbogen earned his Ph.D. in zoology in 2009 from the University of British Columbia. He is now a postdoctoral research fellow at the Scripps Institution of Oceanography at the University of California, San Diego. Address: Scripps Institution of Oceanography, University of California at San Diego, Marine Physical Laboratory (Whale Acoustics), 9500 Gilman Dr., La Jolla, CA 92093-0205. Email: jergold@ucsd.edu



Randy Morse, GoldenStateImages.com

Figure 1. Normally streamlined and efficient swimmers, rorqual whales (such as the blue whale, above) become hugely inflated during feeding. These whales fill their expandable oral cavities with tons of seawater and prey, then filter out the water through the baleen that lines their jaws. After diving deep into the ocean, the whales rapidly lunge through dense patches of prey to engulf mass quantities of food. The biomechanics of this process has been obscured by the ocean depths, but electronic tags have elucidated the whales' feeding mechanism.

linkages between the skull and mandibles, which permits the jaws to open to nearly 90 degrees. Such a feature is required so the whales can engulf as much water as possible during a lunge: Although the mouth area is very large, the proportion of that area directed toward the prey is determined by the gape angle between the skull and jaws.

The rorqual skull also possesses a third jaw joint, the *mandibular symphysis*, which connects the mandibles at the center of the lower jaw. In some mammals this linkage is fused, but in rorquals it also has a fibrocartilage composition that enhances its flexibility. With this third, very flexible jaw joint, the strongly curved mandibles are able to rotate outward and increase the area of the mouth. Mandibular rotation is consistently observed in lunge-feeding rorquals at the sea surface, and also in post-mortem specimens when the mus-

cles that hold the mandibles in place release and allow them to sag open.

By having a kinetic skull with specialized jaw joints, rorquals enhance mouth area and increase the rate of water flow into the oral cavity. This rapid influx of water is facilitated by a most unusual mechanism: a tongue that can invert and form a capacious oral sac that accommodates the engulfed seawater on the ventral side of the body. The rorqual tongue is extremely flaccid and deformable. Although it has some distinct structure reminiscent of a typical mammalian tongue, it is weakly muscularized and composed largely of elastic fatty tissue. A floppy, loose tongue can be easily inverted when water rushes into the oral cavity (also called the buccal cavity). Moreover, there is a specialized intramuscular space, called the *cavum ventrale*, located between the bottom

of the tongue and the walls of the buccal cavity, which extends all the way down to the whale's belly button. During engulfment the tongue inverts into the cavum, retreating through the floor of the mouth and back towards the belly button, forming the large oral sac that holds the incoming seawater.

The extreme distension of the buccal cavity during engulfment presents a problem for the walls of the body, which in cetaceans is composed of stiff blubber and firm connective tissue. All rorquals have a distinct series of longitudinal furrows in the ventral blubber that span nearly half of the whale's body length, from the snout to the belly button. In fact, the name "rorqual" comes from the Norwegian word *røyrkval*, meaning "furrow whale." This ventral groove blubber (VGB) consists of tough ridges separated by deep channels of delicate elastic tissue; when viewed in cross-



Figure 2. A historical whaling photo shows a blue whale's oral cavity sagging after death, when the whale's muscles no longer hold it rigid. The whale's floppy tongue and hairlike filtering baleen can also be seen. (Photograph courtesy of the Shetland Museum and Archives.)

section, the VGB has an accordion-like architecture that could easily unfurl if the underlying muscle became relaxed. The tremendous engulfment capacity of rorquals is clearly dependent on this unique morphological design, and it turns out that the VGB is remarkable not only in its structure but also in its mechanical behavior.

Super-stretchy Blubber and Muscle

The first major breakthrough in understanding the biomechanics of lunge feeding came in the late 1980s from some simple, yet elegant, experiments by Lisa Orton and Paul Brodie. The researchers obtained fresh samples of fin whale VGB from a whaling station in Hvalfjörður, Iceland and performed mechanical tests on the tissue to determine how much strain it could withstand for a given amount of stress. They found that the VGB and associated muscle layers could reversibly extend up to several times their resting length. This extraordinary extensibility was attributed to the vast amounts of elastin, a specialized elastic protein, found throughout the tissue, and the fact that the VGB unfolds like a parachute canopy in the absence of muscle tone.

The extensibility of the VGB is a key component of the engulfment apparatus because it provides the great capacity that is needed for a whale to envelop large amounts of water and prey. In addition, the amount of force that is required to sufficiently stretch

the tissue provided an important clue as to how fast a fin whale must swim to successfully execute a lunge. As a rorqual accelerates and lowers its jaws, dynamic pressure is generated inside the oral cavity and applied against the floor of the mouth. In theory, the dynamic pressure alone could generate enough force to completely extend the VGB and inflate the buccal cavity, but only if the swimming speed is high enough. By approximating the inflated buccal cavity as a thin-walled cylinder, Orton and Brodie predicted that a lunge speed of 3 meters per second would be sufficient to maximally fill a fin whale's buccal cavity. This prediction seemed consistent with sea-surface observations, but there was really no way to accurately measure swim speed during a lunge until very recently.

Lunges in the Deep

Nearly two decades after Orton and Brodie's study, the opportunity came to test their predictions by examining the motion, or kinematics, of rorqual

lunge-feeding in the natural environment. Bill Burgess of Greeneridge Sciences developed a high-resolution digital tag that could be temporarily attached to the backs of whales as they surface to breathe. The tags, equipped with suction cups for attachment and a flotation device for retrieval, contained a variety of sensors, including a hydrophone, a pressure transducer and an accelerometer. The data from these tags provided a short glimpse into the underwater behavior of rorquals, including body orientation, the times when the whale was swimming versus when it was gliding, and dive depth. The application of these tags, using a long fiberglass pole, is not a trivial task—it requires many years of experience at sea and typically a coordinated effort between a large support vessel and a smaller tagging boat.

The tagging operations were led by John Calambokidis and Greg Schorr from the Cascadia Research Collective in Olympia, Washington, and Erin Oleson of the Scripps Institution of Oceanography. For the past seven to eight years, tagging studies were conducted every summer at various locations off the coast of California and Mexico. The tag data showed that many rorquals

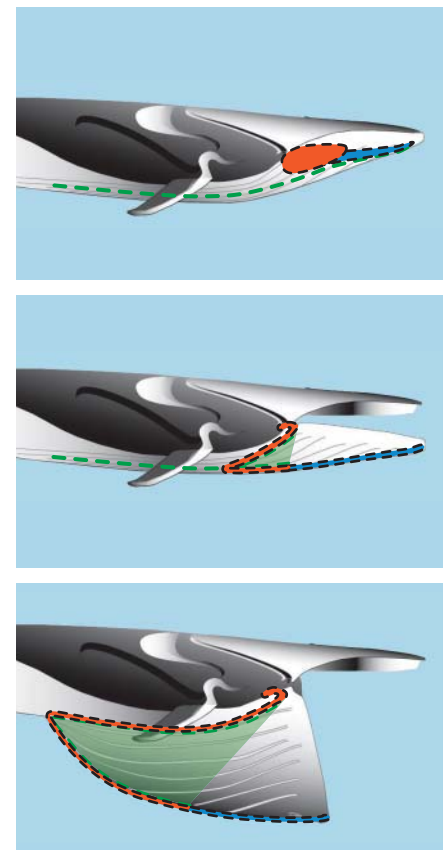


Figure 3. When a whale is not feeding, its tongue (red, top) is furling up along the floor of its mouth (blue) and the cavum ventrale (green dotted line) is collapsed. As the whale engulfs water and prey, the deformable tongue pushes through the cavum ventrale and the oral sac starts to expand (green, middle). At full expansion, the tongue is inverted and flattens out completely to form a large part of the wall of the oral sac; the floor of the mouth also stretches to form part of the cavity (bottom).

made consecutive deep dives, in some cases up to 300 meters in depth. At the bottom of these deep dives, the data showed a series of wiggles or undulations. Each wiggle was accompanied by an intense bout of active swimming strokes and a concomitant decrease in water flow noise, which indicated a rapid decrease in speed that is typically seen during lunges at the sea surface.

Although these data were suggestive of lunges at depth, direct evidence came from another type of suction-cup tag: National Geographic's Crittercam, conceived and developed by Greg Marshall. The video camera within Crittercam was equipped with an infrared light, for the dark conditions during deep dives, and also a time-depth recorder. The video footage shows the whale swimming through dense fields of krill at the bottom of deep dives. The images from one Crittercam deployment shows the whale's lower jaws dropping, followed by a decrease in flow noise, and then an expansion of the ventral groove blubber. This provided visual confirmation of the behavior that we had interpreted from the data recorded by our digital tags: several consecutive lunges at the bottom of deep foraging dives.

After more analyses, we realized that we could use the level of flow noise recorded by the digital tag's hydrophone to calculate the whale's swimming speed throughout each foraging dive. This "flow-noise speedometer" revealed just how rapid the changes in speed were during a lunge. Amazingly, the maximum lunge speed recorded for fin whales was 3 meters per second, precisely the flow speed that Orton and Brodie had predicted to be enough to passively inflate the buccal cavity. Furthermore, the speed data revealed a rapid deceleration of the body even while the whale continued to swim actively, an indication that the whale was experiencing very high drag as it opened its mouth wide. The kinematic data from the tags, it turns out, held the key to determining not only how much drag is incurred, but also how much water the whale engulfs.

Big Gulps and High Drag

As a rorqual lowers its jaws and presents the inside of its mouth to oncoming flow, water that is rushing into the mouth will expand and distend the throat pouch. Such a reconfiguration represents a major departure from the whale's normal

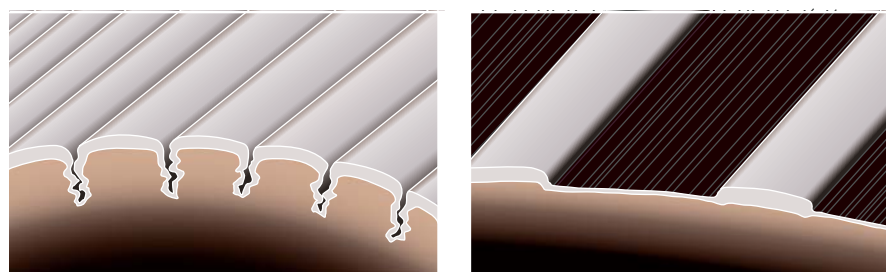


Figure 4. A rorqual whale's ventral groove blubber (VGB) allows its oral cavity to expand enormously. The firm ridges of the accordionlike VGB are connected by deep furrows of delicate elastic tissue (*top*). In cross-section the VGB is made up of layers of muscle, elastin, fatty blubber and epidermis (*bottom left*). Mechanical tests have shown that VGB can expand to more than twice its original length (*bottom right*). (Top photograph courtesy of Nick Pyenson.)

sleek, well-streamlined body profile. The result is predictably high drag as flow is directed around the mandibles and distended buccal cavity, which robs momentum from the whale and causes the body to decelerate rapidly. The size and shape of the mandibles, therefore, has a great influence on how much drag is experienced during a lunge. Because the mandibles determine the size of the mouth, they also largely determine how much water is engulfed.

Recognizing the effects of skull and mandible shape on the mechanics of engulfment, Nick Pyenson of the Smithsonian Institution, Bob Shadwick of the University of British Columbia and I set out to measure as many museum specimens as possible. By integrating our morphological measurements with the kinematic data obtained from the tags, we were able to estimate how much water is engulfed during a fin whale lunge. When the jaws were open to maximum gape, for example, our calculations suggested that the buccal cavity was filling at a

rate of approximately 20 cubic meters per second. At the end of a lunge that lasted six seconds, the accumulated engulfed water mass was about 60 tonnes, which again supported the prediction made by Paul Brodie in 1993. The engulfed water mass is about the size of a school bus, a truly enormous amount. The magnitude of the engulfed water is also colossal relative to the whale, whose own body mass weighs in at around 45 tonnes.

By engulfing a volume of water that is greater than its own body mass, the whale necessarily incurs tremendous amounts of drag. The whale must do work against this drag, and this represents a major source of energy expenditure during a lunge. The work required for engulfment triples over the course of a lunge, whereas the drag increases approximately five-fold. One other important metric in the field of hydrodynamics, called the drag coefficient, is a measure of how streamlined an object is and how effective it is at decreasing drag. High drag coefficients are typical



Figure 5. Attaching an electronic tag to a moving blue whale is no easy task. Here, a Cascadia Research team places a tag via a long fiberglass pole, under a research permit from the National Marine Fisheries Service. The red flotation device on the tag facilitates its recovery once the tag's suction cups fall off the whale. (Photograph by Sherwin Cotler, Cascadia Research.)

of poorly streamlined shapes, whereas low values indicate a highly streamlined shape. Our simple calculations suggested that, over the course of a lunge, the drag coefficient increases by more than an order of magnitude. Thus, a lunge-feeding rorqual undergoes an extreme transformation from a very well-streamlined body to one that is highly susceptible to drag. Interestingly, the maximum drag coefficients were very similar to those for parachutes, another inflating system. The analogy between inflating parachutes and lunge feeding is logical: Both systems must reconfigure in order to generate drag. In other words,

parachutes need drag to inflate and slow down their cargo, whereas rorquals require drag to inflate the buccal cavity.

Of Whales and Parachutes

The realization that rorqual lunge feeding involves incredibly high amounts of drag led us to a most unlikely collaboration with Jean Potvin, a parachute physicist at Saint Louis University in Missouri. Together we developed a new, more detailed model of rorqual lunge feeding inspired by decades of parachute-inflation studies. For a given morphology and initial lunge speed, the model predicted what decrease in velocity to expect for

a passively engulfing whale as it experiences drag. By comparing the model output to the empirical tag data, we could explicitly test particular engulfment mechanisms. The first question we asked was: Is a lunge-feeding whale just like an inflating parachute? If this were the case, a rorqual would inflate *passively*, and the flow-induced pressure that expands the buccal cavity would be met with little resistance because of the extremely compliant properties of the VGB.

Our simulation of passive engulfment in fin whales resulted in a poor match with the tag data because the body was simply not slowing down rapidly enough. In other words, there wasn't enough drag to account for the rapid deceleration that we had observed in the tagged whales. This also meant that water was going into the mouth far too rapidly and, as a consequence, the buccal cavity reached maximum capacity halfway through the lunge (at about the point when the mouth was open to maximum gape). Maximum filling of the buccal cavity would occur because at some point the VGB cannot extend anymore. At that point in time, the entire engulfed mass would have to be immediately accelerated up to the instantaneous speed of the whale (2 meters per second), which would impose unrealistic forces on the walls of the buccal cavity. If the VGB was not strong enough to accommodate these excessively large forces, passive engulfment would cause catastrophic blow-out of the buccal cavity. If the VGB was strong enough to withstand these forces, the engulfed mass would rebound off the buccal-cavity wall and eject back out of the mouth before the jaws closed. In either scenario passive engulfment does not seem to be a feasible mechanism for fin whales. However, such a mechanism might still be possible for lunges involving lower gape angles and smaller engulfed volumes.

If passive engulfment is not possible, how do rorquals execute a lunge successfully? There are two key anatomical characteristics of the VGB that suggested a very different engulfment mechanism. First, we realized that there were several layers of well-developed muscle that adjoin tightly to the grooved blubber. Second, a study by Merijn de Bakker and his colleagues at Leiden University revealed that there were specialized nerves sensitive to mechanical stress, called mechanoreceptors, embedded within both the

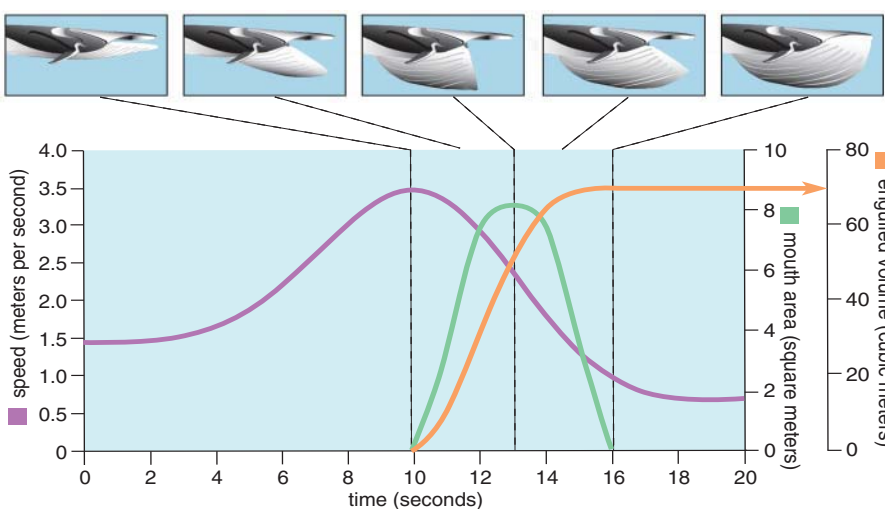


Figure 6. Data from electronic tags on rorqual whales has allowed researchers to break down the mechanics of lunge feeding. After diving to a depth of several hundred meters and accelerating (purple line) into a school of krill, a whale opens its mouth (green line), causing massive drag and deceleration. The oral cavity fills with water (orange line) and the whale closes its mouth, then begins filtering out the water and preparing for the next lunge.

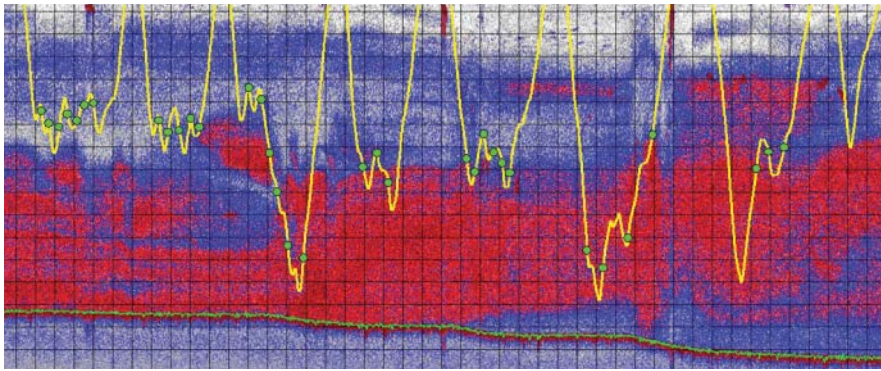


Figure 7. Data from an electronic tag on a feeding humpback whale show that the whale dove to 50 to 140 meters (yellow lines) to feed on patches of krill (medium density in blue, high density in red). The whale performed a series of lunges (green dots) before returning to the surface to breathe. The green line is the seafloor, and each black gridline is 10 meters in depth, with the top starting at five meters below the surface. Each dive lasted several minutes. (Image courtesy of the author.)

muscle and blubber layers of the VGB. These receptors were concentrated within each groove, which is precisely the region of the tissue that would stretch during engulfment. These two lines of evidence suggested that rorquals may be able to gauge the magnitude of the engulfed water mass from the amount of stretch sensed by the tissue and then generate enough force to slowly push the water forward. Such a mechanism is possible if the VGB muscles actively resist lengthening as they are stretched by the incoming flow. By virtue of Newton's third law of motion, demanding equal action and reaction, the whale imparts its momentum to the engulfed water during this "collision"; the whale slows down as the engulfed water, which was initially at rest, speeds up, and eventually both of their speeds become more similar.

When we simulated this type of active engulfment, we found a good match to the velocity profile generated by the digital-tag data. The model output supported our hypothesis of active inflation in rorquals, a very different mechanism than what is observed in parachutes. But why would rorquals push water forward, out of the mouth, when they are trying to engulf it? Indeed, this shove from inside the buccal cavity generates even more drag compared to the case where water is just going around the body and the mouth, which is why the active engulfment simulation better matched the tag data.

Although it seems counterintuitive, pushing water forward during a lunge has some advantages. Gradually pushing water forward over the course of a lunge distributes the drag forces over a longer period. By smoothing out these forces

over lengthier time scales, the peak drag forces experienced by the engulfment apparatus are effectively lower.

Another benefit associated with active engulfment is that it may increase the energetic and mechanical efficiency of filtration by the baleen. If the water is slowly pushed forward, the entirety of the engulfed water mass no longer has to be accelerated from rest. Moreover, because the trajectory of the engulfed water mass inside the buccal cavity is largely parallel to the filter surface of the baleen, rorquals could employ cross-flow filtration; this highly efficient filtration mechanism washes material perpendicularly across the filter surface to prevent clogging. It is used on an industrial scale (for example, in water purification, beer and wine production, and biotechnology processes) and has also been observed in suspension-filter-feeding fish.

If such a mechanism exists in rorquals, as opposed to dead-end filtration where the filtrate gets stuck in the filter itself, it could be very effective at keeping small zooplankton from embedding in the baleen fringes. And further, if krill were to cake the baleen, how would a rorqual scrape it off with such a floppy, weakly muscularized tongue? Maybe one day technology will enable us to visualize the flow inside the mouth of a filtering rorqual and resolve the debate.

Paying the Price to Lunge

The high drag that is required for engulfment has major consequences for rorqual foraging ecology and evolutionary morphology. Not only must rorquals expend significant amounts of energy to accelerate the engulfed water mass, but the high drag also robs the whale of its kinetic energy, bringing the body to a near halt. As a consequence, the body must be reaccelerated from rest in order to execute the subsequent lunge. While holding its breath at the bottom of a dive, the whale must lunge over and over again, and this represents a high energetic cost. Thus, rorquals rapidly deplete their oxygen stores when foraging at depth and must quickly return to the surface to recover. The feeding costs related to high drag during lunge feeding effectively limit the amount of time a large rorqual can spend foraging at depth to about 15 minutes or so per dive. This short timeframe is unexpected because rorquals are so large, and in nearly all other air-breathing vertebrates, diving time usually scales up with increased size, due to a more efficient metabolism.



Figure 8. Specimens at the Smithsonian Institution's National Museum of Natural History include the mandibles of a blue whale (gray) and a sperm whale (yellow). The author (shown twice in this composite photograph) used measurements of such museum specimens to estimate the amount of water engulfed during a whale's lunge. (Photograph courtesy of Nick Pyenson.)

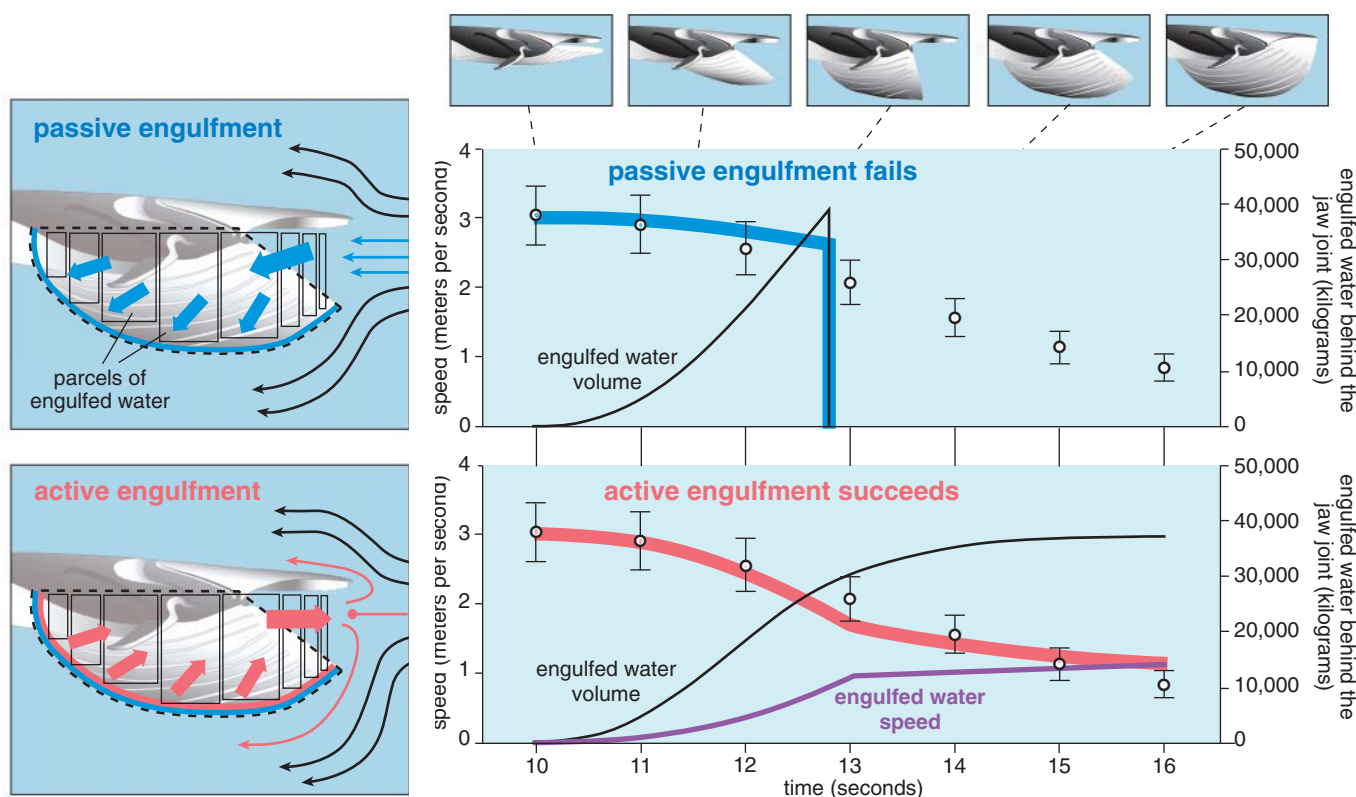


Figure 9. Is a whale like a parachute, expanding its oral cavity passively (*top left*), or does it actively control the process, using its muscles to slowly push the water forward during a lunge (*bottom left*)? Trajectory simulations, which predict the speed of the whale over time, for passive engulfment (*blue line, top graph*) fail to reproduce the data recorded by electronic tags (*white dots with error bars*). However, trajectory simulations of active engulfment (*red line, bottom graph*) produce a close match.

The severely limited diving performance of rorquals was first documented nearly a decade ago by Donald Croll and colleagues at the University of California, Santa Cruz. By attaching simple time-depth recorders to the backs of surfacing blue and fin whales, the researchers discovered that the whales' foraging dives were much shorter than expected for their size. Furthermore, foraging dives that involved more lunges at depth resulted in more surface recovery time after each dive. Using these data, Croll's research group was the first to hypothesize that, due to drag, there was a high energetic cost for each lunge. This hypothesis has been supported by several studies since then, not only for fin and blue whales, but for humpback whales as well.

Because maximum dive time is limited by these high foraging costs, rorquals are particularly dependent on dense aggregations of prey. In addition, it is predicted that a rorqual is morphologically designed to engulf as much water as possible per lunge, which may be why the buccal cavity extends halfway down the body to the belly button and why the jaws make up nearly a quarter of the body length. But why isn't the engulfment apparatus even larger? What are the limits to engulfment capacity and how does it change with body size? These questions led me to a long-forgotten morphometric data set from

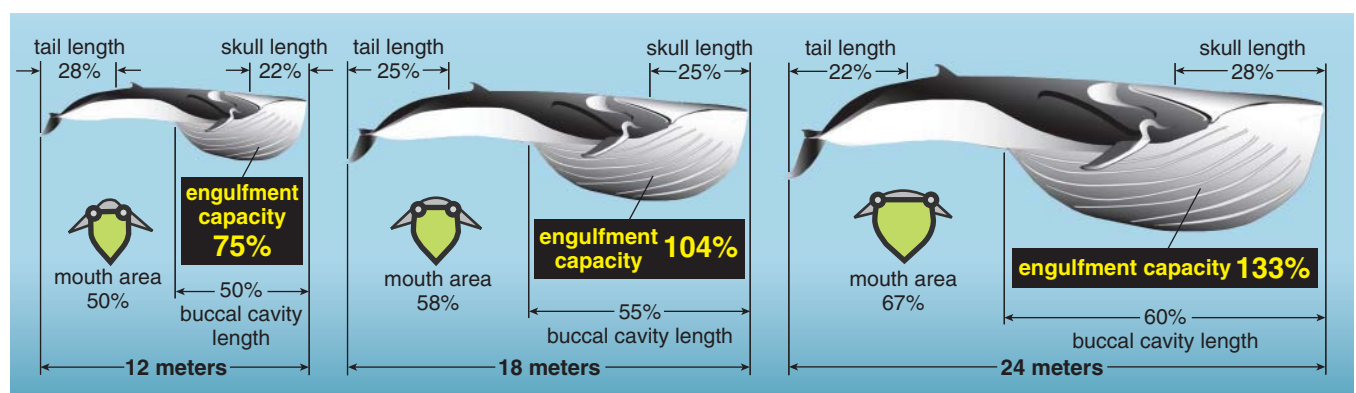


Figure 10. As a fin whale grows, its oral (or *buccal*) cavity does not scale linearly but takes up a larger percentage of its body size. The buccal cavity length increases from 50 percent to 60 percent, and skull length increases from 22 percent to 28 percent, of body length, whereas tail length decreases from 28 percent to 22 percent of body length. The mouth area increases from 50 percent to 67 percent of total projected body area, and engulfment capacity rises from 75 percent to 133 percent of body mass.

the whaling literature, which allowed me to examine the consequences of scale and morphology on lunge-feeding performance.

A Matter of Scale

In an attempt to manage the whaling industry in the 1920s, the British government launched a series of expeditions called the Discovery Investigations in order to learn more about the natural history and biology of large whales in the Southern Ocean. One particular study focused on the body proportions of the two largest rorqual species, fin and blue whales. These species are not only some of the largest animals of all time, but what is often underappreciated is that they also exhibit a wide range in body size. For example, the length at weaning for fin and blue whales is approximately 12 meters and 16 meters, respectively, whereas the maximum size recorded for each species is 24 meters and 28 meters. These expeditions recorded morphometric data, some of which was related to the engulfment apparatus, for hundreds of fin and blue whales over this entire body-size range.

The authors of this study discovered a peculiar pattern related to body size: Larger whales had larger jaws and buccal cavities relative to body size. At the same time, the size of the posterior part of the body (the region from the dorsal fin back towards the tail fluke, or *caudal peduncle*) became relatively smaller. The researchers gave no possible explanations for these bizarre patterns of relative growth (also called *allometry*), probably because the data were collected before we knew how important morphology is in determining lunge-feeding performance. My colleagues and I amassed their complete data set for fin whales in order to estimate engulfment capacity as a function of body size. As we expected, the relative size of the engulfed water mass increased with body size, and this was directly due to the allometry of the engulfment apparatus.

