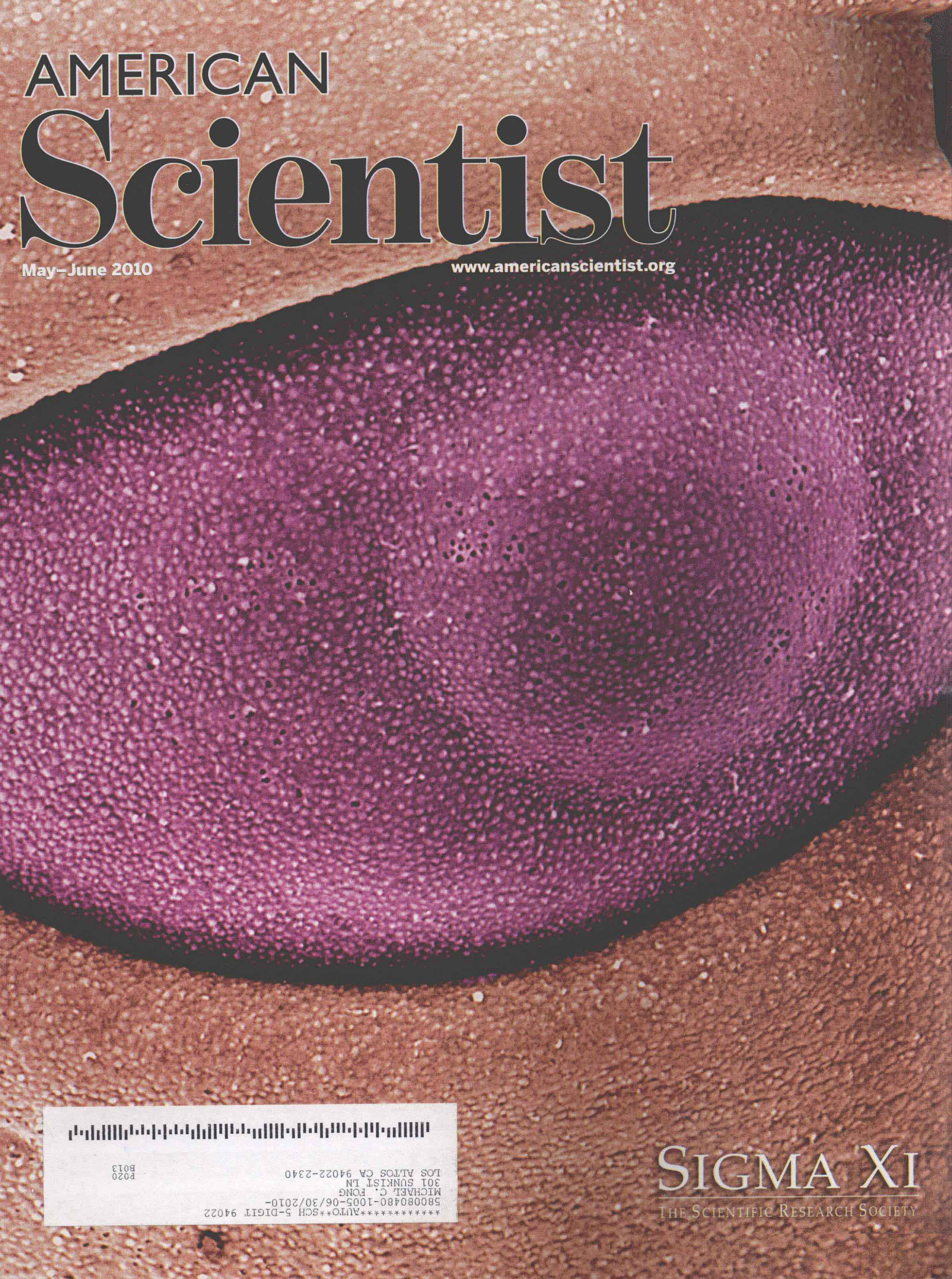
AMERICAN
# Scientist
May–June 2010
www.americanscientist.org