But why did larger whales have relatively smaller caudal peduncles? We hypothesized that this relative shrinking of the tail could represent the cost of devoting all growth-related resources to the anterior region of the body. As rorquals grow, they become morphologically optimized to increase engulfment capacity. The skull becomes relatively longer and wider with body size, and therefore the area of the mouth that is devoted to engulfment is also relatively

greater. In addition, the length of the ventral groove blubber system is also relatively longer in bigger whales, and this effectively increases the relative capacity of the buccal cavity. Given that many other large rorquals also exhibit the same patterns of relative growth, these allometric patterns may represent an adaptation (or exaptation) related to lunge-feeding performance.

The relatively smaller tail should not negatively affect swimming performance in larger rorquals because the actual fluke—the propulsion surface that generates the lift used for thrust—is generally proportional to body size. However, the enhanced engulfment capacity in larger whales does not come without a cost. The active nature of engulfment means that relatively larger water masses must be accelerated forward. Thus, larger rorquals will have to expend relatively more energy to successfully execute a lunge. Considering that high feeding costs limit dive time relative to other diving animals, such rapidly increasing costs for a lunge may limit diving capacity in larger rorquals even more. Such a consequence could be detrimental because sufficiently dense prey patches tend to be very deep. Theoretically, the rate of energy expenditure to feed will increase more rapidly with body size than the rate of energy gained from lunge feeding.

If this scenario is extrapolated to a hypothetical megarorqual that is much larger than a blue whale, we find that the whale would not be able to support its metabolism by lunge feeding. Similar problems associated with large body size were predicted by R. McNeil Alexander for baleen whales that were geometrically similar to one another (all body lengths being proportional to body size). Although rorqual allometry enhances engulfment capacity for a single lunge, the cost associated with it could limit access to food in the deep ocean. From this line of reasoning, we have speculated that the allometric scaling of lunge-feeding energetics has imposed an upper limit on body size in rorquals. It is interesting to think about why an animal isn't, or wasn't, larger than a blue whale, and clearly more studies are needed to explore this hypothesis and others related to limits on big body size. Evolution may have driven the size of these largest of marine mammals to their current scale, but physiological constraints related to filter feeding may also have imposed an upper bound past which they can grow no farther.

References

- Alexander, R. M. 1998. All-time giants: The largest animals and their problems. *Palaeontology* 41:1231–1245.
- Acevedo-Gutiérrez, A., D. A. Croll and B. R. Tershy. 2002. High feeding costs limit dive time in the largest whales. *Journal of Experimental Biology* 205:1747–1753.
- Brodie, P. F. 1993. Noise generated by the jaw actions of feeding fin whales. *Canadian Journal of Zoology* 71:2546–2550.
- Calambokidis, J., et al. 2007. Insights into the underwater diving, feeding and calling behavior of blue whales from a suction-cup-attached video-imaging tag (CRITTERCAM). *Marine Technology Society Journal* 41:19–29.
- Croll, D. A., A. Acevedo-Gutiérrez, B. Tershy and J. Urbán-Ramírez. 2001. The diving behavior of blue and fin whales: Is dive duration shorter than expected based on oxygen stores? *Comparative Biochemistry and Physiology Part A: Molecular and Integrative Physiology* 129A:797–809.
- de Bakker, M. A. G., R. A. Kastelein and J. L. Dubbeldam. 1997. Histology of the grooved ventral pouch of the minke whale, *Balaenoptera acutorostrata*, with special reference to the occurrence of lamellated corpuscles. *Canadian Journal of Zoology* 75:563–567.
- Goldbogen, J. A., et al. 2008. Foraging behavior of humpback whales: Kinematic and respiratory patterns suggest a high cost for a lunge. *Journal of Experimental Biology* 211:3712–3719.
- Goldbogen, J. A., et al. 2006. Kinematics of foraging dives and lunge feeding in fin whales. *Journal of Experimental Biology* 209:1231–1244.
- Goldbogen, J. A., J. Potvin and R. E. Shadwick. 2009. Skull and buccal cavity allometry increase mass-specific engulfment capacity in fin whales. *Proceedings of the Royal Society B*, published online November 25.
- Goldbogen, J. A., N. D. Pyenson and R. E. Shadwick. 2007. Big gulps require high drag for fin whale lunge feeding. *Marine Ecology Progress Series* 349:289–301.
- Lambertsen, R. H. 1983. Internal mechanism of rorqual feeding. *Journal of Mammalogy* 64:76–88.
- Oleson, E. M., et al. 2007. Behavioral context of call production by eastern North Pacific blue whales. *Marine Ecological Progress Series* 330:269–284.
- Orton, L. S., and P. F. Brodie. 1987. Engulfing mechanics of fin whales. *Canadian Journal of Zoology* 65:2898–2907.
- Pivovrunas, A. 1977. Fibro-cartilage skeleton and related structures of ventral pouch of balaenopterid whales. *Journal of Morphology* 151:299–313.
- Potvin, J., J. A. Goldbogen and R. E. Shadwick. 2009. Passive versus active engulfment: verdict from trajectory simulations of lunge-feeding fin whales *Balaenoptera physalus*. *Journal of the Royal Society Interface* 6, 1005–1025.

For relevant Web links, consult this issue of *American Scientist Online*:

<http://www.americanscientist.org/issues/id.83/past.aspx>

The Race for Real-time Photorealism

The coevolution of algorithms and hardware is bringing us closer to interactive computer graphics indistinguishable from reality

Henrik Wann Jensen and Tomas Akenine-Möller

Ever since the emergence of three-dimensional computer graphics in the early 1960s, graphics specialists have dreamed of creating photorealistic virtual worlds indistinguishable from the real world. Product designers, architects, lighting planners, gamers and scientific visualization pioneers have craved real-time reality on a chip; hardware designers and algorithm writers have made spectacular progress, as one can see by looking over the shoulder of teenagers playing the latest games (*facing page*). But as we will see, the computational challenges that remain are immense. Implacable evolutionary progress has been made by software engineers in devising ingenious algorithms, and generations of hardware have been invented to traffic and execute the calculations, yet the sheer scale of the computational task keeps the goal of real-time photorealism at some distance over the horizon. Most office computers consume a small sliver above zero percent of their available computational cycles for routine work; the billions of calculations per second that are available on a modern multi-processor desktop computer are simply not required to process spreadsheets. Compare that to the overwhelming task of the most

advanced 3D applications, churning through the calculations required to produce a scene using the latest algorithms, including the tracking of billions of simulated photons through a scene, and even the tracking of the simulated light penetrating the scene's surfaces to achieve the perfectly convincing photorealistic image.

Such renderings can take today's fast machines hours to produce a single frame.

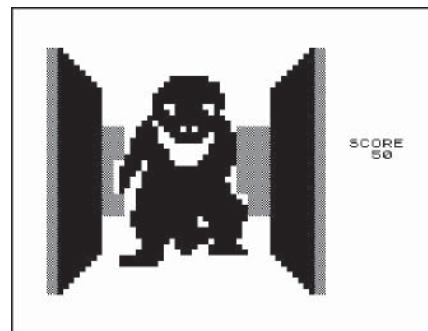
Elite gamers clamor for no less than 60 frames per second. Why so many? Because the pursuit of real-time graphics is driven by the desire for not just visual accuracy but also interactivity. Real-time scenes are created to be interacted with. The television standard of 29.97 frames per second is comfortably convincing for passive viewers; real-time applications, such as gaming and military cockpit simulations, must operate at the speed of human reflexes.

As participants in the enterprise of creating photorealistic graphics (one of us having a research emphasis on greater speed, the other on greater realism), we'll review the kinds of computations required, the schemes that have been invented to moderate the heavy computational chores, and the parallel world of hardware development to support the calculations. The hardware and software of photorealistic graphics have coevolved for several decades. The economics of hardware development, driven mainly by gamers' unquenchable lust for speed, has resulted in affordable graphics cards of awesome power. Computations have been moved from the central processing unit (CPU) of computers to the specialized graphics processing units (GPU) of consumer video cards. The leap in computational prowess then drives the develop-

ment of greedier algorithms for more convincing realism. This cycle has gone on for decades, blossoming into multi-billion-dollar video card and game software industries.

Making the Scene

Current real-time graphics applications, such as games, represent the complexity of virtual environments by converting scene descriptions into millions of geometric primitives—points, lines, and polygons, usually triangles, connected to form polygonal surfaces. Early games represented 3D scenes using a few hundred triangles; in historical context, the experience of interacting with these 3D environments could be quite compelling, but the appeal had little to do with realism.



3D graphics circa 1981: *3D Monster Maze* on the Sinclair ZX81 with 16-kilobyte memory expansion.

More triangles made for more convincing scenes. An obvious first step in accelerating the rendering of a scene was to optimize the number of triangles required. Scene designers are obliged to make judicious decisions about the balance between polygon count and realism. How much detail is enough? Current GPU hardware can process scenes composed of several million triangles while still reaching

Henrik Wann Jensen is an associate professor at the University of California, San Diego, where he specializes in realistic image synthesis and the rendering of natural phenomena. He received his Ph.D. in computer science at the Technical University of Denmark. Tomas Akenine-Möller is a professor of computer science at Lund University who works part-time with Intel, specializing in computer graphics and image processing. He received his Ph.D. in computer graphics at Chalmers University of Technology. Address for Jensen: Computer Science and Engineering, 4116, University of California, San Diego, CA 92093-0404. Email: henrik@graphics.ucsd.edu



Figure 1. Consumer demand combined with algorithmic artistry and muscled-up hardware have driven computer graphics far toward the long-imagined goal of photorealistic animation. The state-of-the-art animated feature movie *Ratatouille*, released by Pixar Animation Studios in 2007, was produced by an arsenal of about 850 computers hosting nearly 3,200 processors. The average rendering time for each frame of animation was about 23,000 seconds per frame. Today's video gamers want the same visual quality—at 60 frames per second. And they are on the road to getting it, as can be seen in the Electronic Arts 2008 action and adventure game *Mirror's Edge*, which delivers dazzling interactive play at more than 60 frames per second on personal computers. (The image above from that game was rendered offline with additional resolution to achieve print quality.) The authors review the roadmap to a future in which advances in speed and photorealism finally achieve the goal of perfectly convincing interactive computer graphics in real time. (Image courtesy of EA Digital Illusions Creative Entertainment.)

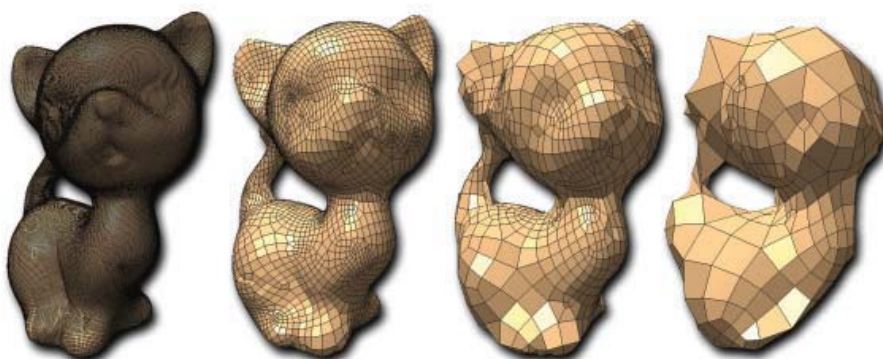


Figure 2. Optimization and approximation are keys to graphics rendering speed. Algorithms optimize how many calculations must be made, and scene elements such as shadows, reflections and even geometry may be approximated, with accuracy surrendered for speed. Above, the polygon count of a 3D model is progressively reduced. For a rendering scheme such as rasterization, which renders individual polygons, the middle models would render much faster than the one on the left. As the distance from the viewer to the cat increases, fewer and fewer polygons can be used with little loss in quality. (Adapted from Daniels, J., C. T. Silva, J. Shepherd and E. Cohen. 2008. "Quadrilateral mesh simplification." *Proceedings of SIGGRAPH Asia* 27(5):1-9.)

the gamer's benchmark of 60 frames per second. If current trends continue, we can expect hardware in the near future that handles hundreds of millions or even billions of triangles with sufficient speed. The question is: how many triangles are required to achieve a photorealistic rendering of a given scene? One of the founders of Pixar Animation Studios, maker of 3D blockbusters from *Toy Story* to *Ratatouille*, concluded that 80 million triangles would be required. It seems that the hardware will soon be up to the job of handling the geometry in real time. However, there is much more to photorealism than polygon count.

Faster with Rasterization

With contemporary hardware and software, the fastest way to *render* a scene (convert the 3D data to a visual image) is *rasterization*, the technique used by today's computer games. An algorithm processes the scene detecting what geometry is visible and what is screened from view (including the back faces of 3D objects facing the viewer). Nonvisible geometry is discarded to speed the calculation, and then the scanner determines which vertices are closest to the viewer. Triangles formed by vertices are painted onto a virtual screen, as shown in Figure 3. The color of each pixel on the screen is determined by the color and surface properties assigned to the triangle, as well as the lighting in the scene. The angles where triangles abut are made to vanish in the image by the neat trick of averaging the color values of adjacent triangles. Color and surface properties (ruggedness, sheen, and so

on) are assigned by software instructions called *shaders*. Most commonly, the surface information is assigned using *texture maps*, which are digital images "glued" onto the 3D object. Texture mapping is an art in itself. In the production pipeline of 3D studios, artists specialize in the creation of texture maps to convey, for example, not just the color of an orange, but also the knobbly surface and the waxy shine. An early breakthrough on the road to photorealism was bump mapping, conceived by the computer graphics pioneer James F. Blinn. (It was said of him quite a few years ago, by the graphics hardware innovator Ivan Sutherland, that "there have been about a dozen great computer graphics people and Jim Blinn is six of them." Blinn has

made many milestone contributions in the field of deriving convincing images from 3D data.) Bump maps convey details of microfine surface structure without adding to the overall geometry load by telling the renderer to handle local lighting *as if* the surface were bumpy, with the bumps defined by light and dark areas on the texture map.

Let There Be Lighting. And Shadows.

The critical element of lighting in a 3D environment comes from virtual light sources placed in the scene. In rasterization schemes, a few simple equations are used to compute how much light emanating from a light source arrives at a given point on each triangle, and how much of this light is reflected towards the observer.

The earliest 3D renderings had a signature, otherworldly look because they lacked shadows, a critical aspect of visual realism. Rendering shadows with rasterization is straightforward using a technique that employs multiple rendering passes. For example, one can use a shadow-mapping algorithm, where the scene is rendered from the light source into a shadow map in a first pass. The shadow map contains information about all the triangles visible from the point of view of a particular light. In a second pass, from the "camera" point of view, which is different from the light source, the color calculation for each triangle queries the shadow map to see if the triangle is visible from the light source or is in shadow. Adjustments are then made to the color of the triangle to account for the shadow.

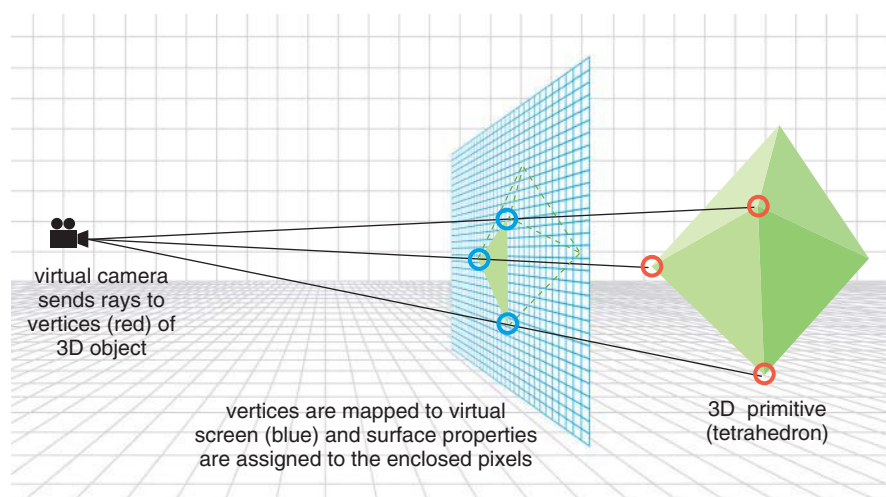


Figure 3. Rasterization algorithms render scenes by projecting rays to the vertices of geometry in the scene, thus defining polygons that are then mapped on pixels. Surface properties assigned to the polygons of the model are then computed and mapped to the pixel screen to create an image.

A decided weakness of rasterization is the rendering of reflections and refractions. Refractions in the real world can be seen as the bending of light when it passes through a transparent medium such as a glass of water. Like shadows, they contribute greatly to realism. For a variety of reasons, reflections and refractions cannot be computed using the triangle-painting technique that is the core strategy of rasterization. Workarounds have been devised to create illusions of reflection and refraction, but the basic problem these lighting effects present has proved intractable for rasterization schemes.

There are other lighting effects that rasterization fails to capture. In real scenes, color bleeding occurs when diffuse surfaces are illuminated by indirect lighting. For example, in a white room with a red carpet, the carpet casts a subtle red glow onto the white walls. Another elusive phenomenon is caustics; when real light is refracted or reflected through a transparent medium, focusing effects can produce blooms of intense brightness. An example of caustics is the shimmering waves of brightness seen on the bottom of a swimming pool. Subsurface scattering is a particularly notable recent development on the road to photorealism that is confounded by the limitations of rasterization. Real materials often have a degree of translucency on their surface. Think of how light penetrates jade. As light crosses

the material's surface, it is scattered, some inward, some back out. The distinctive visual quality of subsurface scattering accounts for the appearance of, among many other things, human skin, and the difficulty of accurately reproducing it accounts for the notoriously unconvincing appearance of many 3D renderings of faces. At present, rasterization is the main player in real-time graphics, but in the opinion of many, for reasons that include its limitations at handling advanced lighting effects like those just mentioned, it will not be the road to real-time photorealism.

Racier Hardware

Hardware is part of the answer. Better graphics is the main reason why average consumers want faster computers, and one of the key technologies driving real-time graphics is the use of specialized graphics processing units that can process and display vast amounts of geometry rapidly. GPUs achieve their performance by using a high degree of parallel processing, in which the task of rendering a scene is divided

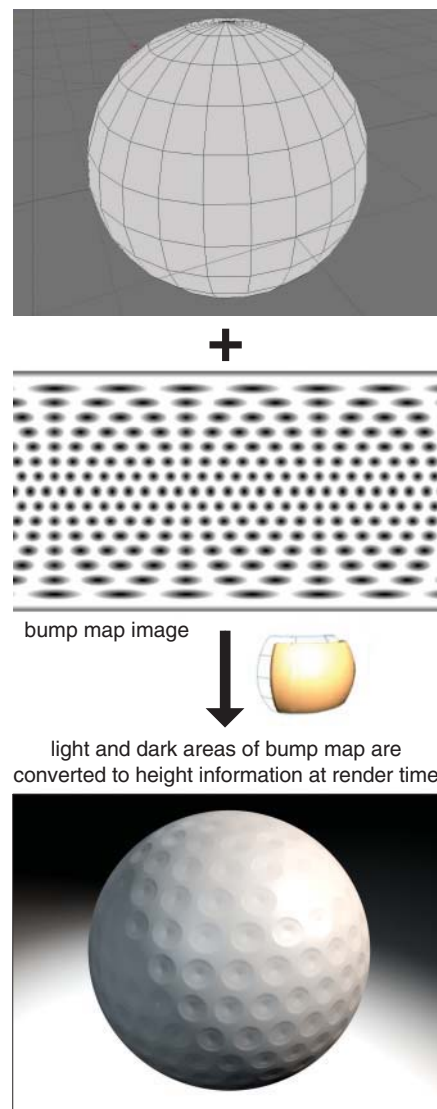


Figure 4. Bump mapping is an extremely efficient scheme for conjuring fine surface detail at render time. A texture map assigned to a 3D model gives shading instructions to the renderer in the form of light and dark regions, which indicate whether regions should cast shadows as if they were slightly raised or lowered from the actual surface of the geometry.

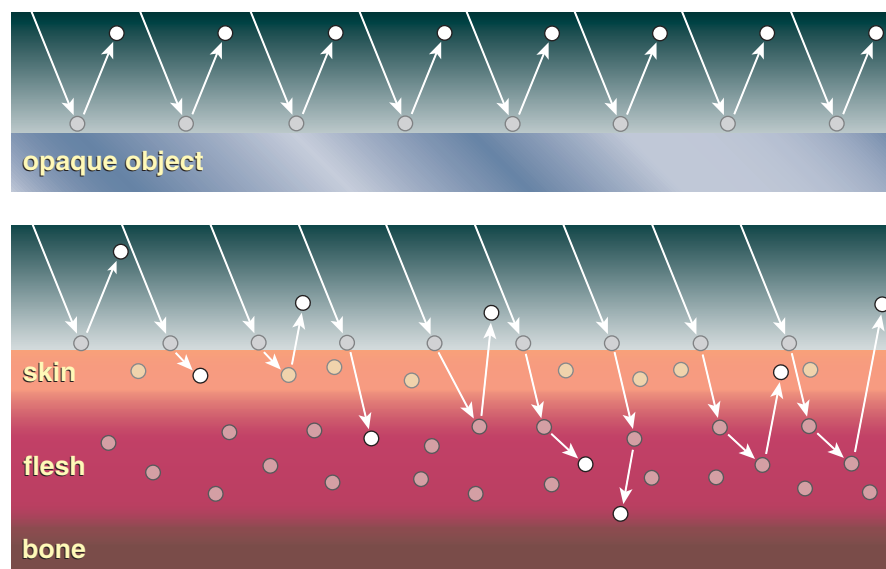


Figure 5. Visual subtleties can be costly in terms of calculation yet necessary for realism. Light hitting a surface such as skin penetrates and scatters, illuminating the surface from within. Subsurface scattering is an algorithm that captures that effect by propagating light rays and tracking their effects, based on material properties assigned to the 3D object. Renaissance painters, such as Vermeer in his *Portrait of a Young Woman*, met the realism challenge with the analogous technique of glazing, applying layers of translucent pigments to capture and scatter light.



Metropolitan Museum of Art, New York/The Bridgeman Art Library Inc.

into smaller tasks that can be processed in parallel by specialized computing units within the GPU. Graphics can be seen as a black hole of computational power—the more power you throw at the problem, the more consumers and developers demand in order to render ever more complex images. Hardware architectures, both CPUs and GPUs, are being designed with these market forces in mind.

A recent development in GPU technology is programmability. Ten years ago, GPUs were essentially fixed-function units with some tweakable parameters to accommodate a few different types of calculations. The rigidity of GPUs greatly limited the types of graphic effects that could be rendered. With the newfound flexibility of programmable GPUs, a programmer can specify advanced lighting models chosen to maximize the potential of the hardware. Programmable hardware has also opened the door to conjuring tricks that overcome the inherent limitations of the rasterization approach. For example, researchers and game developers around the world, in pursuit of ever more realistic game scenarios, have developed approximative multi-pass algorithms that can imitate color bleeding, caustics, and subsurface scattering using rasterization on GPUs. However, this development is starting to hit a wall. The results may be attractive, even entrancing, but by the standards of photorealism they are not convincing. Achieving *true* photorealism will require a fundamental change in the way real-time graphics deals with geometry and lighting.

Realism with Ray Tracing

Whoever solves the riddle of moving beyond rasterization will likely hold the key to the future of real-time graphics. A race is on to develop new hardware capable of supporting new algorithms that can simulate the lighting effects that rasterization cannot handle. One of these algorithms is ray tracing.

Conceptually, ray tracing and rasterization are not that different: Both solve for visibility along a ray. Ray tracing differs in simulating individual light rays that shoot through a 3D environment, including the simulated propagation of new rays when light bounces off scene geometry—multiple new rays, in fact, if the light bounces off diffusely, reflectively, refractively, or in combination as real light gener-

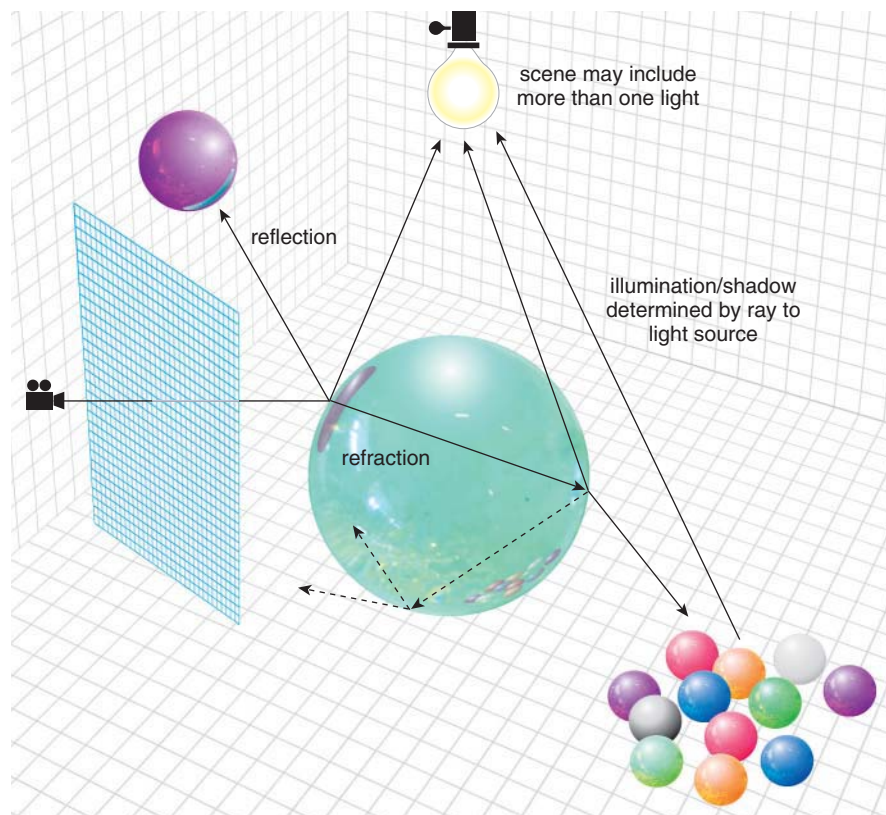


Figure 6. In ray tracing, a ray is shot through each pixel of a virtual screen. Intersection testing between the ray and the geometric primitives in the scene solves for whether the ray hits geometry. An important advantage of ray tracing over rasterization is the ability to represent reflection and refraction, which is done by propagating rays from the points of intersection and tracking their journey through the rest of the 3D scene. In a highly detailed scene, millions of individual rays may be required.

ally does. By tracing individual light rays back to a light source, it is possible to account in a reasonably natural way for the actual physics of light, not just reflection and refraction but also specialized effects like color bleeding and caustics.

Ray tracing is an elegant algorithm, quite simple to specify in code—Paul Heckbert, now a 3D graphics architect at the video card vendor NVIDIA, coded instructions for a functional ray tracer that can be printed, just legibly, on a business card. (The feat was stimulated by a contest in which, it is gleefully reported, “repulsive C code tricks” were unveiled.) The natural way in which ray tracing deals with lighting makes it an obvious candidate to replace rasterization, but a simple algorithm does not necessarily correlate with rapid production of a finished image. The speed of rasterization derives from capturing the visible features of a triangle, then forgetting the triangle as it moves on to the next one. Ray tracing must take account of an entire scene in which light rays bounce

around. In ray-tracing algorithms, it is necessary to process all of the triangles in the scene and then convert the data into an *acceleration structure*, a configuration of the data that optimizes the ability to determine if a given light ray hits a triangle. Different lighting effects may benefit from different acceleration structures. At every step in graphics rendering, researchers are exploring ways to optimize the calculations.

Ray tracing is unavoidably a highly computation-intensive algorithm. Because it tracks the path of every individual ray of light that illuminates a scene, it may be necessary to trace several million rays for a single image. If more advanced effects are incorporated, the number of rays can multiply substantially. The benefit that seduces researchers is the beauty of the images that result. For example, the imperfect lighting of lesser schemes can be replaced by the breathtaking realism of *global illumination*, in which environments are lit, as in reality, not by one or a few light sources, but by all the surfaces that reflect diffuse light back into the scene.

Given the speed advantages of nearly-good-enough rasterization and the computational challenges of better-than-good-enough ray tracing, the next generation GPUs are likely to support both algorithms. The current market for GPUs is thoroughly dominated by three vendors, Intel, NVIDIA and AMD/ATI, which in 2008 represented 97.8 percent of market share. These companies are known to be betting on an evolutionary approach to existing architectures, in which increased programmability will allow ray tracing to be implemented as a complement to rasterization. The world's largest chipmaker, Intel, embarked on the development of a completely new architecture codenamed Larrabee, a "many-core compute engine" based on Intel's highly successful x86 CPU architecture, the processor family used in both PC and Mac computers. The Larrabee architecture has been called a general-purpose GPU, indicative of the blurring boundary between GPUs and CPUs. While supporting traditional GPU functions like rasterized graphics, hybrid CPU features

of the Larrabee can be used to carry out tasks such as ray tracing and advanced physics calculations. (A pleasing side effect of the thriving consumer market, in which competition for the millions of graphics cards purchased each year drives down prices, is the availability of inexpensive, high-performance computing power for other purposes, such as scientific computing.) In December 2009, Intel announced that the first graphics product based on the Larrabee architecture will not be a consumer product as originally planned. Instead, the hardware will be released as a software development platform that will be used by Intel and others to explore the potential of many-core applications. This is a familiar stage in the development of computer graphics over the years, as consumer desires drive the development of more muscular hardware, and hardware developments drive the advance of software applications like real-time ray tracing that come into reach on the new architectures.

The progression from fixed-function to highly programmable GPUs, and

now to architectures with minimal fixed-function hardware, is a sign of the *wheel of reincarnation*, in which functionality is transferred from the CPU to special-purpose hardware for performance reasons, followed by power-craving expansion of the subsidiary unit. The process was first described and named by Todd Myer and Ivan Sutherland as early as 1968:

We approached the task [of creating a graphics processor] by starting with a simple scheme and adding commands and features that we felt would enhance the power of the machine. Gradually the processor became more complex. We were not disturbed by this because computer graphics, after all, *are* complex. Finally the display processor came to resemble a full-fledged computer . . .

To escape the wheel of reincarnation, Myer and Sutherland suggested that if an architecture needs more computational power, it should be added to the core of the system, rather than spurring the creation of special-purpose



Figure 7. Many rendering effects depend on multipass rendering, with information from each pass combined in a final image. The top left image gives a striking view of depths in the scene using a specialized algorithm to capture shadow information. Upper right shows diffuse color without shadows. The two images are combined at bottom left, and at bottom right additional lighting information such as specularities (shininess) dramatically improves the realism of the image. (Images courtesy of Crytek GmbH.)



Figure 8. Computer graphics researchers probe reality for the delicate effects that make or break the realism of an image. Caustics appear when light that is reflected (*left*) or refracted (*right*) accumulates or cancels, generating exotic shapes and hues that the observer may not recognize, but expects. (Photograph courtesy of Tomas Akenine-Möller.)

hardware units. We may be seeing that in the emergence of multiple processors, multiple cores within processors, and enabling architectures that increasingly support parallel processing.

When?

Current graphics hardware is capable of processing several tens of millions of rays per second. Although this sounds impressive, it is still far from the required number of rays for a modern game setup. Modern games rendering at 60 frames per second in high-definition resolution, 1920×1080 pixels, with, let us say, 16 rays per pixel for all lighting effects, require $60 \times 1920 \times 1080 \times 16 = 2$ billion rays per second, which is approximately two orders of magnitude more than current hardware can deliver. One obvious strategy to overcome this challenge is to increase the capability of the hardware. A great advantage of ray tracing is that it is a highly parallel algorithm—it has been called “embarrassingly parallel.” Each ray can be traced independently. This is significant since it allows ray tracing to exploit the parallel nature of GPUs; if 100 processors in parallel cannot complete the job quickly enough, perhaps 1,000 can.

NVIDIA, AMD/ATI and Intel are all betting on parallel computing. The latest GPUs contain hundreds of individual compute units, each capable of tracing individual rays. Intel’s Larrabee architecture uses a hybrid strategy

in which multiple x86-derived processors use specialized vector processing to trace batches of rays simultaneously. This approach is quite challenging to program and it is still unknown if ray tracing can utilize the hardware to its full potential, but promising work has been done on current CPUs. Yet the challenge is not to be underestimated. Moore’s law, which has predicted the progress in computer power over the past 40 years, says that transistor density will double every two years. Due to performance increases in transistors, this can be translated to a doubling in computer performance every 18 months. If Moore’s law holds, then, it will take roughly 10 years before consumer machines are capable of tracing the few billion rays required to render the game setups that are currently available. And in 10 years, the requirements for games and real-time graphics in general might be different, perhaps calling for higher resolutions or yet-to-be-thought-of algorithms.

Hybrid Future

Skeptics may claim, with some justification, that real-time ray tracing is a pipe dream that will never be realized; the hardware will always be too slow. Even if the hardware becomes fast enough to handle 16 rays per pixel in a full-resolution scene, that may not be enough to achieve all the lighting effects that photorealistic ray tracing may call for. With this in mind there is a growing

train of thought that the future may be a hybrid approach that combines both rasterization and ray tracing.

Combining rasterization and ray tracing is an old idea in computer graphics. The basic approach uses rasterization to decide which triangles can be seen on the screen and then uses ray tracing to perform the shading calculations. This method can be used with current GPU hardware, employing ray tracing selectively to add reflections and refractions in strategic places. There is little doubt that future generations of real-time graphics for games will use this approach for as long as the pure ray-tracing approach is unattainable on available hardware.

Pixar uses a hybrid rendering technique to create its movies based on the Reyes algorithm, an advanced form of rasterization. (Reyes is an acronym for “renders everything you ever saw.”) Reyes generates micropolygons—scene geometry is tessellated at render time into pixel-sized triangles or quadrilaterals. The use of micropolygons makes it possible to create complex geometric effects through the use of displacement mapping—similar to the bump mapping described earlier except that it actually displaces the geometry, on a tiny scale, rather than just giving the appearance of displacement. This is a powerful way of creating details such as the pores on human skin, although it can generate significantly more complex geometry than current ray-tracing algorithms can deal with. Micropolygon rendering can be practical on GPUs, and if future games were to use micropolygon rendering, the visual quality of a game could be similar to that of the movie *Toy Story*. However, micropolygon rendering fails at simulating the same lighting effects that limit rasterization. Pixar’s response has been to use ray tracing coupled with micropolygon rendering in a hybrid setup. But when making its movies, Pixar doesn’t have to worry about how long it takes to render a frame.

There is another alternative to ray tracing—trick the human observer. Perhaps it is not necessary to have fully accurate lighting and reflections in the next generation of games. This is the approach that current games use. The real-time graphics community has developed many tricks that deliver great-looking graphic images in real time. For example, NVIDIA has shown a demo of human skin rendered with

subsurface scattering running in real time on a GPU. Clever filtering techniques generated rendered images that looked very convincing; few people could see the difference between their result and a ray-traced image. However, an approach based on tricks has limitations. Each trick is usually highly specialized and often does not mix well with other tricks. For example, it would likely require acrobatic coding to simulate indirect lighting on a human face with simulated subsurface scattering. This ultimately is what makes ray tracing attractive. It scales very well with the addition of processing power, and it is trivial to account for advanced lighting effects by simply tracing more rays.

The annual SIGGRAPH conference (Special Interest Group, Graphics) is the premier venue for computer graphics research. At the August 2009 SIGGRAPH, the crowd-pleasing Computer Animation Festival component of the program presented the debut of a new session, Real-Time Rendering,

in which developers demonstrated their most advanced real-time games and other applications alongside the ground-breaking prerendered works that are the staple of the conference. NVIDIA and Intel both demonstrated real-time ray tracing on their hardware. Intel, using their current-generation CPU architecture, code-named Nehalem and released in late 2008, demonstrated a ray-traced game scenario running at approximately 15 frames per second, featuring a sea bottom visible through the shimmering surface of a lagoon. Progress is being made.

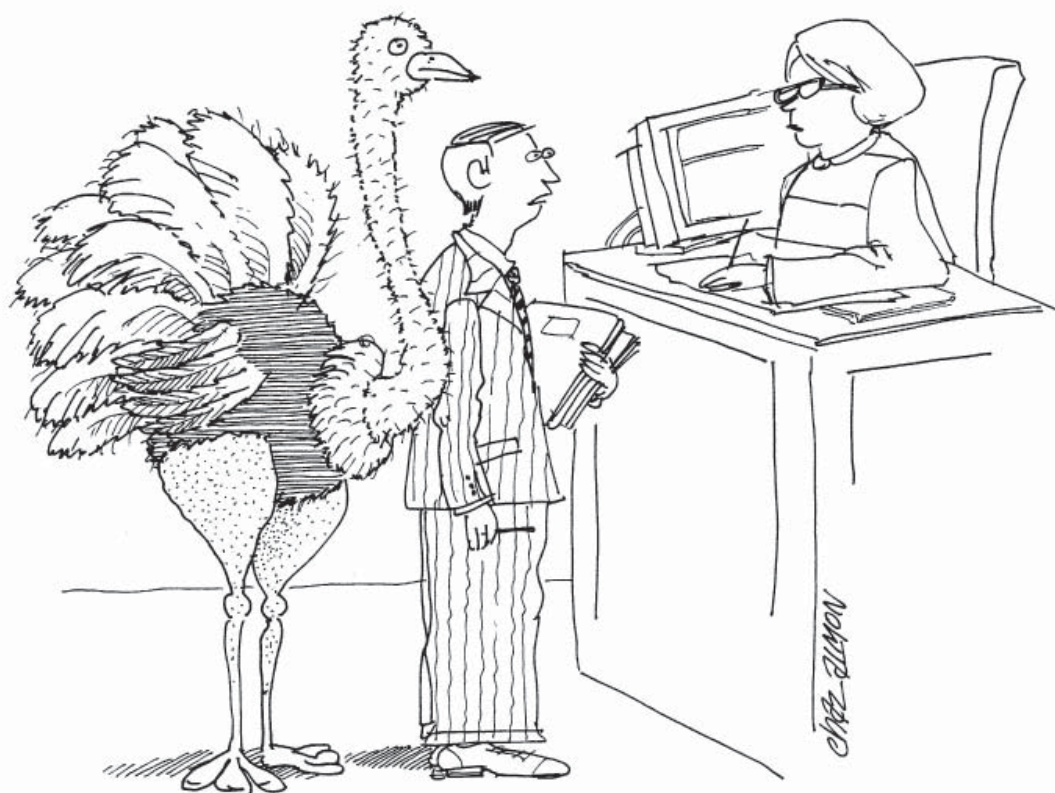
Some years ago, veteran game developer Billy Zelnack said, with hopeful irony, "Pretty soon, computers will be fast." Those words remain as true today as the day they were spoken. We add this, with less ambiguity: "Pretty soon, photorealism will be real-time."

References

- Akenine-Möller, Tomas, Eric Haines and Naty Hoffman. 2008. *Real-Time Rendering*, 3d ed., A. K. Peters Ltd.
- Jensen, Henrik Wann. 2001. *Realistic Image Synthesis Using Photon Mapping*, A. K. Peters.
- Myer, T. H., and I. E. Sutherland. 1968. On the Design of Display Processors. *Communications of the ACM* 11:410–414.
- Pharr, Matt, and Greg Humphreys. 2004. *Physically Based Rendering: From Theory to Implementation*. Morgan Kaufmann.
- Seiler, Larry, Doug Carmean, Eric Sprangle, Tom Forsyth, Michael Abrash, Pradeep Dubey, Stephen Junkins, Adam Lake, Jeremy Sugerman, Robert Cavin, Roger Espasa, Ed Grochowski, Toni Juan, and Pat Hanrahan. 2008. Larrabee: A Many-Core x86 Architecture for Visual Computing, *ACM Transactions on Graphics* 27:18.1–18.15.
- Whitted, Turner. 1980. An Improved Illumination Model for Shaded Display. *Communications of the ACM* 23:343–349.

For relevant Web links, consult this issue of *American Scientist Online*:

<http://www.americanscientist.org/issues/id.83/past.aspx>



"We request low bail as my client is not a flight risk."

Gene-Culture Coevolution and Human Diet

Rather than acting in isolation, biology and culture have interacted to develop the diet we have today

Olli Arjamaa and Timo Vuorisalo

Few would argue against the proposition that in the animal kingdom adaptations related to food choice and foraging behavior have a great impact on individuals' survival and reproduction—and, ultimately, on their evolutionary success. In our own species, however, we are more inclined to view food choice as a cultural trait not directly related to our biological background. This is probably true for variations in human diets on small scales, manifested both geographically and among ethnic groups. Some things really are a matter of taste rather than survival.

On the other hand, some basic patterns of our nutrition clearly are evolved characters, based on between-generation changes in gene frequencies. As Charles Darwin cautiously forecast in the last chapter of *On the Origin of Species*, his theory of natural selection has indeed shed "some light" on the evolution of humans, including the evolution of human diet. The long transition from archaic hunter-gatherers to post-industrial societies has included major changes in foraging behavior and human diet.

Olli Arjamaa received his Ph.D. in animal physiology at the University of Turku in 1983, and his M.D. at the University of Oulu in 1989 (both in Finland). He is adjunct professor at the Center of Excellence of Evolutionary Genetics and Physiology, Department of Biology, University of Turku. His main research interest is the evolutionary physiology of natriuretic peptides. Timo Vuorisalo received his Ph.D. in ecological zoology at the University of Turku in 1989. In 1989–1990 he was a visiting postdoctoral fellow at the Indiana University, Bloomington. He is senior lecturer of Environmental Science and adjunct professor in the Department of Biology, University of Turku. His research interests include evolutionary ecology, environmental history and urban ecology. Address: Department of Biology, 20014 Turun yliopisto, Finland. Email: olli.arjamaa@utu.fi

The traditional view holds that our ancestors gradually evolved from South and East African fruit-eaters to scavengers or meat-eaters by means of purely biological adaptation to changing environmental conditions. Since the 1970s, however, it has become increasingly clear that this picture is too simple. In fact, biological and cultural evolution are not separate phenomena, but instead interact with each other in a complicated manner. As Richard Dawkins put it in *The Selfish Gene*, what is unusual about our species can be summed up in one word: *culture*.

A branch of theoretical population genetics called gene-culture coevolutionary theory studies the evolutionary phenomena that arise from the interactions between genetic and cultural transmission systems. Some part of this work relies on the sociobiologically based theoretical work of Charles J. Lumsden and E. O. Wilson, summarized in *Genes, Mind, and Culture*. Another branch of research focuses on the quantitative study of gene-culture coevolution, originated among others by L. L. Cavalli-Sforza and M. W. Feldman. Mathematical models of gene-culture coevolution have shown that cultural transmission can indeed modify selection pressures, and culture can even create new evolutionary mechanisms, some of them related to human cooperation. Sometimes culture may generate very strong selection pressures, partly due to its homogenizing influence on human behavior.

A gene-culture coevolutionary perspective helps us to understand the process in which culture is shaped by biological imperatives while biological properties are simultaneously altered by genetic evolution in response to cultural history. Fascinating examples of such gene-culture coevolution can be found in the evolution of human

diet. Richard Wrangham's recent book, *Catching Fire: How Cooking Made Us Human*, focused on impacts of taming fire and its consequences on the quality of our food. Some scholars favor a *memetic* approach to this and other steps in the evolution of human diet. Memetics studies the rate of spread of the units of cultural information called memes. This term was coined by Dawkins as an analogy to the more familiar concept of the gene. A meme can be, for instance, a particular method of making fire that makes its users better adapted to utilize certain food sources. As a rule, such a meme spreads in the population if it is advantageous to its carriers. Memes are transmitted between individuals by social learning, which, as we all know, has certainly been (and still is) very important in the evolution of human diet.

In the following paragraphs, we will review the biological and cultural evolution of hominid diets, concluding with three examples of cultural evolution that led to genetic changes in *Homo sapiens*.

The First Steps in the Savanna

The first hominid species arose 10 to 7 million years ago in late Miocene Africa. In particular, *Sahelanthropus tchadensis*, so far the oldest described hominid, has been dated to between 7.2 and 6.8 million years. Hominids probably evolved from an ape-like tree-climbing ancestor, whose descendants gradually became terrestrial bipeds with a substantially enlarged brain. The overall picture of human evolution has changed rather dramatically in recent years, and several alternative family trees for human origins have been proposed.

The major ecological setting for human evolution was the gradually drying climate of late-Miocene and Pliocene Africa. Early hominids re-



Donald Nausbaum/Corbis



Jeremy Homer/Corbis

Figure 1. Genetic and cultural evolution are often thought of as operating independently of each other's influence. Recent investigations, however, show that this is far too simple a picture. Cultural preferences for certain foods, for example, may favor genetic changes that help people utilize them. One example is the practice of animal husbandry for milk production, which can cause the frequency of lactose tolerance—the ability to process this milk sugar as an adult—to vary geographically even within continents. Although only about 3 percent of people in Thailand (*top*) have lactose tolerance, the proportion in northern India, where dairy activity is common (*above*), is about 70 percent.

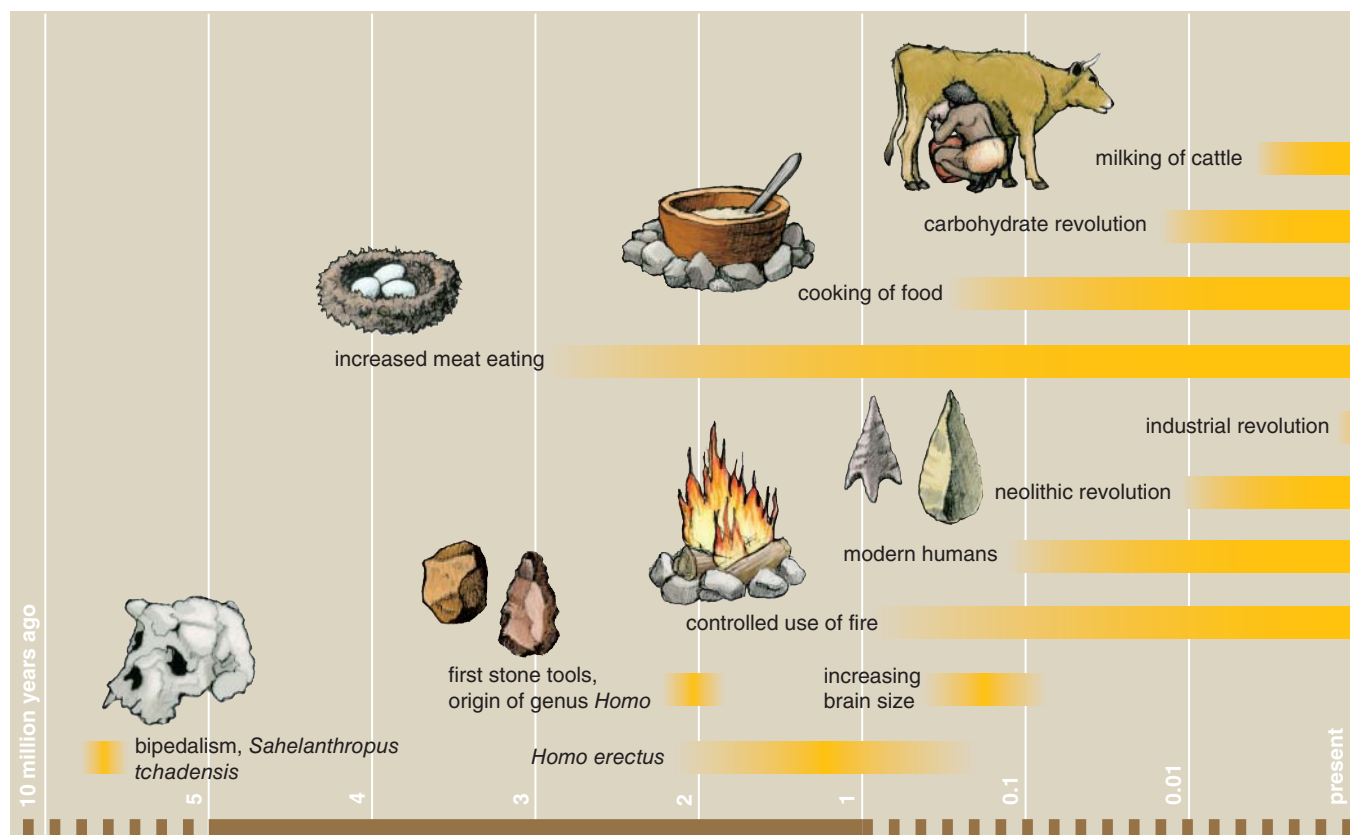


Figure 2. Major events in hominid evolution can be viewed from a gene-culture coevolution perspective. (Note the logarithmic scales at both ends of the time line, *brown dashes*.) Contrary to popular belief, bipedality did not evolve to free hands for manufacturing and use of tools (an example of old teleological thinking not accepted by scientists). In fact, upright posture preceded tool-making by at least 2 million years. Indeed, the celebrated and well-preserved specimen of *Ardipithecus ramidus*, seems to have moved upright already 4.4 million years ago, and the same may have been true for the much older *Sahelanthropus tchadensis*. Bipedality, increasingly complex social behavior, tool-making, increased body size and dietary changes formed an adaptive complex that enhanced survival and reproduction in the changing African environment. Controlled use of fire had a great impact on the diet of our ancestors and helped colonization of all main continents by our species. More recently, the dietary shifts following the Neolithic Revolution provide fascinating examples of the interplay of cultural change and biological evolution.

sponded to the change with a combination of biological and cultural adaptations that together enhanced their survival and reproduction in the changing environment. This adaptive complex probably included increas-

ingly sophisticated bipedality, complex social behavior, making of tools, increased body size and a gradual change in diet. In part, the change in diet was made possible by stone tools used to manipulate food items. The oldest known stone tools date back to 2.6 million years. Stone tool technologies were certainly maintained and spread by social learning, and very likely the same was true for changes in foraging tactics and choice of food.

The main data sources on hominid paleodiets are fossil hominid remains and archaeological sites. Well-preserved fossils allow detailed analyses on dental morphology and microwear, as well as the use of paleodietary techniques that include stable isotope analysis of bone and dentine collagen, as well as enamel apatite. Other useful and widely applied methods include comparisons of fossils with extant species with known dental morphology and diets. The main problem with dental morphology and

wear analyses is that they indicate the predominant type of diet rather than its diversity. Thus, it is always useful to combine paleodietary information from many sources. Archaeological sites may provide valuable information on refuse fauna, tools and home-range areas of hominids, all of which have implications for diet.

Much recent attention has been focused on stable isotope analysis of bone and collagen. These techniques allow comparisons of animals consuming different types of plant diets. This is important, as plant remains seldom fossilize, so the proportion of animals in the diets of early hominids is easily exaggerated. In stable isotope analysis it may be possible to distinguish between diets based on C_3 plants and those based predominantly on C_4 plants. C_3 and C_4 are two different biochemical pathways for carbon fixation in photosynthesis. Plants that utilize the C_3 photosynthetic pathway discriminate against ^{13}C , and as a re-



Museum of Anthropology, University of Missouri

Figure 3. The use of stone tools contributed to the dietary change in our ancestors. Sharp-edged stone tools could slice through the hides of hunted or scavenged animals, thus allowing access to meat. Skulls and bones could be smashed by stone tools, which provided access to nutritious tissues such as bone marrow or brain.

sult C_3 plants have clearly depleted $^{13}C/^{12}C$ ratios. In contrast, plants that utilize the C_4 photosynthetic pathway discriminate less against ^{13}C and are, therefore, in relative terms, enriched in ^{13}C . C_4 plants are physiologically better adapted to conditions of drought and high temperatures, as well as nitrogen limitation, than are C_3 plants. Thus it is very likely that the drying climate of Africa increased the abundance and diversity of C_4 plants in relation to C_3 plants.

The traditional view on early hominids separated them into australopithecines that were considered predominantly fruit eaters, and species of the genus *Homo*—that is, *H. habilis* and *H. erectus*—who were either scavengers or hunters. This traditional separation has been challenged by paleodietary techniques that have highlighted the importance of changes in the makeup of plant diet outlined above. While the ancestral apes apparently continued to exploit the C_3 plants abundant in forest environments, the australopithecines broadened their diet to include C_4 foods, which together with bipedalism allowed them to colonize the increasingly open and seasonal African environment.

This emerging difference in diet very likely contributed to the ecological diversification between apes and hominids, and was an important step in human evolution. The C_4 plants foraged by australopithecines may have included grasses and sedges, although the topic is rather controversial. Interestingly, the use of animals as food sources may also result in a C_4 -type isotopic signature, if the prey animals have consumed C_4 plants. Many researchers believe that a considerable proportion of the diet of australopithecines and early *Homo* consisted of arthropods (perhaps largely termites), bird eggs, lizards, rodents and young antelope, especially in the dry season.

Brain Size, Food and Fire

Progressive changes in diet were associated with changes in body size and anatomy. As Robert Foley at the University of Cambridge has pointed out, increased body size may broaden the dietary niche by increasing home-range area (thus providing a higher diversity of possible food sources) and enhanced tolerance of low-quality foods. A large mammal can afford to subsist off lower-quality foods than a

small mammal. Moreover, increased body size enhances mobility and heat retention, and may thus promote the ability to adapt to cooler climates. All these possibilities were realized in the hominid lineage.

In particular, the origin of *H. erectus* about 1.8 million years ago appears to have been a major adaptive shift in human evolution. *H. erectus* was larger than its predecessors and was apparently the first hominid species to migrate out of Africa. It also showed a higher level of encephalization (skull size relative to body size) than seen in

any living nonhuman primate species today. Increased brain size, in turn, was associated with a change in diet. The increase in brain size probably started about 2.5 million years ago, with gradual transition from *Australopithecus* to *Homo*. Because of the proportionately high energy requirements of brain tissue, the evolution of large human brain size has had important implications for the nutritional requirements of the hominid species. According to the Expensive-Tissue Hypothesis, proposed in 1995 by Leslie Aiello with University College London and Peter

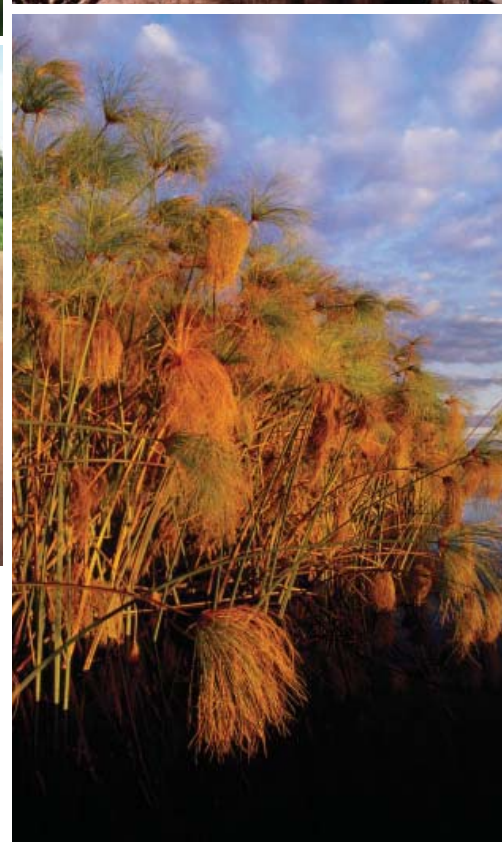


Figure 4. Stable carbon isotope analyses show that early African hominids had a significant C_4 component in their diet. This may have come either from eating C_4 plant foods or from eating animals (for example, termites) that consumed C_4 plants. Common C_3 plants include (clockwise from top left) rice and cassava root. A well-known C_4 plant is the giant sedge *Cyperus papyrus*, which was used as a food source by ancient Egyptians. Teff is a common C_4 plant in Africa today.

Kyle S. Brown, Institute for Human Origins (IHO)



Figure 5. Controlled use of fire was key to changes in hominid diet. Although examples may date back as far as 400,000 years ago, it probably was not common until roughly 50,000 to 100,000 years ago, as shown in these examples of heat-treated silcrete blade tools from the circa 65,000–60,000-year-old layers at Pinnacle Point Site 5-6 (PP5-6) in Africa.

Wheeler with Liverpool John Moores University, the high costs of large human brains are in part supported by energy- and nutrient-rich diets that in most cases include meat.

Increased use of C_4 plants was indeed gradually followed by increased consumption of meat, either scavenged or hunted. Several factors contributed to increased meat availability. First, savanna ecosystems with several modern characteristics started to spread about 1.8 million years ago. This benefited East African ungulates, which increased both in abundance and species diversity. For top predators such as *H.*

erectus this offered more hunting and scavenging possibilities. The diet of *H. erectus* appears to have included more meat than that of australopithecines, and early *Homo*. *H. erectus* probably acquired mammalian carcasses by both hunting and scavenging. Archaeological evidence shows that *H. erectus* used stone tools and probably had a rudimentary hunting and gathering economy. Sharp-edged stone tools were important as they could slice through hide and thus allowed access to meat. These tools also made available tissues such as bone marrow or brain. Greater access to animal foods seems to have

provided the increased levels of fatty acids that were necessary for supporting the rapid hominid brain evolution.

As Richard Wrangham has persuasively argued, domestication of fire had a great influence on the diet of our ancestors. Fire could be used in cooperative hunting, and to cook meat and plants. According to hominid fossil records, cooked food may have appeared already as early as 1.9 million years ago, although reliable evidence of the controlled use of fire does not appear in the archaeological record until after 400,000 years ago. The routine use of fire probably began around 50,000 to 100,000 years ago. Regular use of fire had a great impact on the diet of *H. erectus* and later species, including *H. sapiens*. For instance, the cooking of savanna tubers and other plant foods softens them and increases their energy and nutrient bioavailability. In their raw form, the starch in roots and tubers is not absorbed in the intestine and passes through the body as nondigestible carbohydrate. Cooking increases the nutritional quality of tubers by making more of the carbohydrate energy available for biological processes. It also decreases the risk of microbial infections. Thus, the use of fire considerably expanded the range of possible foods for early humans. Not surprisingly, the spread of our own species to all main continents coincides with the beginning of the routine use of fire.

In relative terms, consumption of meat seems to have peaked with our sister species *H. neanderthalensis*. As

	Paleolithic	Finnish young adults
proteins	30%	17%
carbohydrates	35%	49%
fat	35%	32%
alcohol	none	2%

Figure 6. The carbohydrate revolution began with the domestication of plants and animals about 12,000 years ago. Diets prior to the Neolithic differed considerably from what most people eat today. The contribution of protein to caloric intake (for example, salmon as shown in the Finnish spread at right) declined significantly. In place of the missing protein came carbohydrates such as potatoes. This may have driven an increase in the copies of the human amylase salivary enzyme gene.



Richard Wareham Fotografie/Alamy

Matt Sponheimer and Julia A. Lee-Thorp with Rutgers University and the University of Cape Town have pointed out, on the basis of extensive evidence, “there can be little doubt that Neanderthals consumed large quantities of animal foods.” Remains of large to medium-sized mammals dominate Neanderthal sites. Neanderthals probably both hunted and foraged for mammal carcasses. Perhaps, unromantically, they had a preference for small prey animals when hunting. And in northern areas colonized by the Neanderthals, there was probably no competition for frozen carcasses. The control of fire by the Neanderthals (and archaic modern humans), however, allowed them to defrost and use such carcasses.

The Carbohydrate Revolution

The Neolithic or Agricultural Revolution, a gradual shift to plant and animal domestication, started around 12,000 years ago. For our species this cultural innovation meant, among many other things, that the proportion of carbohydrates in our diet increased considerably. Cereal grains have accounted for about 35 percent of the energy intake of hunter-gatherer societies, whereas it makes up one-half of energy intake in modern agricultural societies—for example, in Finnish young adults (see Figure 6). The Neolithic Revolution also included domestication of mammals, which in favorable conditions guaranteed a constant supply of meat and other sources of animal protein.

Although fire likely played a role in the early utilization of carbohydrates, the big shift in diet brought about by plant domestication has its roots in the interplay of cultural change and biological evolution. Sweet-tasting carbohydrates are energy rich and therefore vital for humans. In the environment of Paleolithic hunter-gatherer populations, carbohydrates were scarce, and therefore it was important to effectively find and taste sweet foods. When eaten, large polymers such as starch are partly hydrolyzed by the enzyme amylase in the mouth and further cleaved into sugars, the sweet taste of which might have functioned as a signal for identifying nutritious food sources. (It is interesting to note that the fruit fly *Drosophila melanogaster* perceives the same compounds as sweet that we do.) Later, in the Neolithic agriculture, during which humans shifted to con-

sumption of a starch-rich diet, the role of the amylase enzyme in the digestive tract became even more important in breaking down starch.

Salivary amylase is a relatively recent development that first originated from a pre-existing pancreatic amylase gene. A duplication of the ancestral pancreatic amylase gene developed salivary specificity independently both in rodents and in primates, emphasizing its importance in digestion. Additionally, its molecular biology gives us a new insight into how evolution has made use of copy number variations (CNVs, which include deletions, insertions, duplications and complex multisite variants) as sources of genetic and phenotypic variation; single-nucleotide polymorphisms (SNPs) were once thought to have this role alone. CNVs may also involve complex gains or losses of homologous sequences at multiple sites in the genome, and structural variants can comprise millions of nucleotides with heterogeneity ranging from kilobases to megabases in size.