SIGMA XI
THE SCIENTIFIC RESEARCH SOCIETY

AMERICAN
# Scientist

Volume 98 • Number 3 • May–June 2010

## Departments



## Scientists' Bookshelf

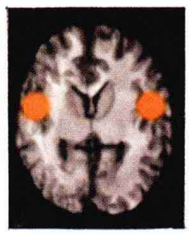


## From Sigma Xi

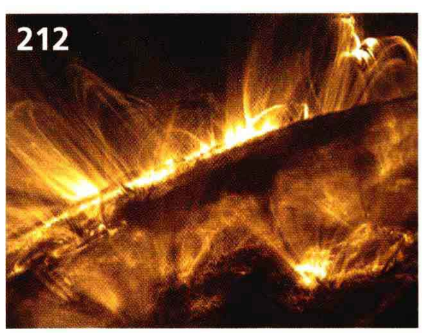


## Feature Articles

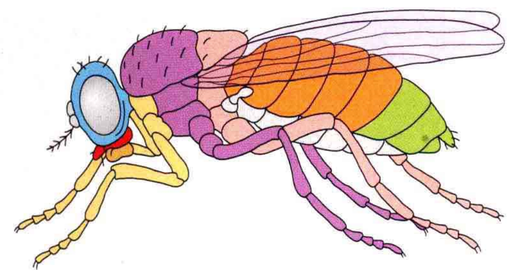




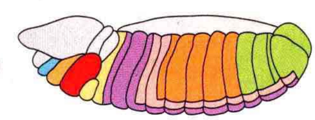
204



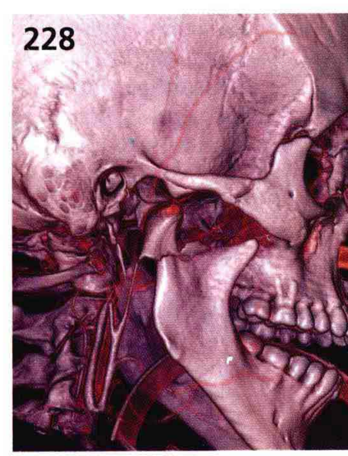


220



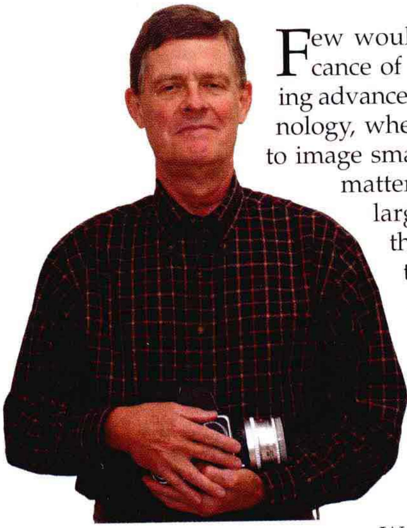





## The Cover

*The developmental paths of the human eye—shown on the cover in a colored scanning electron micrograph during the eighth week of gestation—and the octopus eye are examples of convergent evolution. Although formed by entirely different processes and having somewhat different functions, they are remarkably similar. In "Development Influences Evolution" (pages 220–227), evolutionary developmental biologist Katherine Willmore describes this as one example of developmental forces interacting with selective pressures to determine the course of evolution. In a larger sense, developmental constraints on evolution are the reason that every life form on Earth can be lumped into one of 35 different body plans, all of which originated in the Cambrian Period about 500 million years ago. Known by its practitioners as evo-devo, the study of such interactions is helping science to understand why life has adopted so many variations. (Photograph by SPL/Photo Researchers Inc.)*

# Technology in Its Place



Few would question the significance of scientific and engineering advances made possible by technology, whether it be in the ability to image smaller and smaller bits of matter or analyze larger and larger sets of data. Nonetheless, I often hear—particularly when hanging around field geologists of a certain age—that something is lost when electronics intervene between investigator and subject.

That idea lurks at the core of Darin Wolfe's argument in "To See for One's Self" (pages 228–235). Wolfe documents a decline in the rate of autopsies in the U.S. from about 50 percent of hospital deaths in the 1950s to about 6 percent at present. He points out that many factors have contributed to the decline, including technological developments such as virtual autopsy that are making it less necessary. But he also finds that as anatomical dissection has become a lesser part of medical education, the distance between physician and the physical manifestations of disease has grown, depriving medical students of a valuable source of understanding.

Alex (Sandy) Pentland has made a career of understanding human communication, especially the nonverbal. But curiously enough, as he relates in "To Signal Is Human" (pages 204–211), technology allows his team to quantify such signaling. Employing motion sensors, cameras and recorders, Pentland's group is revealing how it is that we manage to make valid judgments about others at least 70 percent of the time. Moreover, the team has identified a catalog of behaviors that allow them to predict, through observation, who is most likely to be a successful leader.