Analyses of copy number variation in the human salivary amylase gene (*Amy1*) found that the copy number correlated with the protein level and that isolated human populations with a high-starch diet had more copies of *Amy1*. Furthermore, the copy number and diet did not share a common ancestry; local diets created a strong positive selection on the copy number variation of amylase, and this evolutionary sweep may have been coincident with the dietary change during early stages of agriculture in our species. It is interesting to note that the copy number variation appears to have increased in the evolution of human lineage: The salivary protein levels are about six to eight times higher in humans than in chimpanzees and in bonobos, which are mostly frugivorous and ingest little starch compared to humans.

Transition to Dairy Foods

A classic example of gene-culture co-evolution is lactase persistence (LP) in human adults. Milk contains a sugar named lactose, which must be digested by the enzyme lactase before it can be absorbed in the intestine. The ability to digest milk as adults (lactose tolerance) is common in inhabitants of Northern Europe where ancient populations are assumed to have used milk products as an energy source to survive the cold

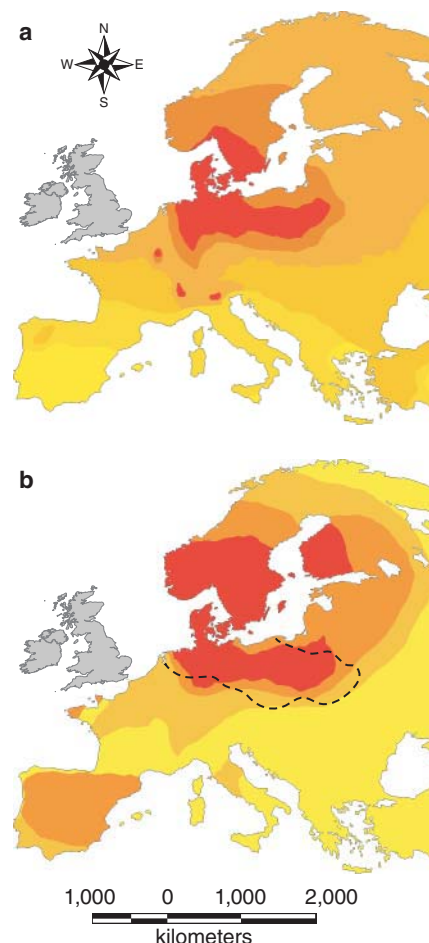


Figure 7. Albano Beja-Pereira and colleagues have done geographic matching between milk gene diversity in cattle, lactose tolerance in contemporary humans and locations of Neolithic cattle farming sites. The dark orange color in *a* shows where the greatest milk gene uniqueness and allelic diversity occur in cattle. In *b* lactase persistence is plotted in contemporary Europeans. The darker the color, the higher the frequency of the lactase persistence allele. The dashed line in *b* shows the geographic area in which the early Neolithic cattle pastoralist culture emerged. (Image adapted from Beja-Pereira et al. 2003.)

and dark winters, whereas in southern Europe and much of Asia, drinking milk after childhood often results in gastrointestinal problems. If the intestine is unable to break down lactose to glucose and galactose—due to lack of lactase or lactase-phlorizin hydrolase (LPH) enzyme, normally located in the villi of enterocytes of the small intestine—bacterial proccession of lactose causes diarrhea, bloating and flatulence that can lead to fatal dehydration in infants. On the other hand, milk provides adults with a fluid and rich source of energy without bacterial contamination, enhancing their survival

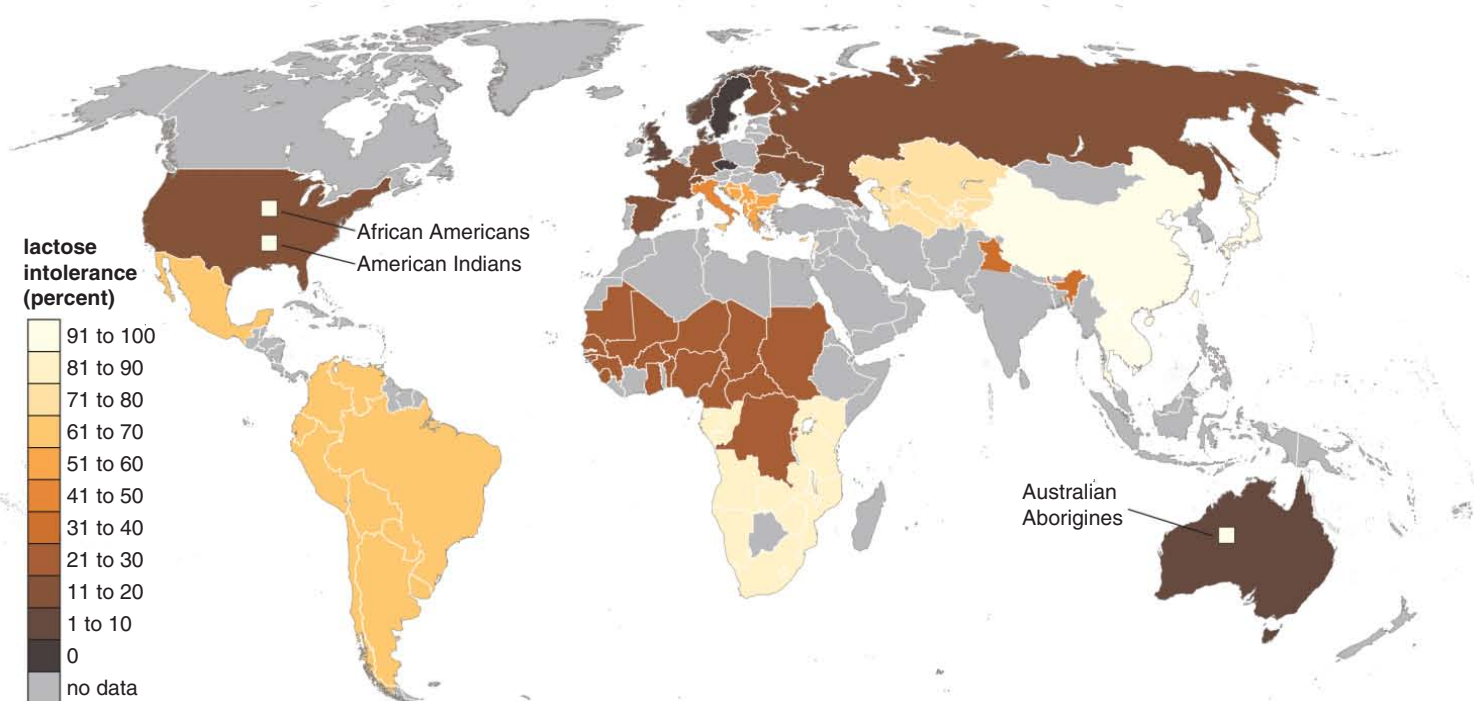


Figure 8. Lactose intolerance in adult human beings is, in fact, the rule rather than the exception, although its prevalence may well be declining as the single nucleotide polymorphism that causes lactase persistence spreads. Note the wide variation in lactose intolerance over short geographic distances. Particularly in African cultures, the prevalence of dairy farming is strongly correlated to lactose tolerance. Gray areas indicate areas where no data are available. (Map adapted from Wikimedia Commons.)

and fitness. Therefore, in the past the phenotype of lactase persistence undoubtedly increased the relative reproductive success of its carriers.

Recent findings of molecular biology show that a single-nucleotide polymorphism that makes isolated populations lactase persistent has been “among the strongest signals of selection yet found for any gene in the genome.” Lactase persistence emerged independently about 10,000 to 6,000 years ago in Europe and in the Middle East, two areas with a different history of adaptation to the utilization of milk. The earliest historical evidence for the use of cattle as providers of milk comes from ancient Egypt and Mesopotamia and dates from the 4th millennium B.C. Still today there are large areas of central Africa and eastern Asia without any tradition of milking, and many adults in these countries are physiologically unable to absorb lactose. The ancient Romans did not drink milk, and this is reflected in the physiology of their Mediterranean descendants today.

The first evidence for a SNP as a causative factor in LP came from a group of Finnish families. A haplotype analysis of nine extended Finnish families revealed that a DNA variant (C/T₋₁₃₉₁₀) located in the enhancer element upstream of the lactase gene

associated perfectly with lactose intolerance and, because it was observed in distantly related populations, suggested that this variant was very old. Later it was shown that this allele had emerged independently in two geographically restricted populations in the Urals and in the Caucasus, the first time between 12,000 and 5,000 years ago and the second time 3,000 to 1,400 years ago. Yet Saudi Arabian populations that have a high prevalence of LP have two different variants introduced in association with the domestication of the Arabian camel about 6,000 years ago. In Africa, a strong selective sweep in lactase persistence produced three new SNPs about 7,000 years ago in Tanzanians, Kenyans and Sudanese, reflecting convergent evolution during a similar type of animal domestication and adult milk consumption.

All these facts indicate that there has been a strong positive selection pressure in isolated populations at different times to introduce lactose tolerance, and this has taken place through several independent mutations, implying adaptation to different types of milking culture. Lactase persistence was practically nonexistent in early European farmers, based on the analysis of Neolithic human skeletons, but when dairy farming started in the early Neo-

lithic period, the frequency of lactase persistence alleles rose rapidly under intense natural selection. The cultural shift towards dairy farming apparently drove the rapid evolution of lactose tolerance, making it one of the strongest pieces of evidence for gene-culture coevolution in modern humans. In other words, the meme for milking had local variants, which spread rapidly due to the positive effects they had on their carriers.

We must bear in mind, however, that the transcription of a gene is under complex regulation, as is the C/T₋₁₃₉₁₀ variant: It contains an enhancer element through which several transcription factors probably contribute to the regulation of the lactase gene in the intestine. In addition, lactose tolerance in humans and the frequencies of milk protein genes in cattle appear to have also coevolved. When the geographical variation in genes encoding the most important milk proteins in a number of European cattle breeds and the prevalence of lactose tolerance in Europe were studied, the high diversity of milk genes correlated geographically with the lactose tolerance in modern Europeans and with the locations of Neolithic cattle farming sites in Europe (see Figure 7). This correlation suggests that there has been a

gene-culture coevolution between cattle and human culture leading towards larger herds with a wider distribution of gene frequencies, resulting in the selection of increased milk production and changed composition of milk proteins more suitable for human nutrition. In the future, we will know even more about the geographical evolution of LP, as it has become possible to rapidly genotype large numbers of individuals harboring lactose tolerance-linked polymorphisms producing various gastrointestinal symptoms after lactose ingestion.

We Are Still Evolving

As shown above, culture-based changes in diet (which can be called memes) have repeatedly generated selective pressures in human biological evolution, demonstrated for instance by the single nucleotide polymorphism of lactase persistence and the copy number variation of amylase. These selective sweeps took place 10,000 to 6,000 years ago when animal and plant domestication started, marking the transition from the Paleolithic to the Neolithic era. Much earlier, genetic changes were certainly associated with the dietary changes of australopithecines and *H. erectus*.

What about the future? Can we, for instance, see any selection pressure in the loci of susceptibility to diet-associated diseases? The answer seems to be yes. The risk of Type II diabetes (T2D) has been suggested to be a target of natural selection in humans as it has strong impacts on metabolism and energy production, and therefore on human survival and fitness. Genome-wide and hypothesis-free association studies have revealed a variant of the transcription factor 7-like (TCF7L2) gene conferring the risk of T2D. Later, in Finns, a similar genome-wide T2D study increased the number of variants near the TCF7L2 to 10. When refining the effects of TCF7L2 gene variants on T2D, a new variant of the same gene that has been selected for in East Asian, European and West African populations was identified. Interestingly, this variant suggested an association both with body mass index and with the concentrations of leptin and ghrelin, the hunger-satiety hormones that originated approximately during the transition from Paleolithic to Neolithic culture. In support of the notion that selection is an on-going process in

human physiological adaptation, the analysis of worldwide samples of human populations showed that the loci associated with the risk of T2D have experienced a recent positive selection, whereas susceptibility to Type I diabetes showed little evidence of being under natural selection.

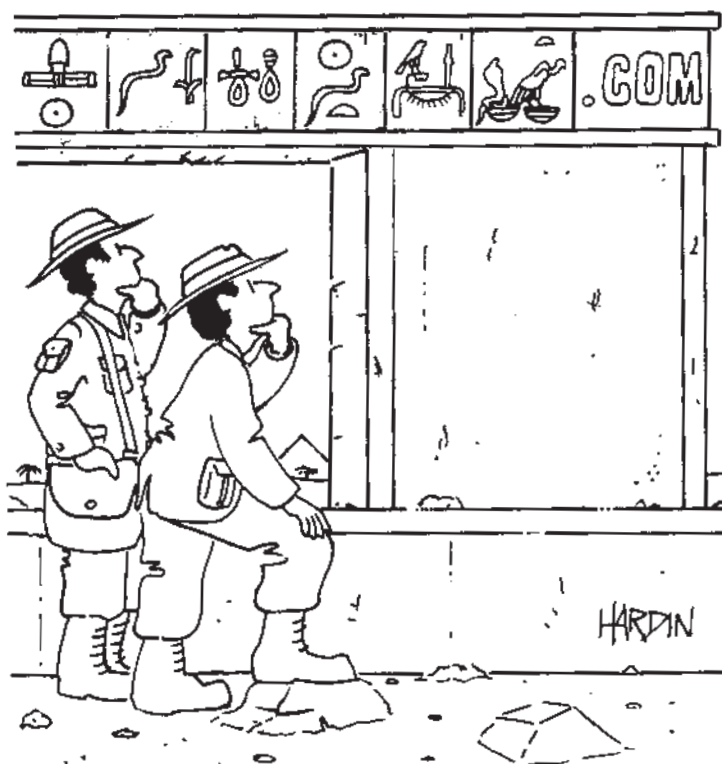
In the near future, genome-wide scans for recent positive selections will increase our understanding of the coevolution between the ancient genome and diet in different populations, projecting to problems in modern nutritional qualities. As has been suggested here, that understanding is likely to be considerably more nuanced than the simple "hunter-gatherer-genes-meet-fast-food" approach so often put forward.

References

- Beja-Pereira, A., G. Luikart, P. R. England et al. 2003. Gene-culture coevolution between cattle milk protein genes and human lactase genes. *Nature Genetics* 35:311–13.
- Bowman, D. M. J. S., J. K. Balch, P. Artaxo et al. 2009. Fire in the Earth system. *Science* 324:481–84.
- Eaton, S. B. 2006. The ancestral human diet: What was it and should it be a paradigm for contemporary nutrition? *Proceedings of the Nutritional Society* 65:1–6.
- Enattah, N. S., T. G. K. Jensen, M. Nielsen et al. 2008. Independent introduction of two lactase-persistence alleles into human populations reflects different history of adaptation to milk culture. *American Journal of Human Genetics* 82:57–72.
- Foley, R. 1987. *Another Unique Species: Patterns in Human Evolutionary Ecology*. Hong Kong: Longman Group.
- Helgason, A., S. Pálsson, G. Thorleifsson et al. 2007. Refining the impact of TCF7L2 gene variants on type 2 diabetes and adaptive evolution. *Nature Genetics* 39:218–25.
- Laland, K. N., J. Odling-Smee and S. Myles. 2010. How culture shaped the human genome: Bringing genetics and the human sciences together. *Nature Reviews Genetics* 11:137–48.
- Leonard, W. R., J. J. Snodgrass and M. L. Robertson. 2007. Effects of brain evolution on human nutrition and metabolism. *Annual Review of Nutrition* 27:311–27.
- Perry, G. H., N. J. Dominy, K. G. Claw et al. 2007. Diet and the evolution of human amylase gene copy number variation. *Nature Genetics* 39:1188–90.
- Pickrell, J. K., G. Coop, J. Novembre et al. 2009. Signals of recent positive selection in a worldwide sample of human populations. *Genome Research* 19:826–37.
- Sponheimer, M., and J. Lee-Thorp. 2007. Hominin paleodiets: the contribution of stable isotopes. In *Handbook of Paleoanthropology, Vol. I: Principles, Methods and Approaches*, eds. W. Henke and I. Tattersall. Berlin: Springer, pp. 555–85.
- Wrangham, R. 2009. *Catching Fire: How Cooking Made Us Human*. New York: Basic Books.

For relevant Web links, consult this issue of *American Scientist Online*:

<http://www.americanscientist.org/issues/id.83/past.aspx>



Finding Alzheimer's Disease

How interest in one patient's suffering and confidence in the physical basis of mental illness led a German doctor to discover the devastating disorder

Ralf Dahm

I have, so to speak, lost myself ...

—Auguste D.

Few illnesses are as devastating as Alzheimer's disease. Memory progressively fails, complex tasks become ever more difficult, and once-familiar situations and people suddenly appear strange, even threatening. Over years, afflicted patients lose virtually all abilities and succumb to the disease. Although there is no cure for Alzheimer's disease yet, scientists have made significant progress toward understanding what goes awry in the brain when neurons die on a massive scale. Recent years have seen a number of promising insights that could lead to effective therapies. But it's been a long journey to this point, one that reaches back more than a century.

The story starts in the autumn of 1901 in the German city of Frankfurt, and centers on two people. The first is Alois Alzheimer, then a 37-year-old doctor at the city's institution for the mentally ill. The second is Auguste D., a woman just over 50 years of age who recently had been admitted to the clinic. At the beginning of that year, Auguste

D.'s personality had begun to change. At first only her memory had occasionally failed her, but as time passed, her behavior changed too. She neglected household chores and, when trying to cook, blundered and ruined food. She was restless, striding around her apartment without direction or purpose, and hiding objects for no apparent reason. Increasingly, she became confused and paranoid, afraid of people she knew well. In the fall of 1901, her husband, a clerk with the railroad authority, could not cope any longer and brought her to Frankfurt's mental institution.

On November 26, 1901, one day after her admission, Alzheimer encountered Auguste D. for the first time. When he entered the room, she sat on her bed with what Alzheimer described in his notes as a "helpless" expression. To get to know her and to find out more about her affliction, Alzheimer asked her questions, writing down their exchange in his file:

What is your name?

Auguste.

Surname?

Auguste.

What is your husband's name?"

I believe Auguste.

Your husband?

I see, my husband...

Are you married?

To Auguste.

Mrs. D.?

Yes, to Auguste D.

How long have you been here?

Three weeks.

Alzheimer showed Auguste D. multiple objects, including a pencil, a pen, a key and a cigar, all of which she could name. When asked after a while what she had been shown, however, she could not remember, a clear sign of trouble forming short-term memories. Several years earlier, Alzheimer had en-

countered patients with similar symptoms and had even published an article on senile dementia in 1898. But these patients had been much older than Auguste D., whose case seemed unique. Brief as it was, this encounter would go down in history. It marks the beginning of scientific investigation into what we now know as Alzheimer's disease.

Intrigued by Auguste D.'s unusual behavior, Alzheimer observed her further. She appeared anxious and very confused. At lunch she ate pork with cauliflower. When asked what she was eating, however, she replied that it was spinach, potatoes and horseradish. Later that day, Alzheimer noticed that she made unusual errors when writing. She would omit or repeat syllables in words, sometimes several times, or abruptly stop in the middle of a phrase or word. For instance, when Alzheimer asked her to note down her name on a little piece of paper, she did not write the full name *Frau Auguste D.*, but broke off after *Frau*. Only when asked to write down every word individually was she able to complete the task. These symptoms were so unusual that Alzheimer decided to follow her case more closely.

On November 29, 1901, he interviewed Auguste D. again, diligently recording her replies:

How are you?

It is always one as the other. Who carried me here?

Where are you?

At the moment; I have temporarily, as I said, I have no means. One simply has to ... I don't know myself ... I really don't know ... dear me, what then is to?

What is your name?

Frau Auguste D.

When were you born?

Eighteen hundred and...

In which year were you born?

Ralf Dahm is director of scientific management at the Spanish National Cancer Research Centre (CNIO) in Madrid and honorary professor at the University of Padua, Italy. Since visiting Tübingen's medieval castle where DNA was discovered, he has been fascinated by the history of the life sciences. He has published on early DNA research, Darwin's theory of evolution and the discovery of Alzheimer's disease. Dahm received his Ph.D. from the Department of Biochemistry at the University of Dundee. He was a postdoctoral scientist at the Max Planck Institute for Developmental Biology in Tübingen and a group leader at the Medical University of Vienna. Address: Department of Biology, University of Padova, Via U. Bassi 58/B, I-35121 Padova, Italy. Internet: ralf.dahm@bio.unipd.it

*This year, no, last year.
When were you born?
Eighteen hundred—I don't know ...
What did I ask you?
Ach, D. Auguste.*

Clearly Auguste D. had great difficulty communicating. Alzheimer continued to test her other cognitive abilities. She performed simple calculations mostly well. But time and again she got lost or stopped speaking right in the middle of a sentence or even a word. Auguste D.'s behavior was also strange. Often she was disoriented, apparently not comprehending situations she was in. She would sometimes touch the faces of her fellow patients or pour water over them, prompting them to strike at her. When asked why, she was apologetic and replied that she was trying to "tidy up."

Taking Pains to Understand

Alzheimer's approach to examining Auguste D. was not standard during his day. At a time when mentally ill patients were often just locked away, Alzheimer and his colleagues in Frankfurt tried to understand their afflictions and help them. They carefully observed and talked to the patients, and tried to alleviate their symptoms as best they could. Instead of restraining restless patients, they encouraged them to exercise in the open air and calmed them down with warm baths or massages. Only when these measures failed did they resort to drugs. In keeping with this approach, Alzheimer early on visited Auguste D. frequently to observe her.

Over time, Auguste D.'s speech became unintelligible. She eventually stopped talking completely, only humming or shouting wildly, often for hours on end. In her final year, her body weakened. She ate only at irregular intervals, often having to be fed. She spent most of her time in bed, hunched up and apathetic. Finally, early in 1906, Auguste D. contracted pneumonia. On April 8 that year, just short of her 56th birthday, she died.

The case of Auguste D., as described by Alzheimer, accurately summarizes the range of progressive changes observed in many Alzheimer's patients today: her deteriorating memory, especially her inability to remember recent events; her disorientation; her decreased ability to speak coherently; her problems understanding and judging situations; and her restless and erratic behavior. Once, when trying and failing to write

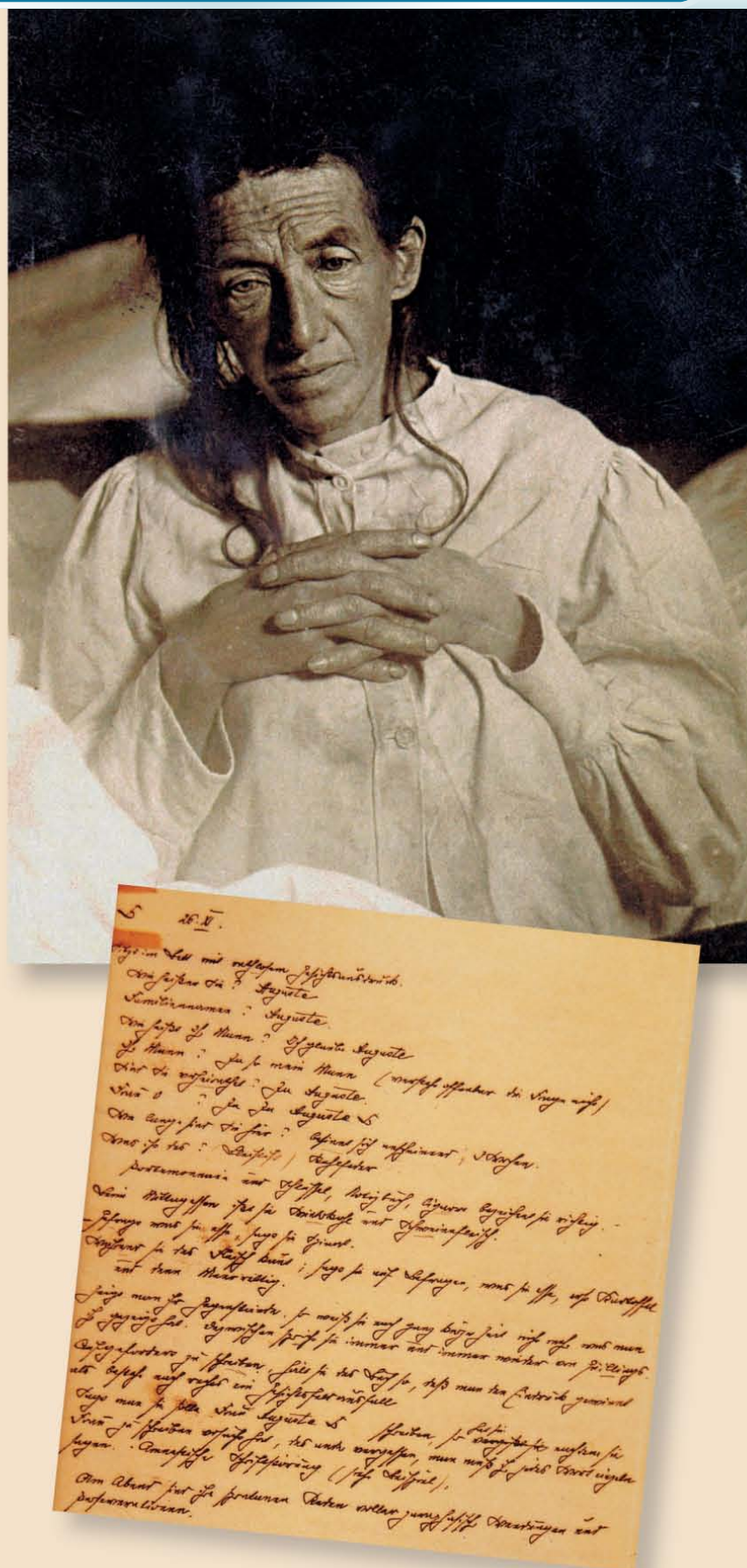


Figure 1. The research that led Dr. Alois Alzheimer to discover what we now call Alzheimer's disease started with his careful observation of a woman named Auguste D. The photograph above is dated 1902, one year after she was admitted to the mental asylum in Frankfurt, Germany, where Alzheimer worked. Auguste D.'s photograph, along with Alzheimer's notes regarding his observations of her, was rediscovered in Germany in the mid-1990s. (Images courtesy of Eli Lilly and Company.)

her name, Auguste D. remarked, "I have, so to say, lost myself." This simple statement is a fitting description of the way many Alzheimer's disease patients experience the disease.

The Right Place, the Right Time

By the time Auguste D. died, Alzheimer was no longer working in Frankfurt. In 1903, after 14 years at the institution for the mentally ill, he had accepted a position as a scientific assistant to Emil Kraepelin in Heidelberg. This was a phenomenal opportunity. Kraepelin was one of the most eminent psychia-

trists of his time. Among other important contributions, he was among those promoting the idea that psychiatric diseases have a biological basis, something acknowledged for many diseases in his day but not yet widely accepted for mental illness. By introducing experimental approaches to understanding mental afflictions, Kraepelin helped transform psychiatry into an empirical science. He developed an innovative system to classify mental disorders, which took into account not only symptoms at any given stage but also changes over time. Kraepelin's system proved

so successful that today's classification of psychiatric disorders remains largely based on it. Alzheimer knew that working with Kraepelin would open up possibilities he could only dream of in Frankfurt. Moreover, Franz Nissl, a close friend and colleague of Alzheimer's in Frankfurt, had also moved to Heidelberg. Alzheimer hoped that together they could substantially advance their studies into the anatomical causes of mental disorders.

It's difficult to pinpoint precisely when Alzheimer became so driven to expand the scientific understanding of neurological maladies. He was an enthusiastic student of the natural sciences throughout his secondary-school days in Franconia. After that, he studied medicine in Berlin, Würzburg and Tübingen, important centers for the medical and biological sciences at that time. During his studies, he had two experiences that must have influenced his later career in psychiatry. While studying in Berlin, Alzheimer came in contact with the new ideas about how mental disorders can correlate to physical changes in the brain. Also, in Würzburg, Alzheimer studied with Albert von Kölliker. A distinguished histologist and pioneer of microscopic anatomy, von Kölliker introduced Alzheimer to microscopy. The solid training in microscopic anatomy he received from von Kölliker equipped Alzheimer with the expertise he would need later to analyze the brains of patients such as Auguste D. Still, his medical thesis focused not on a brain disease but on the histology of the glands that secrete cerumen, or earwax.

After finishing his medical studies with top grades and receiving his license as a medical doctor in 1888, Alzheimer took a position as the private doctor of a mentally ill woman with whom he traveled for five months. Shortly after finishing that assignment, he answered an advertisement seeking an assistant physician at the Frankfurt mental institution, an opening he had seen before taking his first position but had not responded to. The institution's director, Emil Sioli, by then was desperate to recruit someone to help him care for the clinic's 254 patients. Only one day after receiving Alzheimer's application, Sioli sent him a telegram offering him the job. Alzheimer began work in December of 1888. A few months later, Franz Nissl joined him and Sioli as a senior physician. Nissl remains famous today for his discovery of histological stain-



Max-Planck-Institute of Psychiatry Munich, Historic Archives, Portrait collection

Figure 2. Two things especially equipped Alzheimer, pictured here in an undated portrait, to discover the progressive and devastating disease that still bears his name. For one, he embraced a school of thought that argued that many mental ailments could be traced to anomalies in the brain. He also had been trained in microscopy and histology, key skills that would allow him to analyze anatomical abnormalities in diseased brains.



© Historisches Museum Frankfurt/Main Photograph: Horst Ziegenfuss

Figure 3. As a young doctor, Alzheimer treated patients and conducted research at the mental hospital in Frankfurt, shown above. The institution was known at the turn of the last century for the humane treatment of its patients. The German neo-Gothic facility was constructed in 1864 under the auspices of the famous German psychiatrist Heinrich Hoffmann. The institution had several courtyards, gardens and even a grand ballroom. Some Frankfurt citizens called it the “palace of the mad.”

ing techniques that improved scientists’ ability to see structures in neurons and tissue of the human brain. He is also famous for his discovery of the neuronal organelles—called Nissl substance—that are sites of protein synthesis.

The three men were highly compatible. Sioli was a progressive and open-minded director who allowed his two doctors ample time to follow their research interests. Nissl and Alzheimer shared a passion for histopathology and neuropathology. They used microscopes to closely examine tissue to better understand which changes related to a particular brain disease. Encouraged by their environment, the men became close collaborators and friends.

Applying the Best Tools

After Alzheimer left Frankfurt to work with Kraepelin, Sioli kept Alzheimer informed about changes in Auguste D.’s health. When she died, Sioli shipped her brain to Alzheimer, who by then had relocated to Munich, where Kraepelin had been selected to run the Royal Psychiatric Clinic. Alzheimer ran the clinic’s large anatomical laboratory and had set up a state-of-the-art facility for histopathological analyses, which rapidly attracted a number of gifted students and guest scientists. Among them were Hans-Gerhard Creutzfeld and Alfons Maria Jakob. In the 1920s, they would be the first to describe the degenerative neurological disease that

bears their names: Creutzfeld-Jakob disease. Alzheimer’s laboratory was fitted with multiple instruments, including microscopes and a camera lucida, which allowed Alzheimer to make drawings of his histological sections, as well as a room for microphotography, which allowed him to take photographs. He also had several histological

staining methods at his service, including silver stains, which were useful for detecting subcellular structures in neurons due to their high contrast and sensitivity. For that point in history, Alzheimer was in an ideal situation to examine Auguste D.’s brain.

Alzheimer’s initial inspection confirmed his suspicion that hers was an



Men: Warrick Fuller/Schomburg Center for Research in Black Culture/NYPL

Figure 4. Alzheimer, seated to the left of U.S.-based psychiatrist Solomon C. Fuller, was working in Munich by the time Auguste D. died. He had accepted a position with eminent psychiatrist Emil Kraepelin, the person who coined the term “Alzheimer’s disease.” Alzheimer and Fuller are pictured with other psychiatrists at the University of Munich in 1904 or 1905. Only some of the other doctors’ names are legible. They include Baroncini, von Nobert and Ranke.

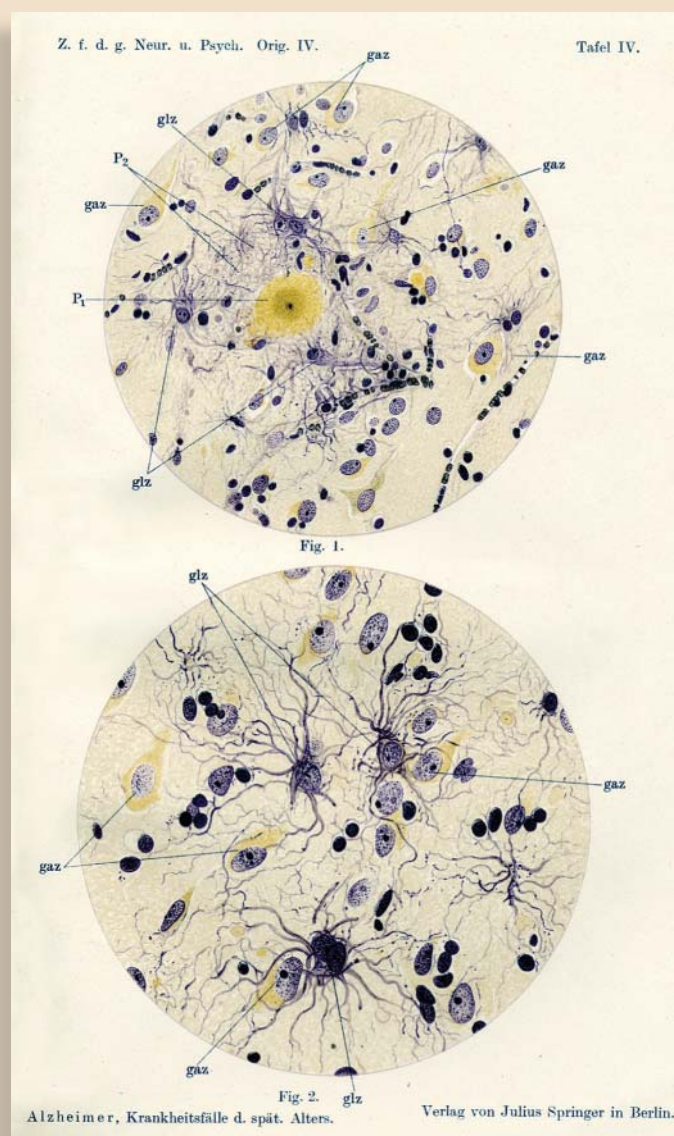
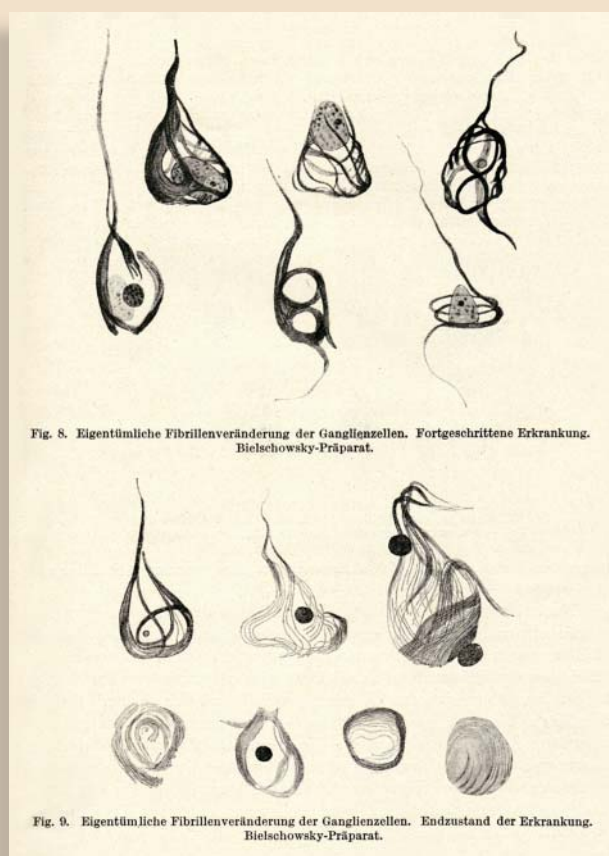
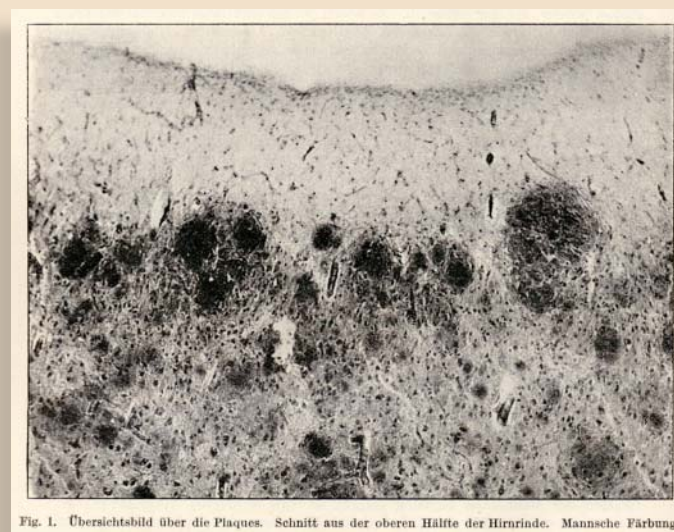
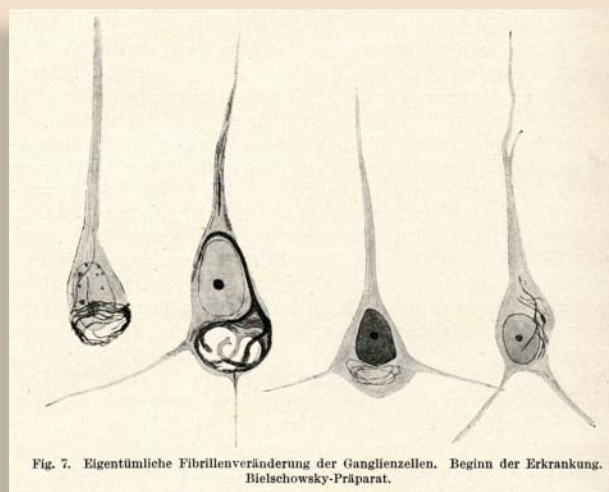


Figure 5. In a 1911 publication, Alzheimer included multiple images of what he and collaborators observed in brain sections of patients suffering from Alzheimer's disease. The panels on the left show stages in the formation of neurofibrillary tangles in the brain of Auguste D. The top panel depicts the beginnings of the process. The middle and bottom panels show intermediate and late stages, respectively. At top right is a photograph of part of a section taken from the brain of patient Johann F. The dark areas represent plaques. Below that are renderings of sections taken from different depths of Auguste D.'s cortex. Numerous plaques are depicted as are cells with strongly staining neurofibrils. In this figure, *P1* refers to the central plaque and *P2* refers to peripheral plaque regions; *glz* refers to glia cells and *gaz* refers to ganglion cells.

extraordinary case. Huge areas of her brain showed a pronounced atrophy. To study the changes in more detail, Alzheimer sectioned parts of the brain and stained them to better reveal the morphology of the tissue under a microscope. Helped by two visiting Italian physicians, Gaetano Perusini and Francesco Bonfiglio, Alzheimer confirmed the atrophy that he had observed in the intact brain. In many regions of the brain, enormous numbers of neurons had died.

In addition to the atrophy, the scientists noticed more subtle changes. Many of the remaining neurons contained peculiar, thick and strongly staining fibrils, or fibers. Throughout the cerebral cortex, they also found deposits of an unknown, gummy substance in the form of plaques. Auguste D.'s brain thus showed what today are generally seen as hallmarks of Alzheimer's disease. First, there was the massive death of neurons, and, second, the presence of neurofibrillary tangles, insoluble aggregates of a protein called *tau* that take the shape of thick, tangled fibers and fill the neuronal cell body. Third were the amyloid plaques, deposits of small peptides called beta-amyloid that form in the spaces between neurons. Although these changes are familiar to any scientist studying the disease today, they were new and exciting to Alzheimer and his colleagues.

The abnormalities were, to some extent, similar to the degenerative changes seen in senile dementia, a pathology observed in elderly patients. But there were two important differences. For one, the changes in Auguste D. had occurred in a woman who was only 51 when she showed the first signs of the disease and who was 55 when she died. Patients with senile dementia generally were in their 70s or 80s. Furthermore, the pathological changes in Auguste D.'s brain were much more dramatic than those Alzheimer had seen in patients suffering from senile dementia. Alzheimer was thus convinced that he had discovered something new.

On November 3, 1906, Alzheimer was ready to present his finding to the scientific community. He was invited to give a lecture at the 37th meeting of the South-West German Psychiatrists in the small university town of Tübingen. What might have been a provincial affair long lost in obscurity was in fact to become a defining moment in the history of neurology. In his talk entitled "On a

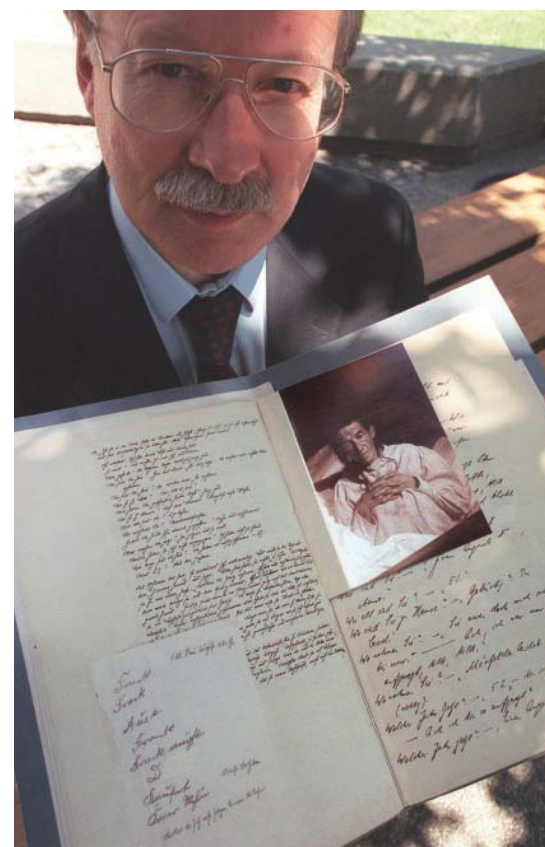
peculiar disease of the cerebral cortex," Alzheimer publicly described Auguste D.'s case for the first time. He began by relating her unusual psychiatric symptoms, noting that they were so unlike any described previously that her case did not fit any known affliction.

He then described the dramatically changed histology of Auguste D.'s brain. By showing images prepared in his laboratory of the widespread cell death, the strange, thick bundles of tangled neurofibrils, and the abundant plaques, Alzheimer hoped to convince the audience of the novelty and importance of his findings. He concluded his talk by repeating his conviction that this case was a new pathology and that histopathological analyses such as he described would allow both a more precise classification and a better understanding of all mental disorders.

Instead of responding enthusiastically to his groundbreaking discovery, however, the 87 scientists and doctors in the audience barely reacted. No one asked questions. There was no discussion. The meeting's organizers, failing to grasp the significance of the findings, noted the talk's title in its proceedings but stated, without explanation, that it "was not appropriate for a short publication." At least the local newspaper, the *Tübinger Chronik*, which published a report of the meeting two days after Alzheimer's lecture, mentioned his talk, but only in one short sentence: "Dr. Alzheimer from Munich reported of a peculiar, severe disease process which, within a period of 4 and a half years, causes a substantial loss of neurons." In the following year, though, the meeting organizers reversed their initial decision and a two-page transcript of Alzheimer's talk, without his figures, was included in the *Allgemeine Zeitschrift für Psychiatrie und psychiatrisch-gerichtliche Medizin* (*General Journal of Psychiatry and Psychiatric-Forensic Medicine*). This report—considered a historic paper today—did not stir much interest in the scientific community either.

The Psychiatrist Persists

Alzheimer was not discouraged. He remained convinced of the importance of his discovery. To gather further data in support of his views and to understand the disease better, he looked for additional cases of younger dementia patients. In 1907 and 1908, Alzheimer obtained the brains of three patients with symptoms much like those he had observed

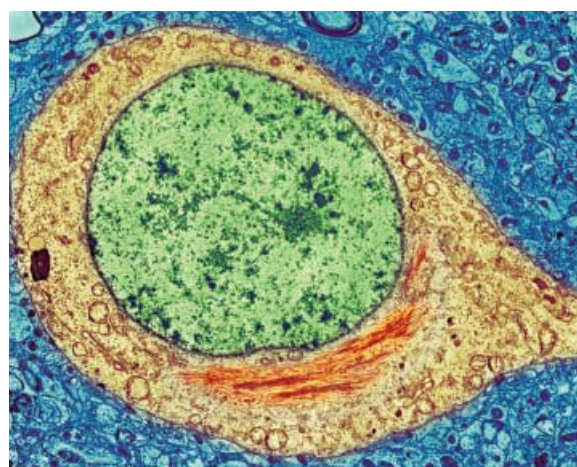
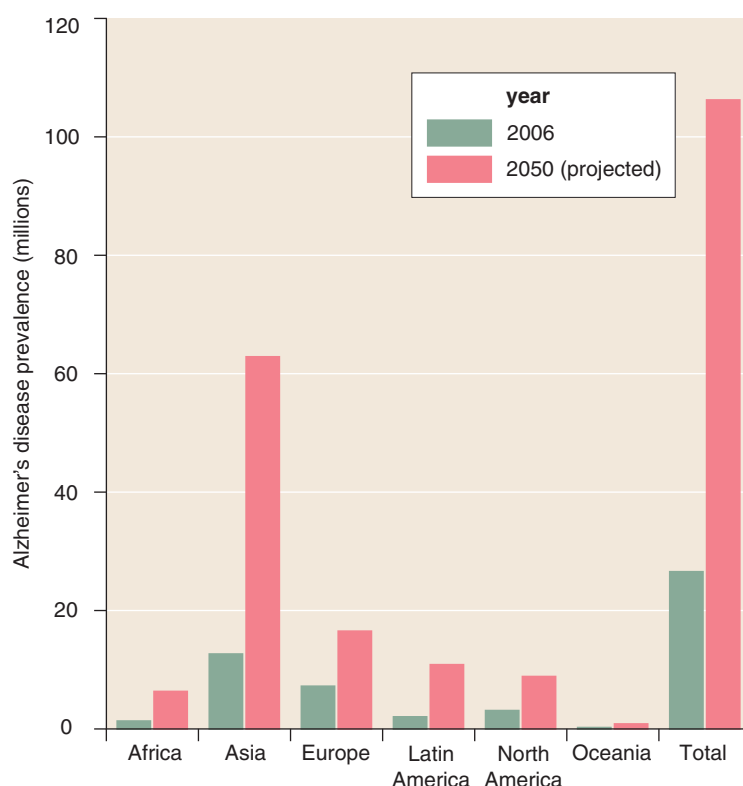


Gino Domenico/Associated Press

Figure 6. Dr. Konrad Maurer of the Johann Wolfgang Goethe University holds samples of the medical file of Auguste D. Rediscovered in Frankfurt in 1995, the file shed light on Alzheimer's observations about the first patient he diagnosed with what is now called Alzheimer's disease.

in Auguste D. Together with Perusini he sectioned the organs and searched for the telltale changes they had seen in Auguste D.'s brain. Once again they found abundant neurofibrillary tangles and amyloid plaques throughout the cerebral cortex. Perusini published the results of their analyses, including the first images illustrating the changes seen in Auguste D.'s brain, in 1909. They appeared in a scientific journal edited by Nissl and Alzheimer himself.

By that time, Kraepelin had started revising his very influential textbook on psychiatry for its eighth edition. In the chapter on senile and presenile dementias, Kraepelin decided to include Alzheimer's new findings. He began his description by noting that "a peculiar group of cases with severe cellular changes has been described by Alzheimer," and continued to relate, in some detail, the clinical symptoms Alzheimer had observed. Then he explained the histological abnormalities of the new disease: "The [plaques] were



Thomas Deerink, NCMIR/Photo Researchers

Figure 7. A century after Alzheimer's disease was discovered, the brain damage caused by the affliction can be imaged more precisely. Above is a colored transmission electron micrograph of a neurofibrillary tangle (red structure) in the cytoplasm of a neuron. Despite such progress, it's not yet clear how to prevent or cure this disease, whose incidence is expected to balloon as the world's population expands and people live longer. At left are worldwide projections for Alzheimer's prevalence developed by Ron Brookmeyer of the Johns Hopkins Bloomberg School of Public Health and colleagues. They were published in *Alzheimer's and Dementia* in 2007.

extraordinarily numerous, and nearly a third of the cortical cells [neurons] appeared to have died. In their place were strangely tangled, strongly staining bundles of fibrils, apparently the last remnants of the perished cell body." To illustrate these points, Kraepelin included figures showing these degenerative changes.

Kraepelin concluded his description by speculating about where in the range of known dementias this new disease might fit in: "The clinical interpretation of this Alzheimer's disease is currently unclear. While the anatomical findings seem to suggest that we are dealing with a particularly severe form of senile dementia, the fact that the disease occasionally already begins in the [patient's] late 40s seems to somewhat contradict this. One would have to presume a 'Senium praecox' [premature aging], if it is maybe not indeed a peculiar disease process, which is more or less independent of age." With these conjectures, Kraepelin appears to have foreseen that, apart from old age, other factors can cause the onset of Alzheimer's disease—genetic factors, for instance, as we know today. This endorsement by Kraepelin finally gave Alzheimer's findings recognition from the scientific community. Importantly, Kraepelin not only described the new disease, he also first used the term *Alzheimer's disease* in his textbook. With

this, Alzheimer's name would forever be associated with his discovery.

Alzheimer published the first comprehensive account of Auguste D.'s case in 1911. In this manuscript he also described another patient, Johann F., who had been admitted to the Munich clinic at the age of 56 with clinical symptoms very similar to those Alzheimer had observed in Auguste D. Interestingly, Johann F.'s brain differed from Auguste D.'s in one important aspect. While it displayed the typical amyloid plaques, there were no signs of changes in the neurofibrils. From today's point of view, Johann F. would be diagnosed with a less common form, the so-called "plaque-only" Alzheimer's disease. Thus, already at this early stage and after examining only a handful of patients, Alzheimer had a glimpse of the range of histopathological symptoms which remain associated with the disease today.

In his second publication on the disease, Alzheimer made clear that he accepted that the brain histology in Alzheimer's disease can vary between individuals. Moreover, he also began working toward describing a disease spectrum that, in addition to the early-onset (presenile) cases, included cases of senile dementia. Those cases had been observed by Alzheimer himself and by other scientists, such as Oskar

Fischer in Prague, and showed very similar histological changes.

Over two decades, Alzheimer poured most of his life into his medical and research pursuits. Working long hours caring for his patients and trying to uncover the causes of their mental afflictions, Alzheimer rarely took time off. During the early years in Munich, when Kraepelin didn't have a funded position for him, Alzheimer labored without a salary and paid substantial parts of the expenses associated with his research from his personal funds. In 1906, his devotion began to pay off. Kraepelin appointed Alzheimer a senior physician, and only three years later he was appointed assistant professor at the University of Munich. In 1910, he was selected as editor of a newly established psychiatric journal. At the same time, psychiatrists worldwide were increasingly recognizing his seminal contributions to neuropathology.

In 1912, the Silesian Friedrich-Wilhelm-University in Breslau offered Alzheimer the position of full professor and director of its Psychiatric and Neurological Clinic. After more than two decades working in the shadows of others, Alzheimer finally had the opportunity to put his own ideas into practice on an institutional level. The Breslau clinic had prestige. Alzheimer succeeded renowned scientists, such

as Heinrich Neumann, Carl Wernicke and most recently Karl Bonhoeffer who had just moved to the Charité Hospital in Berlin. Alzheimer accepted the offer and was appointed on July 16, 1912, certified by the signature of German emperor Wilhelm II himself.

Over time, though, the strain of work began to wear out Alzheimer. During his move to Breslau, he contracted a serious infection and after that endured breathlessness and heart trouble for the rest of his life. Despite failing health, Alzheimer strived to keep up. In addition to running the clinic, he continued publishing research articles and spent a considerable amount of time teaching. In the fall of 1913, he organized the annual meeting of the Society of German Psychiatrists in Breslau. With the outbreak of the First World War, however, psychiatric clinics faced the challenge of treating large numbers of new patients traumatized by the terror of war. For Alzheimer, already weakened by ill health, this came as a heavy blow. He worked hard to cope with a chaotic situation but ultimately it became too much. In October 1915 he was confined to bed and in December, at age 51, he died.

A Broader Legacy

Today Alzheimer is remembered almost exclusively for his discovery of the disease bearing his name. Clearly this was an epochal contribution to neurology, but he also made seminal contributions to understanding a number of other neurological disorders and diseases. He extensively studied other forms of dementia and produced important papers on cerebral atherosclerosis, epilepsy, and psychoses. He worked on brain damage resulting from chronic alcohol abuse and acute syphilis infections, which were very common at the time, and on forensic psychiatry. Besides these achievements, perhaps his most important influence was his instrumental contribution to introducing microscopy techniques into the discipline of psychiatry. That was an essential prerequisite for uncovering the cellular and molecular changes involved in mental disorders.

Over the years, Alzheimer's diagnoses have been questioned and reevaluated. In the 1990s, a team led by Manuel Graeber, then at the Max Planck Institute for Neurobiology near Munich, found about 250 slides of sections of Auguste D.'s brain in a basement at the University of Munich. They examined them

with today's deeper understanding of Alzheimer's disease. They saw a widespread, massive loss of neurons; numerous neurofibrillary tangles; and abundant amyloid plaques in the cerebral cortex, exactly as described by Alzheimer nearly a century earlier. Together with the clinical symptoms Alzheimer had described, these results confirmed that Auguste D. had the dreaded disease.

Given the early onset of the symptoms in Auguste D., it seems likely that she had a genetic predisposition for the disease. And there is even stronger evidence for a genetic contribution in the case of Johann F., whose brain tissue slides were also recovered. An analysis of his family's medical history revealed that several of his close relatives had also suffered from presenile dementia. These include his mother and maternal grandfather, a great-aunt and a great-grandfather, three of his eight siblings and five children of two of his affected siblings.

These observations suggested to scientists that in this case, Alzheimer's disease had a genetic basis. So did the fact that the illness often developed early (as early as a patient's thirties in some cases). Also, variability in the severity of dementia among people with the illness indicates that multiple genes may be involved and that environmental factors may influence those genes.

The scientists in Munich extracted DNA from the recovered brain tissue sections, hoping to identify the mutations that led to the disease in Johann F. and Auguste D. Unfortunately they did not detect any mutations that could have explained the disease. Due to the scarcity of DNA that can be purified from the original sections, the authors decided 10 years ago to postpone any

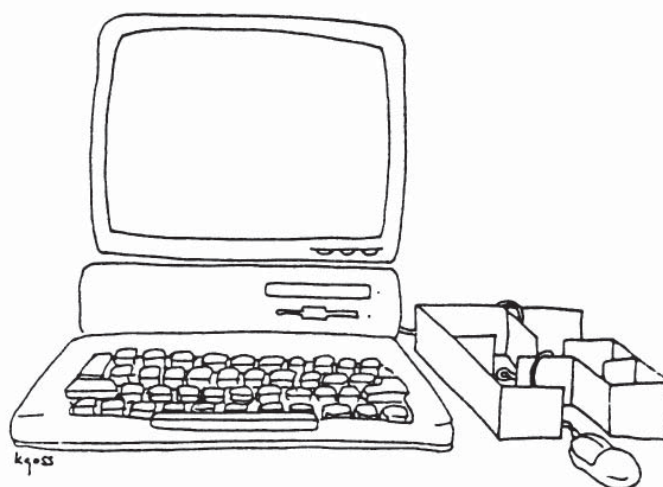
further analyses of Auguste D.'s and Johann F.'s DNA. With whole-genome amplification techniques and sequencing of entire genomes becoming routine, maybe the time has come to look at their molecular makeup again. As doctors and scientists prepare for the growth in Alzheimer's disease diagnoses expected in coming years, the first patients diagnosed with the disease may have more to teach us yet.

References

- Alzheimer, A. 1907 Über eine eigenartige Erkrankung der Hirnrinde. *Allgemeine Zeitschrift für Psychiatrie und psychiatrisch-gerichtliche Medizin* 64:146–148.
- Alzheimer, A., H. Forstl and R. Levy. 1991. On certain peculiar diseases of old age (translation). *History of Psychiatry* 2:71–101.
- Alzheimer, A. 1911. Über eigenartige Krankheitsfälle des späteren Alters. *Zeitschrift für die Gesamte Neurologie und Psychiatrie* 4:356–385.
- Dahm, R. 2006. Alzheimer's discovery. *Current Biology*. 16:906–910.
- Dahm, R. 2006. Alois Alzheimer and the beginnings of research into Alzheimer's disease. In *Alzheimer: 100 Years and Beyond*, eds. M. Jucker, K. Beyreuther, C. Haass, R. M. Nitsch and Y. Christen. Berlin and Heidelberg: Springer-Verlag, pp.37–49.
- Graeber, M., and P. Mehraein. 1999. Reanalysis of the first case of Alzheimer's disease. *European Archives of Psychiatry and Clinical Neuroscience*. 249:10–13.
- Maurer, K., and U. Maurer. 2003. *Alzheimer: The Life of a Physician and the Career of a Disease*. Translated by N. Levi with A. Burns. New York: Columbia University Press.

For relevant Web links, consult this issue of *American Scientist Online*:

<http://www.americanscientist.org/issues/id.83/past.aspx>



SIGHTINGS



Tracking the Karakoram Glaciers

In 2009, Italian outdoor photographer Fabiano Ventura began “On the Trails of the Glaciers,” a project duplicating early 20th-century expeditions to remote mountain glaciers. Ventura’s first stop was the Karakoram range, 16,500 square kilometers of glaciers and peaks that include K2, Earth’s second-highest mountain. Spanning parts of China, India and Pakistan, the Karakoram range is the most heavily glaciated area outside this planet’s polar regions. During a 1909 Italian expedition, photographer Vittorio Sella took numerous much-admired photographs there. Ventura’s new images were taken from the same point of view as the historical images to enable direct comparisons for research purposes. The photographs could help scientists observe developments in glaciers related to global climate change. American Scientist associate editor Catherine Clabby discussed the scientific value of the first stage of Ventura’s project with Kenneth Hewitt of Wilfrid Laurier University in Canada. An expert on Karakoram glaciers, Hewitt is a member of the project’s scientific committee.

A.S. Have glaciers in the Karakoram range been well studied?

K.H. The glaciers have been studied mainly in a few relatively extensive expeditions. The earliest were over 150 years ago and provide some basis for comparison. The 1909 Italian expedition brought a huge leap forward with mapping, observations along many of the high-altitude source areas of Baltoro glacier, and the outstanding photographs of Vittorio.

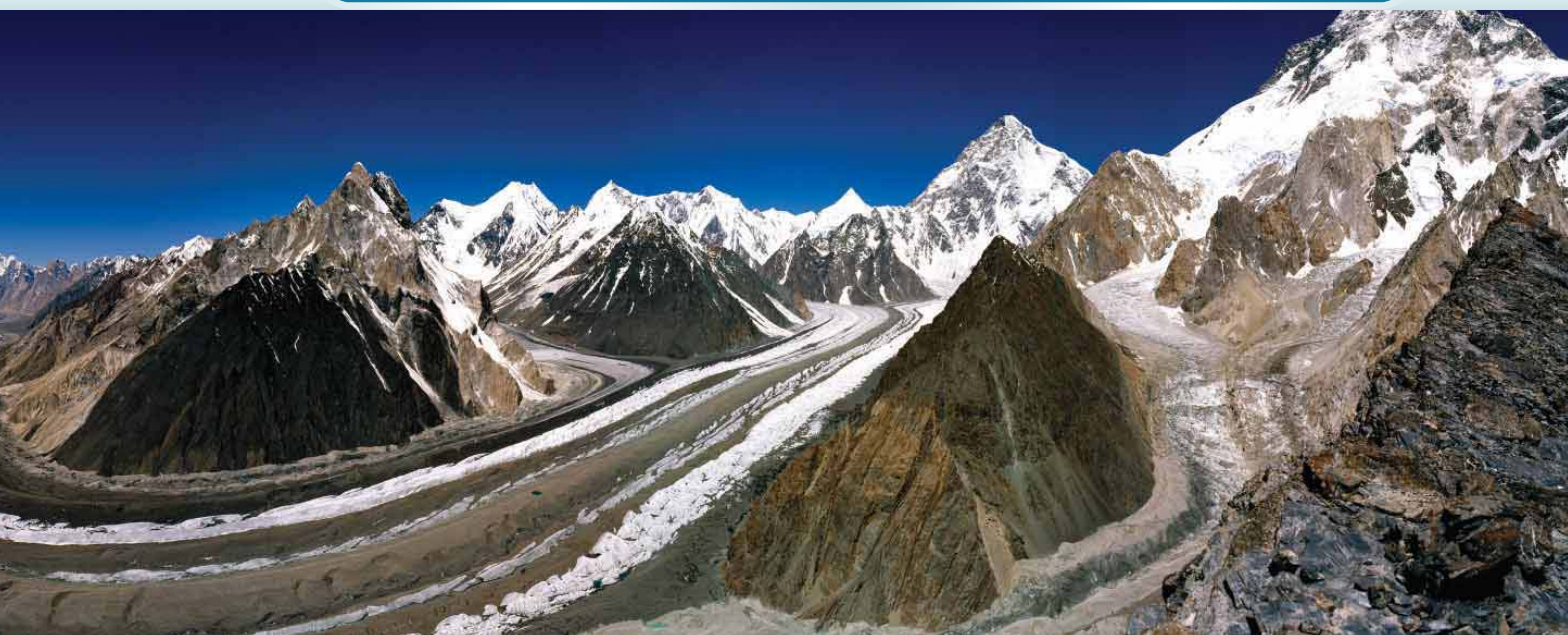
A.S. The Karakoram glaciers display changes since 1909 that you and other scientists associate with climate change. Yet the differences are not the dramatic shrinking and melting observed elsewhere. Can you explain that?