Yet sometimes seeing can be deceiving. As Richard Woo explains in "Revealing the True Solar Corona" (pages 212–219), astrophysicists attempting to understand the Sun's atmosphere may have long been misled by what their eyes told them. Vision, he notes, is far from an ideal empirical tool; the brain constructs images from data in ways that match our experiences and expectations. Because the human eye and its processor are unable to integrate the vast range of brightness in the solar corona, investigators incorrectly concluded that the solar wind emanates from particular regions corresponding to the white light we perceive. Data from instruments, however, now document the broad emission of the wind from the entire solar disk. So we've come full circle: Sometimes it may be best not to see for one's self.

Turn to page 182, and you'll notice among the Letters to the Editors a green box describing some of the magazine's digital offerings. *American Scientist Online* is considerably more than an HTML version of what you hold in your hands. For example, more than 12,000 readers now receive Science in the News Weekly every Monday as an e-mail. SitNW, as we call it, compiles entries from Science in the News Daily and comments on them. One of its more popular features is a ranking of the top three stories of the week, but I personally most look forward to my weekly dose of Mark Heath science cartoon. You may also be interested in online-only interviews with prominent scientist/writers about their communication efforts and reading proclivities. And finally, I confess with some reluctance: We now Tweet too.—*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*
Eric Tolliver *Marketing Associate*

ADVERTISING SALES
Kate Miller *Advertising Manager*
advertising@amsci.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@amscionline.org • subs@amsci.org

PRINTED IN USA

PUBLISHED BY SIGMA XI
THE SCIENTIFIC RESEARCH SOCIETY
Howard Ceri *President*
W. Franklin Gilmore *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

*American Scientist* gratefully acknowledges support for "Engineering" through the Leroy Record Fund.

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.*

### The Wrong Culprit?

To the Editors:

In his interesting "Short History of Hydrogen Sulfide" (January–February), Roger P. Smith mentions strontium sulfide as a potential source of the hydrogen sulfide in homes where Chinese drywall was installed in South Florida, New Orleans and elsewhere. It's hard for this retired chemist to conceive that any source of gypsum contains sufficient strontium to generate so large a problem. That's especially so when a primary raw material is 25 percent sulfur.

Anyone who has ventured into stagnant areas of the many swamps in Florida, for instance, has noted the odor of so-called swamp gas. The origin of this hydrogen sulfide is well known. It comes from the action of anaerobic bacteria on organic material, using the oxygen in sulfate ions for metabolism, reducing sulfur to hydrogen sulfide.

I propose an alternative explanation regarding the drywall. The gypsum there, either intentionally or naturally, likely was contaminated with some proportion of organic material. That contamination in an oxygen-free location, such as the interior of drywall, was then exposed to water, thanks to the high humidity of the subtropics. That mix, at a suitably warm temperature, provides the essential environment for anaerobic bacteria to do their thing.

As a retired chemist, I have no access to a laboratory to conduct simple tests to verify my proposal. I hope someone else will.

Barton Milligan
Fort Lauderdale, FL

Dr. Smith responds:
I am doubly indebted to Dr. Milligan. First he points out the controversy surrounding the assertion that strontium sulfide is the source of hydrogen sulfide in homes damaged by Chinese wallboard. Google "Chinese wallboard" and "strontium sulfide" for a sampling of opinions. Difficulty with the idea that wallboard could contain enough sulfide to account for "so large a problem" is shared by many.

Secondly, Dr. Milligan reminds us of an important source of environmental hydrogen sulfide, namely the reduction of sulfate by anaerobic bacteria, often in swamps. Hence, the origin of the term "swamp gas." Although he proposes an interesting idea, my question is this: How anaerobic do conditions have to be for those bacteria to do their thing? It comes down to how low the oxygen



## Illustration Credits

**Computing Science**
Pages 187, 188, 189 Barbara Aulicino

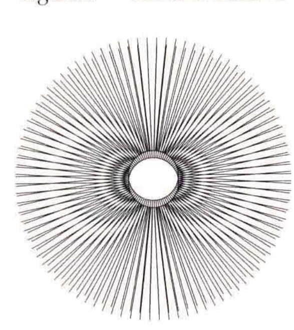
**Engineering**
Page 193 Barbara Aulicino

**To Signal Is Human**
Figures 2, 3, 4, 7 Tom Dunne

**Revealing the True Solar Corona**
Figures 4 *(left)*, 6, 9 Tom Dunne

**Development Influences Evolution**
Figures 2, 3, 4, 7, 8 Barbara Aulicino

**To See For One's Self**
Figure 6 Barbara Aulicino



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.  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.

tension is in the interior of the wallboard? Wallboard must have a certain porosity to let the hydrogen sulfide out, which in turn would allow oxygen to come in.