K.H. Glaciers are diminishing in most parts of the world, and reports of “disappearing glaciers” have come from much of the Himalaya region, but not the Karakoram. Total glacier cover there diminished by about 10 percent during the 20th century, but since the late 1960s there has been little change. Recently, many glaciers in the higher parts of the range have thickened and advanced. There have been exceptional numbers of glacier surges—sudden, rapid advances of some kilometers in a few months. The responses certainly reflect climate change, but they are regionally distinctive responses. The advances appear to relate to negative feedbacks in the glacier environment involving increased snowfall or reduced melting or both. The evidence is largely indirect or model-driven, but several research efforts using satellite imagery suggest both increased snowfall at higher elevations and more storminess and cloud cover in summer, which may re-

duce melting. If the apparent anomalies result from a warming globe, it suggests more moisture is transported from warmer oceans to the highest mountains. Once there, it either nourishes the glaciers with increased precipitation or protects them from some melting with more frequent cloudiness.

A.S. Do these glacier advances and surges pose dangers to people dwelling closest to the Karakoram?

K.H. Given the loss of glaciers elsewhere, sustained glacier mass in the Karakoram may seem good news, and in certain respects it is. But advancing glaciers also bring dangers. They brought hazards and disasters during the Little Ice Age, that period of global cooling that persisted in this region until the early 20th century. Hazards back then included glacial lake outburst floods that reached the heavily populated lowlands. Ice dams from advancing glaciers off the northern flanks of K2 remain a threat on the upper Yarkand River. Surges and terminus advances are confined to the higher parts of the range where they can block paths but rarely reach inhabited areas. The greater danger comes from ponds of water created during glacier advances or within stagnant ice after the surges end. Outburst floods from these ponds are especially destructive where they entrain sediment and become debris flows. Because the mountains straddle three countries, transnational as well as local and national issues arise. About one million people live along the upper Indus streams in the Karakoram and nearby ranges. Tens of millions live downstream in the Indus and Yarkand River lowlands, where snow- and ice- melt waters dominate river flows.



Fabiano Ventura 2009 / www.fabianoventura.it

Above is the gathering of the Baltoro glacier's great ice streams into a single tongue. The tongue extends beyond the center of this vista by more than 30 kilometers. Fabiano Ventura used film rather than digital cameras to create images such as this in the Karakoram range. That allowed him to match the magnification ratio obtained by Vittorio Sella in 1909 and to achieve greater resolution than would be possible using digital equipment. After scanning 4 x 5-inch photographic plates, Ventura produced digital files containing 320 million pixels, roughly 30 times the resolution of a consumer digital camera, for each image. Multiple images were used to produce panoramic views such as this one.

A.S. How are Fabiano Ventura's photographs of these glaciers useful to scientists?

K.H. Observations of changes in glaciers over a century or more are an invaluable first indicator of what has happened and what needs to be explained. We can see detailed variations in the margins and surface features of glaciers in these photographs. The Baltoro glacier is barely 300 meters shorter than in 1909, but the Biafo glacier is 3,500 meters shorter. The two glaciers are in the same part of the range, of similar size and length. However, Biafo is fed mainly by direct snowfall in the huge open basins at its head, while Baltoro is largely avalanche fed and has a much higher, more rugged watershed. There are indications that a higher watershed, avalanche nourishment and heavy debris cover produce a more

conservative response to climate change—and these effects apply to all glaciers advancing at present. High-resolution photography can also extend analysis of glacier changes to otherwise inaccessible areas. It supplements satellite imagery, some of which is amazing, but which is restricted to the recent past and not so good for observing vertical changes. The higher the resolution of images taken now, the greater the usefulness of photography for tracking future changes.

For more information about "On the Trails of the Glaciers," including updates regarding a documentary and a photography exhibit, visit: <http://www.sulletraccedeighiacciai.it>

In Sightings, American Scientist publishes examples of innovative scientific imaging from diverse research fields.



Vittorio Sella 1909 / © Fondazione Sella



Fabiano Ventura 2009 / www.fabianoventura.it

At left is a Vittorio Sella photograph of the terminus of the huge Biafo glacier in 1909. At right is a Fabiano Ventura photograph taken 100 years later. In 1909, the ice reached across the valley, and the Braldu River, carrying the waters of the Baltoro and Panmah glaciers, flowed in tunnels through the ice or was forced against the mountain wall. The contrast between past and present gives a good idea of dramatic changes in the glacier since the Little Ice Age, several centuries of cooling and glacier advances that lasted in the Karakoram until early in the 20th century.

SCIENTISTS' BOOKSHELF

Fellow Feeling

Joan B. Silk

THE AGE OF EMPATHY: Nature's Lessons for a Kinder Society. Frans de Waal. x + 291 pp. Harmony Books, 2009. \$25.99.

The thesis of Frans de Waal's new book, *The Age of Empathy*, is that empathy comes "naturally" to humans, by which he means that it is a biologically grounded capacity that all people share. According to de Waal, empathy has deep evolutionary roots, having originated before the order Primates came into existence. The antiquity of empathy firmly fixes its place in human nature, he believes, making it a robust trait that develops in all societies. De Waal makes an impassioned and eloquent case that understanding the role of empathy in nature can help us build a kinder and more compassionate society. His message will have considerable resonance for many readers.

De Waal has long been a critic of the notion that evolution drives us (and our primate relatives) to express the darker sides of our natures. He has been impatient with colleagues who are fixated on the struggle for existence and give short shrift to the need for cooperation and accommodation among interdependent animals that live in groups. Thus, while many primatologists have focused on evolutionary pressures that generate high levels of competition and conflicts within a group, de Waal has emphasized the importance of the mechanisms that primates use to defuse tension, resolve conflicts and repair the damage caused by them.

De Waal's argument in this book hinges on his claim that empathy is an ancient trait. Emphasizing the continuity in empathic concern across species, he speculates that empathy may be as old as maternal care itself. His reasoning is partly based on the selective advantages that he thinks empathy would have provided for mothers. Females who were sensitive to, and able to anticipate, the needs of their developing offspring would have been more successful mothers than those who were less responsive, he argues. But even if that's the case, it does not necessarily mean that mammals actually evolved the capacity for empathy. After all, it might also have been useful for mammalian males to have the capacity to lactate, because in some circumstances males who could provide nourishment for their young

might have had greater reproductive success than those who lacked this capacity. Nevertheless, except under rare specific conditions, mammalian males do not lactate.

Even though de Waal is firmly convinced that empathy is old and is widespread among mammals, not everyone agrees; there is a lively debate about these matters in the literature. Part of the controversy stems from the fact that the term *empathy* is used to describe a range of phenomena, from emotional contagion (in which one individual "catches" the emotions of another) to what Stephanie D. Preston and de Waal were the first to refer to as *cognitive empathy*—the ability to understand the feelings of others and to appreciate the distinction between their feelings and our own. Emotional contagion is a primitive form of true empathy, de Waal says; when one baby's cry sets off a chorus of cries from the other babies in the nursery, that's emotional contagion. Cognitive empathy is what allows us to understand the anguish of a mother whose child is diagnosed with a terminal illness.

The practical problem is that in any particular case it can be difficult to distinguish between emotional contagion and more elaborate forms of empathy. After all, how do we know what is actually going on in one baby's head when she hears another baby cry? Nevertheless, the distinction is crucial, because an understanding of others' needs is a prerequisite for the transformation of empathy into compassionate action. The contagion metaphor can be used to illustrate this point: If you catch a cold from your partner, you'll share your partner's symptoms. But feeling the same way as someone else is not the same thing as knowing how that person will want to be treated. To take care of your partner, you need to know whether he or she likes to be coddled when sick or prefers being left alone with a good book. If you have that information, you can be helpful even if you don't have a cold yourself.

This means that if we want to understand the capacity that other animals have for com-

passion, we have to figure out what is going on in their heads. Carefully designed experiments have given us some insight into what animals know about the minds of others. For example, Robert Seyfarth and Dorothy Cheney conducted an experiment in which female macaques learned that a box in their enclosure contained a frightening stimulus (a fake snake). Although the mothers were frightened when they came upon the snake and avoided the box afterward, they did not react when their infants approached the box, and they did not warn the infants of the danger the snake presumably represented. Based on these findings, Cheney and Seyfarth concluded that the mothers were unaware that their own knowledge differed from the knowledge of their offspring. The findings of a substantial body of cleverly designed experiments have resulted in a general consensus that monkeys have a less-well-developed understanding of others' minds than do apes.

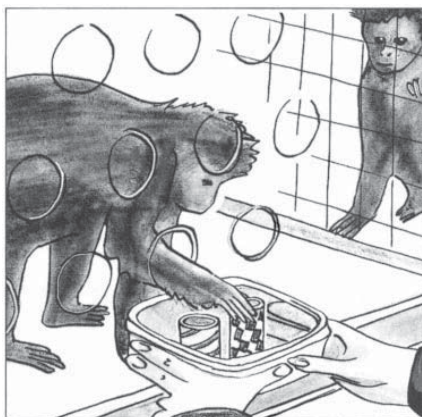
The ability of apes to understand others' minds might allow them to understand others' specific needs and to act compassionately. De Waal believes that apes do understand others' needs and that they act compassionately based on that understanding, a conclusion he bases in part on a number of one-time observations, several of which he describes here. For example, he recounts what happened when a female bonobo found a stunned bird in her enclosure. She carried it to the top of a tree, and then "she spread its wings as if it were a little airplane, and sent it out into the air, thus showing a helping action geared to the needs of a bird."

Although some scientists are dismissive of anecdotal accounts like this one, de Waal argues that they are valuable sources of information, particularly for events that are relatively uncommon in nature. I have no quarrel with this. Richard W. Byrne and Andrew Whiten's compilation of anecdotal observations of tactical deception in primates in the 1980s had a major impact on our understanding of primate cognitive complexity. I am more concerned about the way we make use of these one-time observations. De Waal argues that "If you have seen something yourself, and followed the entire dynamic, there is usually no doubt in your mind of what to make of it." But doubt is a healthy part of science. Doubt leads us to construct alternative hypotheses and to design experiments that will allow us to determine which hypotheses are correct.

Consider, for example, one of the best-known instances of animal altruism, which de Waal mentions in the endnotes for chapter 4. A young child tumbled into the gorilla enclosure at the Brookfield Zoo in Chicago and lay unconscious on the ground. A female gorilla named Binti Jua picked up the child, cradled him in her arms and brought him to the back of the enclosure, where anxious zoo staff were waiting. The event was videotaped by a visitor to the zoo, and Binti

ALSO REVIEWED IN THIS ISSUE

- 160 **PREDICTING THE UNPREDICTABLE: The Tumultuous Science of Earthquake Prediction.** By Susan Hough. Reviewed by Cosma Shalizi. *As recently as the 1970s, it seemed feasible that scientists would soon be able to say precisely when and where earthquakes would strike and what their impact would be, but most geologists now believe that that goal is almost certainly unattainable. Perhaps we should focus instead on organizing society so that when the earth shakes, it's not a catastrophe, says Shalizi*
- 162 **STEPHEN JAY GOULD: Reflections on His View of Life.** Edited by Warren D. Allmon, Patricia H. Kelley and Robert M. Ross. Reviewed by Kim Sterelny. *Because Stephen Jay Gould was ambivalent about or perhaps even hostile toward cladistics, population genetics and ecology, he was only partially connected to the mainstream of developing evolutionary thought, says Sterelny, who wishes these essays had more to say about the connections that Gould made or failed to make between his own ideas and the rest of his discipline*
- 164 **NURTURESHOCK: New Thinking about Children.** By Po Bronson and Ashley Merryman. Reviewed by Ethan Rettel. *Bronson and Merryman point to scientific findings that challenge some common assumptions about young people and parenting*
- 166 **BOYLE: Between God and Science.** By Michael Hunter. Reviewed by Jan Golinski. *Hunter places Boyle's scientific accomplishments in a context of lifelong piety and serious moral concerns, says Golinski. Dense with factual detail, the book covers every aspect of Boyle's life and work*
- 168 **MAPPING THE WORLD: Stories of Geography.** By Caroline and Martine Laffon. Reviewed by Brian Hayes. • **STRANGE MAPS: An Atlas of Cartographic Curiosities.** By Frank Jacobs. Reviewed by Anna Lena Phillips. *Quick glimpses into two new map books*
- 169 **SEASICK: Ocean Change and the Extinction of Life on Earth.** By Alanna Mitchell. Reviewed by Rick MacPherson. *Mitchell sets out on a personal voyage of discovery, accompanying top ocean scientists on expeditions that reveal the toll various assaults are taking on the global ocean*
- 170 **NOT BY DESIGN: Retiring Darwin's Watchmaker.** By John O. Reiss. Reviewed by John Dupré. *Reiss aims to reassert a thoroughgoing materialism and remove teleology from our vision of nature, says Dupré. Part of the problem, Reiss believes, is the gap that many biologists have assumed between existence and adaptedness*
- 172 **BIRDSCAPES: Birds in Our Imagination and Experience.** By Jeremy Mynott. • **THE BIRD: A Natural History of Who Birds Are, Where They Came From, and How They Live.** By Colin Tudge. Reviewed by Aaron French. *In addition to covering such topics as the behavior, morphology and conservation of birds, both of these books explore what birds mean to us and what we can learn from living with them*
- 174 **NANOVIEWS.** *Short takes on two books: Fordlandia: The Rise and Fall of Henry Ford's Forgotten Jungle City • Crow Planet: Essential Wisdom from the Urban Wilderness*



A capuchin monkey reaches through an armhole to choose between two differently marked pieces of pipe that can be exchanged for food. One of these tokens gets a reward only for the chooser, but the other token is prosocial—it “buys” food for both the chooser and the monkey who is looking on. Capuchins usually select the prosocial token. From *The Age of Empathy*.

Jua became famous. De Waal describes this as an “act of sympathy” prompted by Binti Jua’s concern for the welfare of the child. But there is more to the story. Binti Jua had been neglected by her own mother, and as a result she was hand-reared by humans. In an effort to improve the chance that she would be a better mother herself, her keepers gave her operant training with a doll; zoo staff rewarded her for holding the doll correctly and bringing it to them for inspection. All of this helped Binti Jua become a competent mother when she had her own infant. However, this piece of her history also raises the possibility that Binti Jua’s behavior during this incident reflected the training she had received rather than her sympathy for the child’s plight. I don’t know which interpretation is correct, but it is important to acknowledge that there are alternative explanations for Binti Jua’s behavior.

More systematic efforts to assess chimpanzees’ concern for the welfare of others have had mixed results. In some experimental settings, chimpanzees do provide appropriate instrumental help to their fellow chimpanzees, but in others they do not—for example, in one experiment chimpanzees failed to deliver food rewards to familiar group members even when they could have done so at no cost to themselves. De Waal endorses the experiments in which the chimpanzees were helpful and dismisses the others as examples of “false negatives.” As the author of one of the

sets of experiments in which chimpanzees failed to be helpful, I may not be entirely objective about the value of that work. However, I am convinced that if we really want to understand the nature of empathic concern and compassion in other apes, we need to figure out why chimpanzees respond helpfully in some circumstances and unhelpfully in others. De Waal himself made this argument in his 1996 book, *Good Natured: The Origins of Right and Wrong in Humans and Other Animals*, saying that “for a research program into animal empathy, it is not enough to review the highlights of succorant behavior; it is equally important to consider the absence of such behavior when it might have been expected.”

For de Waal, the debate about whether apes are motivated to help others matters because of its implications for humans. The continuities in empathy between humans and other creatures give him confidence about the prospects for creating a kinder human society:

I derive great optimism from empathy’s evolutionary antiquity. It makes it a robust trait that will develop in virtually every human being so that society can count on it and try to foster and grow it.

If empathy were limited to humans, de Waal says, that would mean that it was a trait that evolved only recently. This concerns him: “If empathy were truly like a toupee put on our head yesterday, my greatest fear would be that it might blow off tomorrow.”

But the existence of emotional contagion in rodents and cognitive empathy in apes is neither a necessary nor a sufficient condition for there to be cogni-

tive empathy and compassion in humans. Important traits are transformed over the course of evolutionary time. Monkeys and apes have lost function in about one-third of their olfactory receptor genes, greatly reducing the sensitivity of their sense of smell; the apes have lost their tails; brachiating gibbons have greatly shortened thumbs in their hook-like hands; and we humans have lost the ability to grasp things with our feet. At the same time, fundamental traits such as bipedal locomotion, spoken language and cumulative cultural change were all greatly elaborated after the human lineage diverged from the lineage of modern apes. All of these traits, despite their relatively recent origins, have left an indelible mark on our species.

Here de Waal misses the opportunity to explore what makes us different from other apes. We cooperate in larger groups, solve collective action problems, adhere to social norms and possess moral sentiments. Whether or not we inherited the capacity for empathy from our primate ancestors, we have developed these capacities much further than other apes have. Recently, a number of scholars have given a great deal of thought to how and why human societies have become more cooperative than the societies of other primates, but de Waal does not discuss their ideas here. That is a pity, because we need to know the answers to those questions if we want to create kinder societies.

Joan B. Silk is professor and chair of the department of anthropology at the University of California, Los Angeles. She is coauthor with Robert Boyd of *How Humans Evolved*, which is now in its fifth edition (W. W. Norton, 2009).

GEOLOGY

Ready or Not

Cosma Shalizi

PREDICTING THE UNPREDICTABLE: The Tumultuous Science of Earthquake Prediction. Susan Hough. viii + 261 pp. Princeton University Press, 2010. \$24.95.

Earthquake prediction is, in an important sense, a solved problem. Earthquakes are vastly more common in certain parts of the world than others, and they occur at a reasonably steady statistical frequency in a given location. We even know why this is so. Earthquakes are most frequent

in those parts of the world where the tectonic plates run up against each other and try to move past each other. Where the plates meet, we get fault lines. When the material on one side of the fault sticks to that on the other, strain builds up and gets released in sudden movements: earthquakes. So the

baseline prediction is that earthquakes occur near faults, with frequencies about equal to their historical frequencies, because the mechanics of tension and relaxation change very slowly.

This lets us say things like “Once in about every 140 years, the Hayward fault in northern California has a quake of magnitude 7.0 or greater.” But some people, including some seismologists, are not content with this level of understanding and these actuarial “forecasts”; they want to be able make highly accurate *predictions*—to be able to say precisely when and where an earthquake will occur and what its impact will be (“magnitude 7.1, directly beneath the stadium at the University of California, Berkeley, the day after the Big Game with Stanford in 2010”). As recently as the 1970s, this goal seemed feasible to professionals and the U.S. government, but now most geologists believe that it is extremely unlikely ever to be accomplished. In *Predicting the Unpredictable*, Susan Hough tries to explain both the initial enthusiasm for precise predictions and how and why that enthusiasm dissipated.

The enthusiasm came at the end of the plate tectonics revolution, which gave us our current understanding of, among many other things, earthquakes. After millennia of speculation and superstition, we finally knew why the earth shakes and why earthquakes happen where they do. It really didn’t seem too much to hope that this triumph of science would soon extend to knowing *when* they would happen. Moreover, the authorities in the People’s Republic of China had apparently been able to predict the magnitude 7.3 earthquake that occurred in Haicheng in northwest China in 1975. (The real story of the Haicheng prediction, as Hough explains in chapter 6, is far murkier; as one of her sources puts it, “the prediction . . . was a blend of confusion, empirical analysis, intuitive judgment, and good luck.” But the details were deliberately kept from the rest of the world for many years.) Eminent geologists saw earthquake prediction as a reasonable scientific aim, and by the end of the 1970s, they managed to get it inscribed into U.S. policy, along with hazard reduction. They also established an official body for evaluating earthquake predictions.

Chapters 9 through 13 are mostly about various prediction efforts since that time, ranging from the serious to the crackpot. None of these efforts has

been really successful, although Hough is careful to say that some of them are only ambiguously failures. Evaluating the success of the predictors is harder than it first seems, because earthquakes are not just concentrated around plate boundaries at characteristic, though irregular, intervals; they are also clustered in space and especially in time. Earthquakes tend to happen near where other earthquakes have happened recently.

This clustering invalidates what has been a common method of evaluating earthquake predictions, which is to assess how well the predictions match the actual record of quakes and then compare that with how well the *same* predictions match a simulated record in which earthquakes occur at random on each fault at the historical rate (technically, according to a homogeneous Poisson process). Matching the real data better than the simulated data is supposed to be evidence of predictive ability.

To see the flaw here, think of trying to predict where and when lightning will strike. We know that lightning strikes, like earthquakes, are clustered in both space and time, because they occur during thunderstorms. So a basic prediction rule might state that “within 10 minutes of the last lightning strike, there will be another strike within 5 kilometers of it.” If we used this rule to make predictions and then were evaluated by the method described in the preceding paragraph, we would look like wizards. If we made predictions only after the lightning had already begun, we’d look even better. This is not just an idle analogy; the statisticians Brad Luen and P. B. Stark have recently shown that, according to such tests, the following rule seems to have astonishing predictive power: “When an earthquake of magnitude 5.5 or greater occurs anywhere in the world, predict that an earthquake at least as large will occur within 21 days and within an epicentral distance of 50 km.” Earthquake prediction schemes that do no better than this baseline predictor have little value, they observe. And no prediction method yet devised does any better than that.

Of course it’s possible that there is some good way of making detailed predictions, which we just haven’t found yet. To continue the lightning-strike analogy, we’ve learned a lot about how thunderstorms form and move; we can track them and extrapolate where they will go. Perhaps earthquakes are preceded by similar signals and patterns



In the wall to the right of this archway in Memorial Stadium at the University of California, Berkeley, is an open crack caused by steady creep in the Hayward Fault, which runs directly beneath the stadium. From *Predicting the Unpredictable*.

that are, as the saying goes, patiently waiting for our wits to grow sharper. But it’s equally possible that any predictive pattern specific enough to be useful would involve so many high-precision measurements of so much of the Earth’s crust that it could never be used in practice.

Suppose, however, that that’s not true; suppose we are someday able to make predictions like the one above about Berkeley’s stadium. We could perhaps evacuate Berkeley and its environs, but every building, power line and sewer pipe there would still go through the quake. It would be a major catastrophe if they all went to pieces, even if no loss of life occurred. If we insist on living in places like Berkeley, where we *know* there will continue to be earthquakes, why not work on hazard reduction—on building cities that can survive quakes and protect us during them—rather than on quake prediction? As Hough puts it:

If earthquake science could perfect the art of forecasts on a fifty-year scale, we would know what structures and infrastructure would be up against. For the purposes of

building a resilient society, earthquake prediction is largely beside the point. Whether the next Big One strikes next Tuesday at 4:00 p.m. or fifty years from now, the houses we live in, the buildings we work in, the freeways we drive on—all of these will be safe when the earth starts to shake, or they won't be.

One might almost say that the real problem isn't predicting when the

earth will shake, it's organizing society so that it's not a catastrophe when that happens.

In the end, whether through hope, caution or diplomacy, Hough declines to dismiss the prospect of prediction altogether. The current state of a lot of the science she reports on is frustratingly inconclusive. Hough's book, however, is not frustrating at all; it offers an enlightening, fair and insightful look at how one science has dealt with the

intersection of an extremely hard problem with legitimate public demands for results. Those of us in other fields who read it may find ourselves profiting from the example someday.

Cosma Shalizi is an assistant professor in the statistics department at Carnegie Mellon University and an external professor at the Santa Fe Institute. He is writing a book on the statistical analysis of complex systems models. His blog, Three-Toed Sloth, can be found at <http://bactra.org/weblog/>.

BIOLOGY

Explicating Gould

Kim Sterelny

STEPHEN JAY GOULD: Reflections on His View of Life. Warren D. Allmon, Patricia H. Kelley and Robert M. Ross, editors. xiv + 400 pp. Oxford University Press, 2009. \$34.95.

Stephen Jay Gould was an immensely charismatic, insightful and influential, but ultimately ambiguous, figure in American academic life. To Americans outside the life sciences proper, he *was* evolutionary biology. His wonderful essay collections articulated a vision of that discipline—its history, its importance and also its limits. One of the traits that made Gould so appealing to many in the humanities and social sciences is

that he claimed neither too much nor too little for his discipline. In his books, evolutionary biology speaks to great issues concerning the universe and our place in it, but not so loudly as to drown out other voices. He had none of the apparently imperialist ambitions of that talented and equally passionate spokesman of biology Edward O. Wilson. It is no coincidence that the humanist intelligentsia have given a much friendlier reception to Gould than

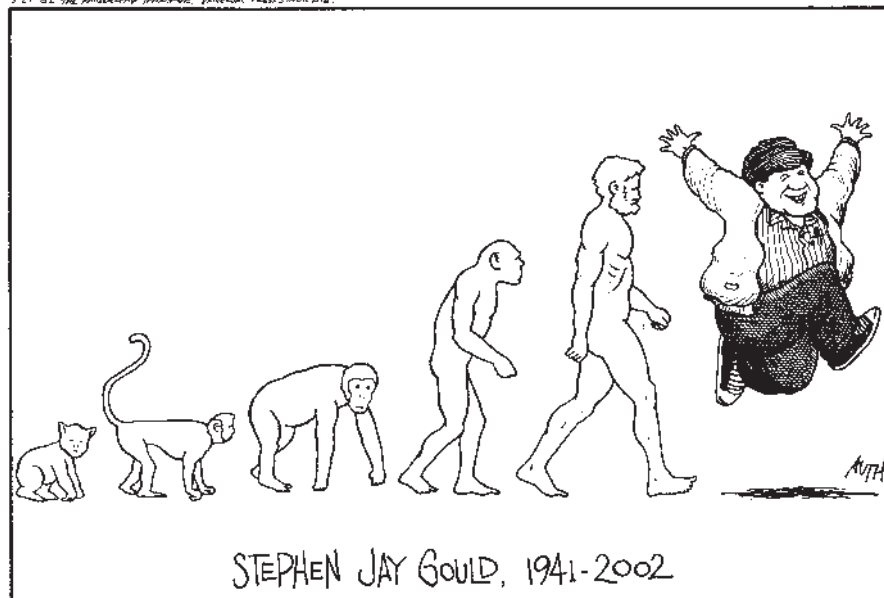
to Wilson. Gould's work is appealing to philosophers like me because it trades in big, but difficult and theoretically contested, ideas: the role of accident and the contingency of history; the relation between large-scale pattern and local process in the history of life; the role of social forces in the life of science.

Within the life sciences, Gould is regarded with more ambivalence. He gets credit (with others) for having made paleobiology again central to evolutionary biology. He did so by challenging theorists with patterns in the historical record that were at first appearance puzzling; if received views of evolutionary mechanism were correct, Gould argued, those patterns should not be there. The first and most famous such challenge grew from his work with Niles Eldredge on punctuated equilibrium, but there were more to come.

Despite this important legacy, Gould's own place in the history of evolutionary biology is not secure. In late 2009, I attended an important celebration of Darwin's legacy at the University of Chicago, in which participants reviewed the current state of evolutionary biology and anticipated its future. Gould and his agenda were almost invisible. No doubt this was in part an accident of the choice of speakers. But it is in part a consequence of Gould's ambivalence regarding, or perhaps even hostility toward, core growth points in biology: cladistics, population genetics, ecology.

Gould was an early force in one of the major recent developments in biology: the growth of evolutionary developmental biology, and the idea that the variation on which selection works is channeled by deeply conserved and widely shared developmental mechanisms. Gould's first book, *Ontogeny and Phylogeny* (1977), was about the antecedents of this movement, and in his essays and monographs he regularly

5.17 — 2. THE ASCENDING STAIRCASE. HUMANOIDAL PLEISTOCENE



This cartoon by Tony Auth is reproduced in *Stephen Jay Gould: Reflections on His View of Life*, where it is captioned "The (punctuated) Ascent of Stephen Jay Gould, or Portrait of the Evolutionist as a Provocateur."

returned to this developing set of ideas. He did so to articulate his vision of natural selection as an important but constrained force in evolution. But in the past 20 years, evolutionary biology has been transformed in other ways too. Perhaps the most important is that cladistics—systematic, methodologically self-conscious, formally sophisticated phylogenetic inference—has become the dominant method of classification. This phylogenetic inference engine has made it possible to identify trees of life with much greater reliability and to test adaptationist hypotheses and their rivals far more rigorously. Even though Gould had been early to see the problems of impressionistic adaptationist theorizing, his own work responded very little to these changes in evolutionary biology.

Likewise, Gould showed very little interest in the evolving state of population genetics; his last book, *The Structure of Evolutionary Theory*, barely mentions it (W. D. Hamilton, for example, is not even in the index). This is surprising, because one recent development in population genetics had been the growth of multilevel models of selection. In *The Structure of Evolutionary Theory*, Gould does nod to these models, but he does little to connect them to his own ideas on hierarchical models of selection, which get very little formal development of any kind, let alone the kind of development that would connect them to the extending mainstream of evolutionary theory. Finally, Gould showed extraordinarily little interest in ecology and the processes that link population-level events to patterns in the history of life.

In short, although Gould was clearly an immensely fertile thinker whose ideas were deeply informed both by contemporary paleobiology and by the history of biology, in other respects his work is only partially connected to the mainstream of developing evolutionary thought. There is, therefore, room for a work that explores and reflects on the explicit connections that Gould made between his own ideas and the rest of his discipline, and that makes some implicit connections more explicit. *Stephen Jay Gould: Reflections on His View of Life*, edited by Warren D. Allmon, Patricia H. Kelley and Robert M. Ross, is an interesting collection of essays, but it does not quite do that, in part because it is an insider's perspective and in part because some of the essays are of only

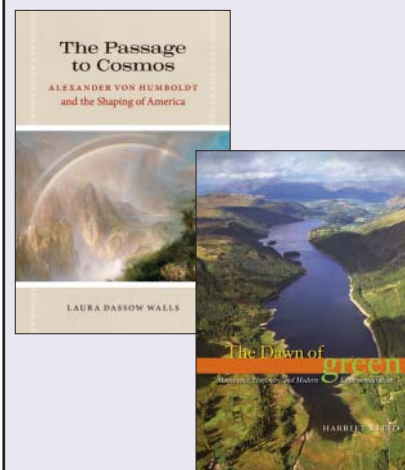
local significance. For example, the chapters on Gould's status as an educator, his role as an iconic left-liberal American intellectual, and his relations with those of his students who were religious will likely be of interest to no one outside the U.S. milieu and I would guess to very few people within it. The collection is not a Festschrift, but it does show an occasional tendency to decay in that direction.

One of the strongest chapters in the collection, to my mind, is "A Tree Grows in Queens: Stephen Jay Gould and Ecology," by Allmon, Paul D. Morris and Linda C. Ivany. Allmon, Morris and Ivany try to explain the lack of an ecological footprint in Gould's work. In their view, this is best explained by his skepticism about natural selection. It is true that the more strongly one believes that the tree of life is basically shaped by mass extinction (while thinking that the extinction in mass extinction does not depend on adaptation), the less important ecology is. The theory of punctuated equilibrium, too, plays down the importance of ecology through most of the life of a species. Although I am sure this must be part of the story, something is missing. Gould remained committed to the truth and importance of the punctuated equilibrium model of the life history of the typical species. He developed that model in partnership with Eldredge and (later) Elizabeth Vrba. But ecological disturbance remains central to Eldredge's and Vrba's conception of punctuated equilibrium and, more generally, to evolutionary change. Equilibrium is not forever. So for example, in contrast to Gould, Eldredge has written extensively on ecological organization and its relation to evolutionary hierarchy. Allmon, Morris and Ivany have put their finger on a crucial problem in interpreting Gould, but the problem remains unsolved.

Allmon is also the author of the fine chapter that opens the collection—"The Structure of Gould: Haptenstance, Humanism, History, and the Unity of His View of Life." Long and insightful, this essay is one of interpretation rather than assessment. It explores the relation between content and form in Gould's work. The current norm of science is that research consists in the publication of peer-reviewed papers in specialist journals. As one nears retirement, it may be acceptable to switch to writing reflec-

THE NATURAL WORLD

FROM CHICAGO



The Passage to Cosmos

Alexander von Humboldt and the Shaping of America

Laura Dassow Walls

"Laura Dassow Walls leads the reader on a fascinating, breathless chase after the explorer-naturalist who anticipated planetary ecology and inspired both Darwin and Thoreau. Alexander von Humboldt was a pioneer environmentalist whose sympathies crossed nations, races, and cultures; his friendships included Jefferson and Goethe, Simón Bolívar, Moses Mendelssohn, and John C. Frémont."—Daniel Walker Howe, Pulitzer Prize-winning author of *What Hath God Wrought*

CLOTH \$35.00

The Dawn of Green

Manchester, Thirlmere, and Modern Environmentalism

Harriet Ritvo

"The tensions between the needs of cities and the protection of rural landscape were defined in the late Victorian struggle between the Manchester City Council and defenders of the English Lake District. . . . Harriet Ritvo has beautifully analyzed this classic confrontation, setting it in context, but also showing its importance for continuing debates around the use of the earth."—John V. Pickstone, University of Manchester

CLOTH \$26.00

The University of Chicago Press

www.press.uchicago.edu

tive what-does-it-all-mean review papers, and even a book or two: This is known as going through philosopause. Gould went through philosopause early, and Allmon attempts to explain why, connecting the form of Gould's work as an essayist and book author with his humanism, his liberalism and his interest in exploring murky, large-scale questions. He broke with conventional norms of science writing not just because he wanted to reach more people, but because of what he wanted to say. The essay is interesting, but I think Allmon lets Gould off too lightly, especially in his discussion of the supposed early misreading of punctuated equilibrium. Gould's early rhetoric on the revolutionary impact of that idea and his repeated flirtations with Richard Goldschmidt's metaphors made it easy to read Gould as rejecting standard neo-Darwinian gradualism; at the time, I read him that way myself. And the discipline of peer review would certainly have improved *The Structure of Evolutionary Theory*.

Richard Bambach's essay, "Diversity in the Fossil Record and Stephen Jay Gould's Evolving View of the History of Life," is also rewarding, although it is less ambitious than Allmon's. Bambach offers a chronological overview of the most important feature of Gould's purely scientific ideas: his emerging view of the basic patterns of life's history. It is good to have this material assembled and presented so coherently. I would, however, have liked to see a rather franker assessment of these ideas. In *Wonderful Life: The Burgess Shale and the Nature of History* (1989), Gould argues that if "the tape of life" were replayed from very slightly different initial conditions, the resulting tree of life would probably in no way resemble our actual biota. This was surely one of the most provocative of his ideas, and Simon Conway Morris replied at length in *Life's Solution*, reaching utterly the opposite conclusion. At the end of the chapter, Bambach touches on this debate but says almost nothing to assess it.

In general, the other essays in this book have the same virtues as Bambach's chapter. The editors, I suspect, think that Gould has been much misread and misunderstood, so most of the essays seek to state his views simply and without distracting polemic. As a result the collection is stronger on description than evaluation. I am

not convinced that Gould has been so much misunderstood, and I would have preferred more assessment and less exposition. That said, I did enjoy reading the book.

Kim Sterelny divides his time between Victoria University of Wellington, where he holds a Personal Chair in Philosophy, and the Research School of Social Sciences at Australian National University.

DEVELOPMENTAL PSYCHOLOGY

The Science of Parenting

Ethan Rimmel

NURTURESHOCK: New Thinking about Children. Po Bronson and Ashley Merryman. xiv + 336 pp. Twelve, 2009. \$24.99.

Does praise undermine a child's confidence? Can gifted children be reliably identified in preschool? Why do siblings fight, and how can they be discouraged from doing so? Are popular children more aggressive? Do videos like those in the *Baby Einstein* series help infants learn language?

NurtureShock: New Thinking about Children addresses such questions, examining how recent research in developmental psychology challenges conventional wisdom about parenting and schooling. Aimed at laypeople rather than academics, the book made the *New York Times* nonfiction bestseller list last year and was listed as one of the year's best by Barnes and Noble, *Discover Magazine*, *Library Journal* and others.

The authors, Po Bronson and Ashley Merryman, are not researchers themselves. Bronson has written several books on other topics, including the bestselling *What Should I Do with My Life?*, about career choices. Together, Bronson and Merryman have written about parenting and social science in online columns for *Time* and *Newsweek* and in articles for *New York* magazine. Three chapters in *NurtureShock* are adapted from their *New York* articles.

The title evokes Alvin Toffler's 1970 book *Future Shock*. But Bronson and Merryman explain in the introduction that they are using the term *nurture shock* to refer to "the panic—common among new parents—that the mythical fountain of knowledge is not magically kicking in." And they warn that the

information in the book will deliver a shock, by revealing that "our bedrock assumptions about kids can no longer be counted on." Somewhat confusingly, the authors also assert that what the subtitle calls "new thinking about children" is actually a "restoration of common sense."

Each of the 10 chapters focuses on a different topic: praise, sleep, racial attitudes, lying, intelligence testing, sibling conflict, teen rebellion, self-control, aggression and language development. Bronson and Merryman did their homework, talking to many researchers and attending academic conferences. The book's endnotes include citations for many of the empirical statements in the text, and the list of selected sources and references is extensive. The coverage is somewhat skewed toward the work of the researchers who were interviewed, but Bronson and Merryman talked to leading experts on every topic.

In some places additional information could have been helpful. For example, in the chapter on self-control, the authors focus on a preschool program called "Tools of the Mind," which successfully teaches self-regulation. However, they don't explain the theoretical work that inspired the program, that of the Russian psychologist Lev Vygotsky. Nor do the authors mention research by Angela Duckworth and Martin Seligman showing that self-discipline predicts academic achievement better than IQ does. As far as I know, though, nowhere in the book have they neglected evidence that would undermine their arguments.

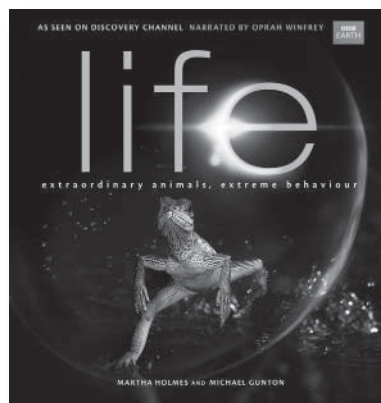
Bronson and Merryman make child development research accessible and even exciting; *NurtureShock* is an easy and enjoyable read. By academic standards, the writing style may be a bit melodramatic in some places, but I would recommend the book to any parent. All of the advice has empirical support, and readers will almost certainly emerge thinking differently about some aspect of parenting. Some sections do seem geared toward American parents of middle to high socioeconomic status. For example, cognitive testing for competitive admission to prestigious private preschools is an issue in only a few urban areas of the United States; it's unheard of elsewhere.

As a developmental psychologist, I appreciate the attention that Bronson and Merryman are attracting to the field. At the risk of nitpicking, however, they do get some things wrong. For example, they describe brain development as a process in which "gray matter gets upgraded to white matter." This metaphor is not quite correct. White matter is added as nerve axons are covered in whitish myelin, but it does not replace gray matter, which is made up of nerve cell bodies and dendrites. In the chapter about racial attitudes, the authors describe a 2006 study by Meagan Patterson and Rebecca Bigler in which preschool children were randomly assigned to wear either red or blue T-shirts in their classrooms for three weeks. Bronson and Merryman write that "the teachers never mentioned their colors and never again grouped the kids by shirt color." In fact, that was only true of classrooms in the control condition. In the article reporting their findings, Patterson and Bigler state that

Teachers in the experimental classrooms made frequent use of the color groups to label children (e.g., "Good morning, Blues and Reds") and to organize the classrooms. For example, teachers in the experimental classrooms decorated children's cubbies with blue and red labels and lined up children at the door by color group.

Although children in both the experimental and control conditions developed some bias for their own group, children in the experimental group showed greater in-group bias. So Bronson and Merryman are correct that young children can form

AMAZING NATURE



Life

EXTRAORDINARY ANIMALS,
EXTREME BEHAVIOUR

Martha Holmes and Michael Gunton

Life, the spectacular companion volume to the new Discovery Channel/BBC series, tells a compelling story of survival and of the amazing behaviors animals and plants adopt to stay alive and pass their genes to a new generation.

\$39.95 hardcover

Adventures among Ants

A GLOBAL SAFARI WITH
A CAST OF TRILLIONS

Mark W. Moffett

"Certainly *Adventures among Ants*, with its detailed account of his fieldwork, makes an invaluable contribution to our scientific knowledge of these creatures. But it does more. It is so well written and captures his excitement so wonderfully."

—Jane Goodall, PhD, DBE,
Founder, the Jane Goodall Institute, and
UN Messenger of Peace

\$29.95 hardcover

The Atlas of Global Conservation

CHANGES, CHALLENGES, AND
OPPORTUNITIES TO MAKE A DIFFERENCE

Jonathan Hoekstra, Jennifer L. Molnar, Michael Jennings, Carmen Revenga, Mark D. Spalding, Timothy M. Boucher, James C. Robertson, and Thomas J. Heibel, with Katherine Ellison

"It's exciting to see so much real information about nature and its fate presented so beautifully and accessibly. The *Atlas*' fact-backed case for urgent action builds steadily to a compelling conclusion. And the writing is every bit as good as the excellent graphics and photographs."

—James Gustave Speth,
author of *Red Sky at Morning*

Published in cooperation with The Nature Conservancy, \$49.95 hardcover

At bookstores or www.ucpress.edu

The Encyclopedia of Weather and Climate Change

A COMPLETE VISUAL GUIDE

Juliane L. Fry, Hans-F. Graf, Richard Grotjahn, Marilyn N. Raphael, Clive Saunders, and Richard Whitaker

This comprehensive, and up-to-date volume covers all aspects of the world's weather. Liberally illustrated with more than 2,000 color photos, maps, diagrams, and other images, *The Encyclopedia of Weather and Climate Change* takes the reader beyond simple definitions.

\$39.95 hardcover

Leopold's Shack and Ricketts's Lab

THE EMERGENCE OF ENVIRONMENTALISM

Michael J. Lannoo

"Brings fresh insight to the fertile ideas and writings of two innovators of early twentieth century ecology. In this insightful and important book, Lannoo enriches the legacies of Leopold and Ricketts as early conservation-minded environmentalists and suggests that there is still much to be learned from them."

—Katharine A. Rodger, editor of
Breaking Through

\$24.95 hardcover



UNIVERSITY OF CALIFORNIA PRESS

prejudices without adult labeling, but their text gives the impression that adult labeling was not a factor, whereas this study and others by Bigler and colleagues demonstrate that labeling does matter.

More worrisome are some signs that the authors misunderstand statistics. For instance, they convert all correlation coefficients to percentages, an error that will annoy readers who are knowledgeable about statistics and could potentially mislead those who aren't. A naive reader, seeing a correlation expressed as 40 percent, may focus on how far that number falls short of 100 percent and fail to recog-

nize that even a correlation of 0.40 has some predictive value.

Bronson and Merryman also don't seem to understand effect sizes. For example, they write that

Among scholars, interventions considered to be really great often have an effect size of something like 15%, which means that 15% of children altered their targeted behavior, and therefore 85% did not alter it.

But that's not what it means. It could mean that all the children altered their behavior a little bit, which moved their average by 15 percent of a standard deviation.

The few things that Bronson and Merryman get wrong, however, are far outweighed by the things they get right. They have done a service to developmental science by making its findings accessible to a wider audience, and to parents by providing insight into children as well as practical suggestions for child rearing. For those achievements, the book deserves the accolades it is receiving.

Ethan Remmel is a cognitive developmental psychologist at Western Washington University in Bellingham. His research focus is the relationship between language experience and children's understanding of the mind.

HISTORY

The Godly Scientist

Jan Golinski

BOYLE: Between God and Science. Michael Hunter. xiv + 366 pp. Yale University Press, 2009. \$55.

A short entry in a single-volume encyclopedia will tell you the achievements for which Robert Boyle (1627–1691) is most commonly remembered. He discovered air pressure and formulated Boyle's law, which shows that the pressure and volume of a gas are inversely related to one another. He studied the workings of the barometer and designed an air pump to investigate the effects of a vacuum. He was a founding member of a group organized in 1660 to encourage and communicate scientific research; in 1662 it became the Royal Society of London. His most famous book, *The Sceptical Chymist* (1661), challenged the prevailing theories about chemical composition held by the alchemists of the time and by the followers of Aristotle's natural philosophy.

Michael Hunter's new biography will add greatly to readers' knowledge of Boyle and may correct some misimpressions. It turns out, for example, that although Boyle developed the concept of what he called the "spring of the air," he never wrote down his law in the algebraic form that is now familiar: $pV = k$, where p is the pressure of the system, V is the volume of the gas, and k is a constant. Much of the hands-on work with the air pump—and some

of the crucial interpretation—was performed by Robert Hooke and others, whom Boyle employed as his assistants. And although Boyle was present at the first meeting of what became the Royal Society and had previously associated with some of its members in Oxford in the late 1650s, he did not attend with consistent regularity in the following years. He was more important as an inspiration for the leading ideas and values of the society than as an institutional organizer, which was a role he never assumed. And, despite his criticisms of alchemy in *The Sceptical Chymist*, Boyle was never disillusioned with the subject. In fact, it was the first experimental field to draw his attention, and he never ceased to be fascinated by the alchemists' vision of transmuting base metals into gold and producing wonderful new medicines.

Hunter's goal is to explain the complexities surrounding Boyle's life and work, and thereby to tell the story of how he became the person he was. In the first paragraph of the book, Hunter notes that Boyle was the most eminent "scientist" of his day but explains that this word was not one Boyle could have applied to himself, because it was not coined until the 19th century. Boyle called himself a "naturalist," an "experi-

mental philosopher" and a "Christian virtuoso." Three elements—experimental science, moral philosophy and Christian devotion—were essential to his self-formation. Hunter gives all of them their due weight, placing Boyle's scientific accomplishments in a context of lifelong piety and serious moral concerns.

These aspects of Boyle's identity have emerged more clearly in recent years as the result of a huge scholarly enterprise in which Hunter has been the driving force. Previously unpublished writings of Boyle have been rescued from the archives and put into print, multivolume editions of his works and correspondence have been compiled, and an extensive Web site (<http://www.bbk.ac.uk/boyle/>) reports the progress of Boyle studies. When the whole body of Boyle's writings and the whole documentary record surrounding him is taken into account, it becomes impossible to fit him into the role of scientist as this would now be understood.

Consider, for example, his enthusiasm for what in the 17th century was called "chymistry." Boyle first glimpsed the potential benefits to be derived from chemical investigations in the late 1640s, after he had completed his formal schooling and settled in Dorset on an estate acquired by his father, the Earl of Cork. Boyle arranged for chemical furnaces and vessels to be shipped to him there, inspired by the idea that medicines could be prepared in the laboratory to alleviate human suffering. A particularly important influence was exerted by the American chemist George Starkey, whom Boyle supported in his research into the possibility of metallic transmutation. For Boyle, the

quest was justified in terms of a moral obligation to exploit the resources of nature for human benefit, but it led him far afield from topics recognized today as scientific. In the late 1670s, he again engaged in intense experimental work, leading to what he thought was success in making gold on at least one occasion. After Boyle's death, Isaac Newton, who had his own obsessive interest in the subject, thought that Boyle might have left information among his papers about the secrets of transmutation. He was to be disappointed.

It is not possible, therefore, to understand all of Boyle's activities in terms of the ideas that prevail in modern science. Hunter makes the case that a more fundamental theme in Boyle's life was his personal piety and the associated preoccupation with leading a moral life. This concern preceded Boyle's scientific work and it survived into his last days, when he was anticipating his own death and consulting with leading churchmen about matters that weighed on his conscience. In the course of his life, he studied scripture and religious doctrine intensely, and he supported the work of Protestant missionaries in America, Asia and his native Ireland. He justified his scientific interests by reference to this religious outlook, sometimes in a rather tortuous manner when it came to arcane matters like alchemy. In general at that time the study of nature was defended as the study of God's works. Boyle was a pioneer in allying natural theology—the belief that God's attributes could be discerned in the natural world—with empirical scientific inquiry. The marriage of Christian faith with experimental science was to hold firm long after Boyle's time and was seriously challenged only in the 19th century.

This biography shows the centrality of Boyle's religious faith to his work, but Hunter makes no grand claims for an underlying unity in his subject's worldview. In fact, he rarely steps back

to make larger interpretive claims at all, sticking rigorously to the documented facts of Boyle's life and to the texts of his writings. Hunter's basic argument is that Boyle was a more complicated individual than has been realized hitherto, which is no doubt true but scarcely distinguishes him from many other people. It is unfortunate that the book does not make a strong case for Boyle's importance to readers who are not already convinced of it.

One reason for this is that Hunter seems to share one of his subject's abiding characteristics, what Hunter calls his "scrupulosity." Boyle was notorious for the convolutions and hesitations of his writings, for torturing his own con-

science on subtle moral questions, and for his reluctance to endorse theoretical speculation that went beyond the certified facts. Hunter writes intriguingly about these personal qualities, which he clearly admires, but which he acknowledges had adverse effects on Boyle's prose style. That style, Hunter declares, reflects the author's sense of "the complexity of issues and a concomitant desire to multiply testimony in order to reinforce his case." Stylistically, Hunter follows a similar path, citing evidence abundantly for the verifiable facts but scrupulously avoiding going beyond them into what he sees as the realm of speculation. The choice is responsible both for the strengths of this book and for its limitations.

There is little reason to doubt that Hunter has written what will be the first point of reference for future inquiries concerning Boyle. It covers every aspect of his life and work, with comprehensive citations of primary and secondary sources in the endnotes and in an extensive bibliographical essay. The very thorough index and the table of Boyle's whereabouts at each stage of his life will increase the book's value for specialists, who will surely come to regard it as indispensable. Some readers, however, may find them-

selves overwhelmed by the density of factual detail and will long for a few more interpretive—even speculative—remarks that would help make sense of Boyle's career as a whole. It would be unfortunate if this led nonspecialist readers to overlook the merits of this authoritative study of a very significant figure in the history of science.

Jan Golinski is professor of history and humanities at the University of New Hampshire, where he currently serves as chair of the Department of History. His books include Making Natural Knowledge: Constructivism and the History of Science (2nd edition, 2005) and British Weather and the Climate of Enlightenment (2007), both published by University of Chicago Press.



Boyle's air pump, which was constructed for his use by Robert Hooke in 1659, was a key piece of equipment in the experiments Boyle describes in his 1660 book *New Experiments, Physico-Mechanical, Touching the Spring of the Air and its Effects*. These experiments strikingly demonstrated the physical properties of air, showing that it had the capacity to exert pressure and to expand. This drawing of the pump was used to illustrate the book. From Boyle.

Land Portraits

Modern practice in cartography favors the plan view—the landscape seen as if from an infinite height—but earlier mapmakers were more flexible about perspective. In this map of the Portuguese colony of Macao, drawn in 1646 by Pedro Barreto de Resende, an oblique view conveys information about both the horizontal layout of the town and the vertical scale of the terrain. This technique of land portraiture was known at the time as *chorography*. The view of Macao is one of about 90 maps reproduced in *Mapping the World: Stories of Geography*, by Caroline and Martine Laffon (Firefly Books, \$39.95). The Laffons emphasize that maps bring us more than the geographic coordinates of a place; they tell us stories about the landscape. For example, in the Barreto map of Macao, the most conspicuous features are fortifications, churches and houses built by the Portuguese. “As for the local population,” the Laffons write, “as on many colonial maps, they seem to be overshadowed by the new ruling class.”—*Brian Hayes*



Heading South

It's refreshing to see an S above the compass arrow on a map—and a little disconcerting. This map of South Asia, made by the editors of *Himal* magazine, places south at the top and north at the bottom, giving visual importance to features and countries that don't always receive it. India, dwarfed by China on conventional maps, is prominent here, and Sri Lanka takes center stage. The map appears in the collection *Strange Maps: An Atlas of Cartographic Curiosities* (Viking Studio, \$30). Frank Jacobs, the author of the book and of a blog with the same name, reminds us that the convention of placing north at the top of a map is just that—a convention. He also notes that maps made in the Middle Ages often place east at the top, which is why we speak of orientation. Reversed maps such as this one are good reminders of how the representations of the world that we create shape our perceptions of place. *Strange Maps* contains many more thought-provoking maps, with engaging commentary. While we are turning southward, it's worth noting another example: a map of the varieties of barbecue sauce favored across the American state of South Carolina.—*Anna Lena Phillips*

MARINE BIOLOGY

Cruising for a Bruising

Rick MacPherson

SEASICK: Ocean Change and the Extinction of Life on Earth. Alanna Mitchell. x + 161 pp. University of Chicago Press, 2009. \$25.

At the conclusion of his Darwin Medal Lecture at the 11th International Coral Reef Symposium in 2008, Terry Hughes, who is director of the Australian Research Council's Centre of Excellence for Coral Reef Studies at James Cook University, projected two side-by-side images onto massive screens in the darkened hall. On the left was an image of a canoe in which two passengers sat, comfortably dry and smiling. On the right was the same canoe, only upside down, with the passengers in the water. Hughes explained that these are the two equilibrium states for a canoe: upright and capsized. At equilibrium, the canoe resists shifting from one state to the other. But with enough forcing, a tipping point is reached at which the canoe can shift rapidly into the opposite state of equilibrium, sometimes to the dismay of the passengers.

Hughes's apt metaphor underscored a key message of his lecture: that coral reefs have tipping points as well. And although they may resist change at first, showing few outward signs of stress, when shifts do take place, they can occur more rapidly than anyone had previously predicted and are tremendously difficult to reverse.

This warning forms the backbone of *Seasick: Ocean Change and the Extinction of Life on Earth*, by veteran science journalist Alanna Mitchell. Mitchell trawls the oxygen-depleted oceanic dead zones in the Gulf of Mexico, counts the days after the full moon in Panama to figure out when to search for signs of coral spawn, questions what a souring ocean chemistry holds for the future of marine plankton communities, and recounts the actions that have depleted global fisheries, documenting the toll that one frightening assault after another has taken on our ocean. Their cumulative effect has pushed us across a threshold. It appears that global systems may already be unable to return the ocean to its former state and are beginning instead to interact to create a new, far less hospitable state.

Faced with the myriad ways humans are changing the ocean, Mitchell admits that giving in to despair would be easy. Instead, she chooses a personal voyage of discovery in an effort to get to the bottom of things—in some instances literally (more on that later). Immersing herself in what Richard Feynman called “the pleasure of finding things out,” she goes straight to the primary sources, traveling with top scientists and taking part in their fieldwork. Nancy Rabalais, Ken Caldera, Joanie Kleypas, Nancy Knowlton, Boris Worm, Jerry Blackford—her list of mentors and guides reads like a fantasy lineup of ocean-science all-stars.

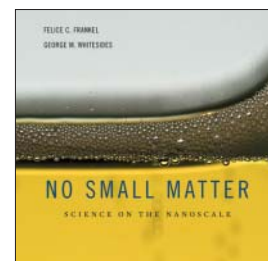
Mitchell's quest for reasons to be hopeful is daunting. At one point, on a grueling 11-day oceanographic cruise near New Orleans, she works to sample and map a small portion of the dead zone, a 17,000-square-kilometer area of water south of Texas and Louisiana, where the Mississippi River discharges into the Gulf of Mexico. Heavy agrochemical runoff into the Mississippi eventually spills into the Gulf, where it acts as fertilizer for phytoplankton, creating massive algal blooms. The blooms eventually die and sink, and bacterial decomposition effectively depletes any available oxygen from the surrounding water. Over time, layer by layer, dead zones stack up atop the continental shelf. Mitchell notes that as a result of climate change, dead zones are both increasing in number (there are now more than 400 of them globally) and thickening, as the top of the stack moves closer to the surface.

Mitchell finds connections between ocean distress and climate change nearly everywhere she goes. Looking for spawning coral in Panama, she discovers that its reproductive cycle has been weakened as a consequence of coral bleaching caused by increased sea-surface temperatures. She climbs the Pyrenees in Spain with geologists who are searching for evidence of climate disruptions during the Paleocene-Eocene Thermal Maximum, a dramatic warming of the Earth's atmosphere that took place 55 million years ago.

NO SMALL MATTER

Science on the Nanoscale

Felice C. Frankel &
George M. Whitesides



“It's exceptionally rare that science is rendered in such lucid, thoughtful,

ful, charming fashion. But I am not sure I've ever encountered a beautiful book as important as this one.” —Kurt Andersen, author of *Heyday*, host of public radio's Studio 360

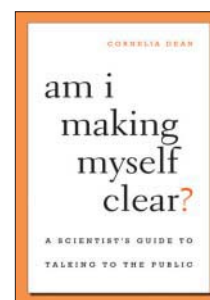
Belknap Press / new in cloth / \$35.00

★ ★ ★ ★ ★

AM I MAKING MYSELF CLEAR?

A Scientist's Guide to
Talking to the Public

Cornelia Dean



“Highly regarded New York Times science reporter Cornelia Dean presents a handbook for any scientist called upon to talk to a reporter, go on

television, lobby legislators or in general answer that age-old question, What exactly is it you do?” —*Publishers Weekly*

New in cloth / \$19.95



HARVARD UNIVERSITY PRESS

WWW.HUP.HARVARD.EDU

BLOG: HARVARDPRESS.TYPEPAD.COM

But perhaps of greatest concern to her is the insidious threat to oceans posed by high levels of carbon dioxide in our atmosphere. When atmospheric carbon dioxide dissolves in seawater, it forms carbonic acid. The more CO₂ there is in the atmosphere, the more acidic seawater becomes; ultimately this reduces the amount of carbonate that is available in the water. Carbonate is critical for the formation and maintenance of calcium carbonate, which makes up the shells of mollusks and planktonic foraminiferans as well as the limestone that coral polyps produce to create reef architecture.

Here Mitchell's scientist guides can offer little comfort. No one has come up with a way to mitigate the threat posed by ocean acidification. Mitchell writes hopefully of the possibility that the nations of the world will set targets that maintain atmospheric CO₂ levels near 380 parts per million. But news from the recent Copenhagen Climate Summit makes that seem unlikely.

Yet despite the book's barrage of grim realities, and setting aside for the moment the fact that Mitchell overestimates the effectiveness of both the International Convention for the Regulation of Whaling and the Convention on International Trade in Endangered Species, I found the argument for hope and change that she presents compelling.

At the start of the final chapter, overwhelmed by the thought that the ocean may be terminally ill, Mitchell finds herself on the verge of despair. Nevertheless, she resolves to go through with a trip to a depth of 3,000 feet in a submersible. There she experiences a resurgence of hope:

Shivering in my undersea womb, peering at these wondrous, ancient life forms, it occurs to me that we are in an era that holds out the potential of magnificent regeneration. We could, if enough of us wanted to, form a new relationship with our planet. We could become the gentle symbionts we were meant to be instead of the planetary parasites we have unwittingly become.

As Mitchell emphasizes in the epilogue, the future is in our hands.

Rick MacPherson is a marine ecologist and is Conservation Programs Director for the Coral Reef Alliance, an international biodiversity conservation organization working exclusively to protect coral reefs. His interests include the history and philosophy of science and evolutionary theory.

BIOLOGY

The Conditions for Existence

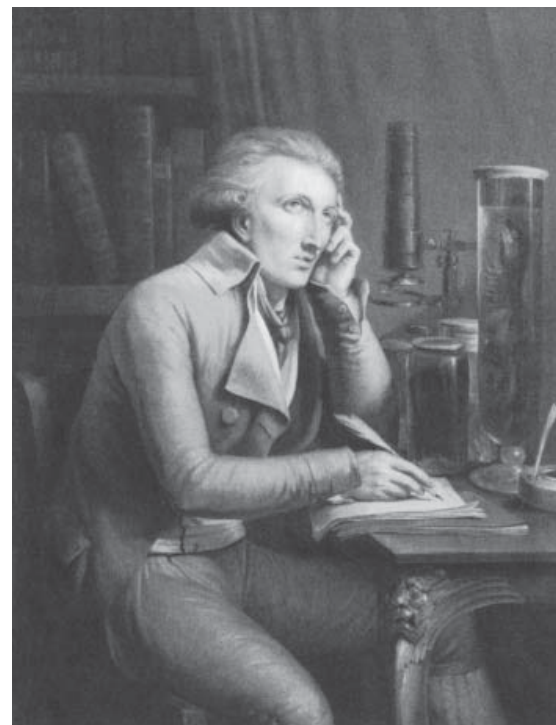
John Dupré

NOT BY DESIGN: Retiring Darwin's Watchmaker. John O. Reiss. xviii + 422 pp. University of California Press, 2009. \$49.95.

Followers of the debate between evolutionists and various waves of creationists, most recently the advocates of "intelligent design," will have been struck by one curious convergence between the views of the opposing parties. Both sides agree that life, whether or not literally designed by an intelligent agent, seems just as if it had been designed. Richard Dawkins intentionally picks up William Paley's famous example of the watch that could only have come about through deliberate design, adding to it the suggestion that the designer—for Dawkins, natural selection—is a blind watchmaker. Daniel Dennett, another prominent scourge of the creationists, is equally sure that design is a fundamental and inescapable concept for analyzing life. It has always seemed to me that this notion is a mistaken one, but it is far from easy to explain exactly why. John Reiss's *Not by Design: Retiring Darwin's Watchmaker* provides the best-worked-out explanation I've encountered.

The book opens with an extended journey through the history of biology. The specific focus of this journey is the dialectic between those who see the world and the living things within it as saturated with design and purpose, and the truly committed naturalists and materialists who have no truck with any of this. The former group includes the majority of the leading luminaries in standard accounts of biology, starting with Plato and Aristotle and concluding with no less a personage than Darwin. On the other side of the debate are the Epicureans in antiquity and the 18th-century French *philosophes*, among others. Reiss quotes what is probably the most widely cited version of the Epicurean view, that given by Philo in David Hume's *Dialogues Concerning Natural Religion*, part 8:

Is there a system, an order, an oeconomy of things, by which matter can preserve that perpetual agitation, which seems essential to it, and yet maintain a constancy



Georges Cuvier (1769–1832) was a self-taught naturalist whose interest in comparative anatomy and fossil bones led him to try to reconstruct the history of life on Earth, focusing on the extinction of species by catastrophes. His legacy became one of orthodoxy, but his conservatism was that of any good scientist, says John O. Reiss. This portrait, painted by Mathieu-Ignace van Brée, shows Cuvier at about age 29, a few years after his rapid rise to fame. From *Not by Design*.

in the forms, which it produces? There certainly is such an oeconomy: For this is actually the case with the present world. The continual motion of matter, therefore, in less than infinite transpositions, must produce this oeconomy or order; and by its very nature, that order, when once established, supports itself, for many ages, if not to eternity. But wherever matter is so poised, arranged, and adjusted as to continue in perpetual motion, and yet preserve a constancy in the forms, its situation, must, of necessity, have all the same appearance of art and contrivance which we observe at present.

Reiss's goal is to reassert such a thoroughgoing materialism and remove teleology from our vision of nature.

The somewhat surprising hero of this historical narrative is the early 19th-century naturalist Georges Cuvier. Cuvier's particular importance is in his development of the idea of the "conditions of existence"—or, as Reiss prefers to translate this, the conditions *for* existence. These must not be confused with the concept familiar to modern readers of Darwin as the conditions of life, the external circumstances to which an organism must be adapted if it is to survive. The conditions for existence are, rather, those features of a living thing without which it could not survive. In Philo's words, they are the ways that matter is "poised, arranged, and adjusted . . . to . . . preserve a constancy in the forms." The central idea of *Not by Design* is that the demonstration that some feature is part of the conditions for existence of an organism, together with the observation that the organism does indeed exist, is in general as much explanation of the presence of this feature as we can expect.

The book begins, appropriately, with an examination of the scope and limits of teleology—the explanation of the existence of a thing (or of a property or behavior of a thing) in terms of a future state toward which it conduces. The upshot of this is that teleology is acceptable under only three conditions: first, when there are deterministic laws that bring about the end in question, as when the temperature of a house changes after the thermostat is adjusted; second, when the connection to the future state is mediated by the intention of an agent; or third, in instances of what Reiss calls *conditional/functional explanation*, the variety of explanation illustrated by the conditions for existence: Given that there is a system that does X (survives, for example) and has such and such a feature that enables it to do X, then the system must have this feature (or some functionally equivalent feature) if it is to exist at all. In this last case, of course, the end (survival) is an observation that explains the current state of things; there is no question of a future state figuring into the explanation.

Where is the objectionable teleology to be found in the history and current practice of biology? Reiss's most illuminating formulation of an answer to this question is found in his objection to the gap that so many biologists, most no-

tably Darwin, have assumed between existence and adaptedness. For Reiss, an organism cannot exist—by definition—unless it fully satisfies its conditions for existence. The only sense in which a kind of organism may be said to increase in adaptedness is that its population may be growing. It may, for example, increase the size of its fundamental niche (the set of conditions in which it could in principle exist)—by displacing a competitor, say. One must not, however, suppose that the niche exists externally to the organism, as something that somehow creates a target to which an organism is attracted as an end. This is the kind of teleology that Reiss is consistently attacking. One source of it is the misleading analogy Darwin draws between natural selection and artificial selection. In the latter, there is indeed a goal, the intention of the breeder. (Reiss identifies the shifting-balance theory of Sewall Wright as another perspective on evolution led astray by the same analogy.)

Where does this leave natural selection? Reiss distinguishes several modes of selection. *Broad-sense selection* is that which maintains the satisfaction of the conditions for existence by individuals; tautologically, individuals that fail to

satisfy those conditions do not survive. *Medium-sense selection* is the average differential survival and reproduction of genotypic or phenotypic classes of organisms within a population and can be measured by the rate of increase of the class. And *narrow-sense selection* is differential survival and reproduction among classes to the extent that this is caused by the distinguishing characteristics of these classes. According to Reiss, it is narrow-sense selection that was the (necessary) contribution made by Darwin and Alfred Russel Wallace to our understanding of evolution. Narrow-sense selection is essential to explaining some changes in populations that track changes in the environment, but of course it does not imply a constant move toward some externally given optimum or state of better design.

An especially interesting consequence of the rejection of the distinction between existence and adaptedness is that it puts the topic of genetic drift in quite a different light. To summarize very crudely Reiss's detailed discussion, the way drift can be distinguished from the effects of selection is that the latter involves a move to a more adapted state. But this, of course, assumes that there is a distinction

Science is Based on Evidence! Shouldn't Science Heroes Be?!

Who Saved the Most Lives in History?

How Many Do You Know?

Biostatistics by
Amy R. Pearce, PhD
Published by Quill Driver Books

What is more important than saving a life?

Scientists Greater Than Einstein makes an evidence based argument that the greatest scientists of the twentieth century were health scientists. The evidence – over 1.6 billion lives saved! Coming from the fields of agronomy, chemistry, epidemiology, medicine, microbiology, and ophthalmology, these individuals had more impact upon humanity over the past 150 years than any other scientists. See the results of the first extensive survey undertaken to discover who saved the most lives in history. At ScienceHeroes.com over 100 lifesaving scientists are profiled, and 10 are featured in the book where each chapter illuminates their formidable research in rich detail.

How many on the Top 10 lists do you know? Scrutinize the lists of medical scientists, living scientists, women scientists, and more at ScienceHeroes.com

between existence and adaptedness, whereas Reiss regards that distinction as illegitimate. Conceptually, selection and drift are quite different processes, but in practice they can be extremely difficult to separate. Once we see that the trajectory of a population through time is one in which adaptedness is always maintained—the conditions for existence are continuously met—it is very difficult to distinguish among the causes of this maintenance.

What is most fundamental—and encompasses selective processes, drift and much else besides—is the meeting over time of the conditions for existence of a lineage. (This, incidentally, is a concept that appears several times in the book, but it seemed to me that it might helpfully have been separated more sharply from the parallel concept for an organism.) The existence of an organism requires that it be part of a lineage that meets the conditions for the existence of the lineage of which it is a part—the survival and reproduction of its sequence of members. The

conditions for the existence of the lineage would seem, therefore, to be the most fundamental concept.

This is a difficult, sometimes dense and sometimes frustrating book—and my attempt to summarize its main theses probably shares those characteristics. Anyone interested in evolutionary biology is likely to disagree with some of the claims that Reiss makes. It is, however, an important book that should be widely read and discussed. As we gradually recover from the orgy of Darwin adulation that has marked the year of his anniversaries, nothing is more needed than a reminder that evolution remains a topic about which we are far from knowing all the answers. The Darwinolatry of some popularizers has suggested that the discovery of natural selection—perhaps with a subsequent assist from Gregor Mendel—left little more to be done than a tedious filling in of details. The enduring debates with creationists have also undoubtedly tended to discourage admission that major conceptual issues about evolution remain

unresolved. On the contrary, however, the decisive point that needs to be made again and again in these debates is that the openness to advance, the progressiveness, of scientific thought is precisely what distinguishes it most significantly from creationist dogma. Reiss's book contributes much to this goal.

It is a great pity that a book such as this cannot be written at the same level of accessibility as the popular neo-Darwinist works that it explicitly or implicitly opposes. It may be that an anthropomorphic understanding of nature by analogy to design is difficult for the human mind to avoid. But this book is a good illustration that the effort is worth making.

John Dupré is professor of philosophy of science and director of the ESRC Centre for Genomics in Society (Egenis) at the University of Exeter. He is the author of, among other books, Darwin's Legacy: What Evolution Means Today (Oxford University Press, 2003) and The Constituents of Life (Van Gorcum, 2008). He is also coauthor, with Barry Barnes, of Genomes and What to Make of Them (University of Chicago Press, 2007).

ORNITHOLOGY

Avian Appreciation

Aaron French

BIRDSCAPES: Birds in Our Imagination and Experience. Jeremy Mynott. xiv + 367 pp. Princeton University Press, 2009. \$29.95.

THE BIRD: A Natural History of Who Birds Are, Where They Came From, and How They Live. Colin Tudge. xvi + 462 pp. Crown Publishers, 2008. \$30.

The beauty and mystery of birds have inspired thousands of books about all aspects of their diversity, behavior, morphology, conservation and identification. Yet as two recent arrivals, *The Bird* and *Birdscapes*, demonstrate, those topics have not yet been exhausted.

The Bird, by science writer Colin Tudge, is the more typical book by far. A full 20 percent of the text is devoted to chapter 4, "All the Birds in the World: An Annotated Cast List." Other sections describe what makes a bird a bird, how birds live and how we live with birds. Fortunately, the writing is lively and appealing, and the text is filled with interesting tidbits of information. Readers learn, for example, that modern broiler chickens, "raced from egg to puffed-up oven weight in six weeks," don't live long enough to grow a sturdy wishbone.

Tudge's chapter describing the orders and families of all the world's birds will be of great interest to the casual bird-watcher or wildlife enthusiast—it reads more like a catalog of wonders than an ornithological manual. However, by necessity each entry is extremely brief, and to knowledgeable readers some of his omissions are glaring. When discussing the Hawaiian honeycreepers, for example, Tudge fails to mention that they are among the most critically endangered birds in the world.

Despite this lack of comprehensiveness, *The Bird* will be a welcome addition to the library of any bird lover because it is so enjoyable to read.

Jeremy Mynott's *Birdscapes* is much less conventional. Mynott, a lifelong birder and former publishing executive, writes in the preface that the book "has been in the nature of an explora-



tion for me, a journey whose sights and sounds I did not fully foresee when I started and whose destination was unclear." And a strange journey it is. Mynott discusses not just species differences, birdsong, conservation and nomenclature, but such matters as how humans have used images and metaphors of birds to piece together ideas, which birds people profess to like the most, and how our interest in birds is affected by conceptions of rarity and

beauty. The book's recurring themes, he says, are "the snares of sentimentality, the pros and cons of anthropomorphism, the interplay between what we perceive in birds and what we project onto them, and the power of metaphors, names, and symbols to express or distort our vision." The tone varies, ranging from playful, conspiratorial and poetic to dryly academic, thoughtful and poignant. Mynott is erudite and insightful, but his meandering was not always to my taste. The text sometimes struck me as self-indulgent and strangely lacking in focus.

Opening the book at random, you might find a passage from *Romeo and Juliet* (was it a lark that Romeo heard, or a nightingale?), an exploration of how to see nature properly (with an allusion to Oscar Wilde's suggestion that nature is just an unsatisfactory imitation of art), or a discussion of how French bird names differ from English ones. You could also come across something that seems at first glance to have nothing whatever to do with birds—illustration 24, for example, which consists of photographs of four nude actresses on stage at the Windmill Theatre in London. Mynott explains that British authorities decreed in 1940 that onstage nudity was acceptable as long as actresses were in poses that were "motionless and expressionless." His point is that animation has a key effect on the observer and is fundamental to our reactions to birds. This may well be, but for me he has strayed too far off topic here.

Who is the intended audience for *Birdscapes*? Mynott certainly has enthusiasm to spare, but his style is too pedantic for a popular audience. The book appears to be aimed at intellectual birders who love literature. They are likely to delight in Mynott's erudition and find the book's idiosyncrasies charming.

To illustrate the differences in approach between *The Bird* and *Birdscapes*, let's examine the way Tudge and Mynott cover similar ground. In chapter 3 of *The Bird*, "Keeping Track: The Absolute Need to Classify," Tudge jumps right into the fray, asserting,

It's a simple question of the kind six-year-olds ask: How many kinds of birds are there? But as with most of the questions that six-year-olds ask, the answer is that nobody knows, and nobody can ever know—at least not exactly.



Some bird names "are little more than vague expressions of admiration," notes Jeremy Mynott; the Australian fairy wren on the lower left is known as the splendid fairy wren, and the one at lower right is the superb fairy wren. More commonly, names describe the bird's physical appearance in some way, noting colors, patterns or features; the bird at top left is a red-backed fairy wren, for example, and the one at top right is a variegated fairy wren. Size, activity, voice, preferred food, favored habitat and the name of the person who discovered the species are other factors often used as the basis for bird names. From *Birdscapes*.

What follows is a fairly standard 30-page overview of taxonomy and systematics, progressing in a predictably breezy fashion from Linnaeus to Darwin and beyond, ending with the revolutionary DNA-DNA hybridization studies of Charles Sibley and Jon Edward Ahlquist. Tudge stays within the traditional boundaries of the topic and provides a discussion that newcomers will welcome.

Mynott is more circuitous. He opens his chapter on nomenclature and classification, "Seeing a Difference," with an anecdote about a birding trip to the Isles of Scilly, where he sees but fails to recognize a semi-palmated sandpiper, "a five-star rarity." He misidentifies it as a stint and is then corrected by ornithologist Peter Grant. "What had Peter noticed that I had missed?," he wonders. He then goes off on one tangent after another as he discusses distinctions and differences, species and individuals, observing and perceiving, illusion and self-deception, and patterns and pro-

files. He squeezes in just three lines from Darwin's *On the Origin of Species* but expounds for pages on Sherlock Holmes, master observer and perceiver. Mynott brings the discussion back around to birds by proposing that the three most important attributes of birdwatchers are "1) active attention, 2) informed expectation, and 3) ambition of imagination." It is clear that attribute number three is where his interest lies.

Neither of these books gives a full picture of birds and birding, but both are entertaining and contain much that's worth knowing. Choose *The Bird* if you like to devour a book in one or two sittings. *Birdscapes* is best read piecemeal—you'll want to consume it in small bites, in birdlike fashion.

Aaron French, who has a master's degree in ecology, spent two years living with the Baka pygmies in Cameroon, studying birds and monkeys. He is now the chef of the Sunny Side Café in Albany, California. His Web site can be found at <http://www.eco-chef.com/>.

NANOVIEWS

FORDLANDIA: The Rise and Fall of Henry Ford's Forgotten Jungle City.

Greg Grandin. Metropolitan Books, \$27.50.

In 1927, Henry Ford bought a Connecticut-sized piece of the Amazon and built an authentic American town in the Brazilian jungle, complete with electric lights and indoor plumbing. In Ford's conception, Fordlandia would be an independent source of raw materials for his burgeoning auto empire, and a way to preserve the vanishing America of his Michigan childhood. His enormous wealth and willpower enabled him briefly to establish a utopia in the jungle, complete with golf courses, ice cream parlors, movie theaters and Victrolas. But these were succeeded by brothels, bars and disease. Like the Lincoln Zephyr in the photograph at right, stuck in Fordlandia mud, Ford's experiment finally foundered in the wilderness, and in 1945 he sold the whole property back to Brazil.

In *Fordlandia: The Rise and Fall of Henry Ford's Forgotten Jungle City*, New York University historian Greg Grandin recounts the whole tale, from Ford's first searches for an independent source of rubber to his culminating dream of a civilizing engine in the Amazon. One U.S. diplomat in



Brazil, trying to explain why Ford was committed to a venture unlikely to be profitable, wrote to his superiors in the State Department that "Mr. Ford considers the project as a 'work of civilization.' . . . Nothing else will explain the lavish expenditure of money."

The story of the jungle suburb is so outlandish that many writers would be tempted to reduce it to a fable—an ecological parable or a screed against imperialism—or to draw analogies to Joseph Conrad or El Dorado. But Fordlandia was a

real, complex endeavor, and Grandin refuses to simplify its lessons. For him, ultimately, it's a window into the curious character of Henry Ford, a self-made titan whose own factories had begun to cast a shadow across the America that had produced him. That his Amazon adventure failed shows the limits of what his hubris could accomplish, but that he embarked on the adventure at all reveals his tortured idealism, and that elevates *Fordlandia* to a quixotic tragedy.—Greg Ross

CROW PLANET: Essential Wisdom from the Urban Wilderness. Lyanda Lynn Haupt. Little, Brown, \$23.99.

"How, exactly, are we connected to the earth, the more-than-human world, in our lives and in our actions? And in light of this connection, how are we to carry out our lives on a changing earth?" These are the hard questions that Lyanda Lynn Haupt sets out to explore in the memoir *Crow Planet*. Crows, for Haupt, represent both the continued presence of the wild in places dominated by humans, and the narrowing down of ecological diversity as those changes to the landscape make it harder for some species to exist.

One crow she watches collects shells, dried berries and shiny bits of trash; in a similar fashion, Haupt gathers the work of many fine writers around her. There are references to

David Buddbill's poetry and Jennifer Price's environmental history along with plenty of corvid science, as well as the requisite (and still relevant) quotes from Rachel Carson. All this is woven in with Haupt's own musings as she hangs clothes on the line, talks with her young daughter, fights depression, and works to learn more about crows and urban ecosystems. For the most part she succeeds in pulling together the narrative threads of her personal life, specific ecological communities and humans' impact on ecosystems as a whole. Some sections fly by; others, such as her hours-long observation of a dead crow in the tradition of Louis Agassiz, allow the reader to luxuriate in the writer's deep contemplation of the natural world. Throughout, her descriptions of how crows communicate, nest, mate and live are fascinating.



Haupt's approach is that of the flaneur, so it's fitting that the answers she comes to, if not always conclusive, feel useful and encouraging of more exploration.—Anna Lena Phillips

sigma xi today

March-April 2010 · Volume 19, Number 2

2010 Sigma Xi Awards Announced



Michael J. Spivey, a professor of cognitive science at the University of California, Merced, known for his innovative studies of language and visual perception, will

receive Sigma Xi's 2010 William Procter Prize for Scientific Achievement, the Society's highest honor.

The Procter Prize and other top annual awards will be presented at the Sigma Xi Annual Meeting and International Research Conference next November in Raleigh, North Carolina.

The 2010 John P. McGovern Science and Society Award will go to **Barbara Gastel** at Texas A&M University. A professor of veterinary integrative biosciences and of humanities in medicine, she has devoted much of her career to improving scientific communication.



Howard R. Moskowitz, an expert on sensory psychology and its commercial application, will receive the Walston Chubb Award for Innovation. He is president and CEO of Moskowitz

Jacobs Inc. in White Plains, New York.

And **Kevin R. Gurney** will be honored with Sigma Xi's Young Investigator Award. He is an associate professor of earth and atmospheric science at Purdue University whose work on tracking CO₂ emissions has been groundbreaking.



Procter Prize winner Michael Spivey has a long history of studying language and visual perception. He was the driving force in creating a new line of research in psycholinguistics.

He uses eye-tracking and computer mouse-tracking equipment to study how humans perceive and respond to what they hear and see.

Motion-tracking software and hardware document not only the subjects' final

(continued on next page)

www.sigmaxi.org

From the President



In Support of Sigma Xi

Giving back has been a recurrent theme and even the title of my last installment of "From the President." I suppose the value I see in supporting the scientific enterprise is derived from my own gratitude to all of those who have helped me and shared in the joys of my own experience as a teacher, mentor and researcher.

Sigma Xi has always stood for values that I respect and has provided me with a means to give back. I now wish to encourage you to give back and to support Sigma Xi in its mission to enhance and promote the scientific enterprise.

A common complaint about our annual meeting has been the level of political wrangling that may seem to dominate the meeting. Therefore, it was very heartening to me to have so many of the delegates at this year's annual meeting approach me about taking a more active role in Sigma Xi.

I believe this likely occurs at each annual meeting; however, because I presided as president at this meeting I became more aware of this response. There are in fact many ways to serve Sigma Xi and at many different levels of commitment.

The easiest way to start is at the chapter level. Is your chapter as active as you would like? Are the current programs of your chapter in line with your interests and commitments? Is your chapter not serving you well, or do you just want to get more involved?

You can support the activities of your chapter or serve as an agent for change. Most of us started our service to Sigma Xi in just that way. We stepped up, took on the responsibility for programs we wanted to make happen and evolved as leaders of our local chapters. Yes, it was work, but clearly to many of us the rewards were significant enough to cause us to seek greater involvement in a Society whose value we hold dear.

Sigma Xi, at the international level, operates through a committee system where our members bring knowledge, experience and chapter-wide perspectives of the many issues that the Society must deal with. Membership on one of our committees is often the starting point for many in the governance of the Society.

Our Web site, www.sigmaxi.org, lists the committees of Sigma Xi and provides an e-mail link to volunteer for service on a specific committee or on any committee in general. You should know that to maintain continuity on committees the turn over of membership is cyclical, so please be patient and your opportunity will come.

Finally, those who wish to dedicate themselves to the future of the Society can become directly involved in its governance. Serving on the nominations committee for your region or constituency group is often an introduction to this level of commitment.

Taking on the directorship of a region or constituency group places you on the Board of Directors. This is a significant commitment that requires your time to work in your region or constituency to support chapters and promote Sigma Xi. It also brings with it the fiduciary responsibilities of serving on the board.

Most importantly it brings you the satisfaction of taking a stand for what you believe in and the friendship that grows out of working closely with like-minded individuals who share a common cause. I highly recommend that you become involved; it will be a rewarding experience both for you and the Society.

Howard Ceri

award winners

Thirty-Five Students Receive Medals at Sigma Xi Conference

Thirty-five student researchers received medals and cash awards for their poster presentations at the 2009 Sigma Xi International Research Conference in Texas.

More than 200 students presented their research at this year's conference, representing nearly 100 academic institutions.

The winners of a special Student Choice Award were **Aditya Kaddu**, **Zhao Kong** and **Daniel Rist** of Rice University. The award was sponsored by the Washington, D.C., Chapter of Sigma Xi and carried a \$250 cash prize. Medalists for superior presentations were as follows:

Doctoral Candidates

Interdisciplinary Research

Pearce Creasman—Texas A&M University

Physics & Astronomy

Derek Nowak—Portland State University

Math & Computer Science

Faisal Reza—Duke University

Graduate Students

Ecology & Evolutionary Biology

Anna Coleman-Hulbert—Portland State University

Geo-Sciences

Ruth Mullins—Texas A&M University

Engineering

David Kvale—University of Toledo

Undergraduate Students

Biochemistry

Valeria Gonzalez—University of California, Irvine

Behavioral Sciences

Jamar Whaley—Queens College; **Michael Gonzalez**—University of California, Irvine

Cellular & Molecular Biology

Jing Han—Northwestern University; **Franklin Garcia** and **Mayra Carrillo**—University of California, Irvine; **Vineet Singal**—Stanford University; **Hatim Thaker** and **Danny Jandali**—Northwestern University

Chemistry

Derek Rhoades—Ohio Northern University; **Abdul Jangda**—University of Houston, Downtown

Ecology & Evolutionary Biology

Elizabeth Lavoie—State University of New York, Plattsburgh; **Krystle Minear**—Weber State University

Engineering

David Garland and **Kenneth Davis**—Rice University; **Maha Haji**—University of California, Berkeley

Environmental Science

Chang Woo Lee—University of Texas at M.D. Anderson Cancer Center

Geo-Sciences

James Burnes—Lamar University

Interdisciplinary Research

Patricia Troy—Ohio Wesleyan University; **Michael Chien**—University of Pennsylvania; **Wee Leow**—Texas A&M University

Math & Computer Science

Kyle Pounder—Saint Mary's College of California

Physics & Astronomy

Jake Connors—The Ohio State University

Physiology & Immunology

Erick Maravill—University of California, Irvine

High School Students

Cellular & Molecular Biology

Mirza Shabbir—Harlem Children Society

Ecology & Evolutionary Biology

Gabriel Joachim—Cibola High School

Chemistry

Rodney Agnant—Harlem Children Society •

Sigma Xi Leaders in Washington, D.C.



In Washington, D.C., recently, Sigma Xi President-elect Joseph Whittaker (left) and Executive Director Jerome Baker flank Sigma Xi member and U.S. Congressman Rush Holt after discussing increasing budgets for federal science agencies.

2010 Sigma Xi Awards

(continued from previous page)

answers but also the answers they considered along the way. The end result is a more accurate representation of how the human brain processes information.

McGovern Award winner Barbara Gastel is Knowledge Community Editor for AuthorAID, a major project of the International Network for the Availability of Scientific Publications. She co-authored the sixth edition of Robert Day's *How to Write and Publish a Scientific Paper* and is now co-authoring the seventh edition.

These new editions speak to globalization and digitizing of publishing. Gastel wrote the *Health Writer's Handbook*. She is chief editor of *Science Editor*, the periodical of The Council of Science Editors.

Chubb Award winner Howard Moskowitz created a new technology, called Mind Genomics, to better understand the way consumers think about products and about social issues.

The technology creates and links scientific based databases into a system called Rule Developing Experimentation (RDE). RDE helps companies worldwide to optimize products, messaging and graphic design.

Young Investigator Award winner Kevin Gurney focuses his research on the global carbon cycle, understanding sinks for atmospheric CO₂, how CO₂ changes connect to climate change and how to connect good climate science to development of sound public policy.

He was the lead author on a 2002 paper addressing CO₂ inversions that is listed in the top 1 percent of *Nature* papers. He received a grant from NASA to build a CO₂ emissions inventory for the U.S. and led a project to create a high-resolution, interactive map of U.S. carbon dioxide emissions from fossil fuels.

The 2010 Sigma Xi Honorary Members will be announced at a later time. Profiles of Sigma Xi award winners will appear in upcoming issues of *American Scientist*. •



© 2008 JupiterImages Corporation.

How Has Christianity Changed over 2,000 Years?

In the first centuries after Christ, there was no “official” New Testament. Instead, early Christians read and fervently followed a wide variety of scriptures—many more than we have today.

Relying on these writings, Christians held beliefs that today would be considered bizarre. Some believed that there were 2, 12, or as many as 30 gods. Some thought that a malicious deity, rather than the true God, created the world. Some maintained that Christ’s death and resurrection had nothing to do with salvation while others insisted that Christ never really died at all.

What did these “other” scriptures say? Do they exist today? How could such outlandish ideas ever be considered Christian? If such beliefs were once common, why do they no longer exist? These are just a few of the many provocative questions that arise from **Lost Christianities: Christian Scriptures and the Battles over Authentication**, an insightful 24-lecture course taught by Professor Bart D. Ehrman, the Chair of the Department of Religious Studies at the University of North Carolina at Chapel Hill and the author and editor of 17 books, including *The New York Times* best-seller *Misquoting Jesus*.

This course is one of The Great Courses®, a noncredit, recorded college lecture series from The Teaching Company®. Award-winning professors of a wide array of subjects in the sciences and the liberal arts have made more than 300 college-level courses that are available now on our website.

ACT NOW!

1-800-TEACH-12
www.TEACH12.com/4amsc

Lost Christianities: Christian Scriptures and the Battles over Authentication

Taught by Professor Bart D. Ehrman,
 The University of North Carolina at Chapel Hill

Lecture Titles

- | | |
|---|---|
| 1. The Diversity of Early Christianity | 13. The Acts of John |
| 2. Christians Who Would Be Jews | 14. The Acts of Thomas |
| 3. Christians Who Refuse To Be Jews | 15. The Acts of Paul and Thecla |
| 4. Early Gnostic Christianity—Our Sources | 16. Forgeries in the Name of Paul |
| 5. Early Christian Gnosticism—An Overview | 17. The Epistle of Barnabas |
| 6. The Gnostic Gospel of Truth | 18. The Apocalypse of Peter |
| 7. Gnostics Explain Themselves | 19. The Rise of Early Christian Orthodoxy |
| 8. The Coptic Gospel of Thomas | 20. Beginnings of the Canon |
| 9. Thomas' Gnostic Teachings | 21. Formation of the New Testament Canon |
| 10. Infancy Gospels | 22. Interpretation of Scripture |
| 11. The Gospel of Peter | 23. Orthodox Corruption of Scripture |
| 12. The Secret Gospel of Mark | 24. Early Christian Creeds |

Order Today!
Offer Expires Tuesday, May 11, 2010

Lost Christianities: Christian Scriptures and the Battles over Authentication
Course No. 6593

24 lectures (30 minutes/lecture)

DVDs ~~\$254.95~~ **NOW \$69.95**
 + \$10 Shipping, Processing, and Lifetime Satisfaction Guarantee

Audio CDs ~~\$179.95~~ **NOW \$49.95**
 + \$10 Shipping, Processing, and Lifetime Satisfaction Guarantee

Priority Code: 40363

THE TEACHING COMPANY®
The Joy of Lifelong Learning Every Day™
 GREAT PROFESSORS, GREAT COURSES, GREAT VALUE
 GUARANTEED.™



Partners
In Health

Over 20 years of health and social justice



STAND WITH HAITI

Within 72 hours of the earthquake, PIH had deployed **several teams of medical personnel** and **several tons of supplies** to people desperate for food, water, and medications. **Within a week**, we had treated **thousands of patients** in **twenty operating rooms** run by PIH staff, partners, and volunteers. **Additional volunteers and supplies** are being sent **every day** as we continue to provide **care and comfort** to patients who have lost everything.



Donate now www.pih.org