At least we agree that hydrogen sulfide is the most likely culprit; such accord is a good and sometimes rare thing.

#### Harness DNA Memory

To the Editors:

I was quite intrigued by Kurt D. Bollacker's recent column, "Avoiding a Digital Dark Age" (March–April). The article pointed out two substantial shortcomings in storing information digitally. First, a single error in a digital recording can have rather dramatic effects. Thus, one needs not only to continually back up and copy digital information but also to do this in an error-correcting fashion. Second, when copying digital information, one has to be mindful that computer formats keep changing. It is important to keep current with formats and remain backwards compatible.

I would like to propose a somewhat whimsical suggestion for overcoming these two issues. Perhaps we could encode digital information that we wish to preserve in the DNA of bacteria—more specifically, in the regions between genes. Note, first of all, that the information would be recorded in a most fundamental and universal format, the natural DNA code of A, G, C and T. Secondly, because bacteria naturally replicate, backups of the information would be made in an error-correcting fashion using DNA polymerase. Although this copying is subject to some random mutation that evades correction, we could increase the fidelity by averaging the "readout" over a population of bacteria, rather than taking it from a single individual.

Mark Gerstein
Yale University

To the Editors:

Kurt Bollacker's problem with restoring his file backups is an excellent example of the importance of free and open-source software (FOSS). Dr. Bollacker lost the ability to recover his files because he had lost his copy of the proprietary software that created the backup. Also, the company that created the software no longer exists and he wasn't able to locate the software on the Internet.

While much FOSS software comes without a price tag, the term "free" in this context refers not to the cost but to the principle that the software can be copied, studied, changed and improved. The availability of source code avoids the common problem that the type of machine that ran the original software no longer exists. The source code can be recompiled or rewritten for another machine.

Devlin Gualtieri
Ledgewood, N.J.

Dr. Bollacker responds:

I believe that open-source software is a necessary part of solving the problem of digital data preservation. There are many benefits and advantages to FOSS that help the world of software development and data handling. The mutability, lack of licensing cost and didactic qualities of FOSS make our world a better place. But these benefits are not directly related to digital data preservation. The main virtue of open-source software in data preservation is that it can be thought of as highly precise (if not easy to read) documentation for digital formats. We don't need a manual if the software can

## online @ AMERICAN Scientist.org

Find feature articles, columns, book reviews and Science Observer reports from our latest issue at *amsci.org*. Also, see previous issues dating back to 1998. (Access to some material is limited to Sigma Xi members and subscribers.)

### Exclusively online:

**Science in the News**
A concise roundup of science stories collected each weekday from multiple media outlets, with links to the original articles

**On the Bookshelf**
*American Scientist* book reviews and interviews with scientist authors about their work and their favorite books

**Pizza Lunch Podcasts**
Listen to *American Scientist* Pizza Lunch talks and hear scientists present research to nonscientists

**E-Newsletters**
subscribe at *www.amsci.org/enews*

**Science in the News Daily:** A roundup of science news stories delivered every weekday

**Science in the News Weekly:** A weekly summary the top science news

**Scientists' Bookshelf Monthly:** Science book news, author interviews and reviews

***American Scientist* Update:** An early peek at each new issue, with descriptions of feature articles, columns, Science Observer news and book reviews

***American Scientist* on Twitter**
twitter.com/AmSciMag

be recompiled to read old data and, ideally, convert it into new formats. Even if ancient software can no longer be run, the source can allow its functionality to be duplicated in new code. The need for such documentation, and thus the need to preserve source code, is diminished if the data formats are very simple and/or are highly "self documenting" (as are some XML schemas). Ideally, the formats would be self documented, and the source code would still be around, giving us a nice bit of redundant protection.

### Can't Break that Law

To the Editors:

The passive engulfment model for lunge feeding in "The Ultimate Mouthful" by Jeremy Goldbogen (March–April) appears to violate the law of conservation of momentum because it shows the whale's speed dropping to zero. As a most basic first approximation of the process, if an average whale engulfs a mass of water equal to its own mass, the velocity after engulfment should be one half the velocity before engulfment. Factoring in active swimming and changes in drag during engulfment will modify that result, but any model that stops a whale dead in its tracks should have been stopped dead in its tracks.

Peter Kaiteris
West Hempstead, NY

Dr. Goldbogen responds:

Peter Kaiteris is correct. The "passive engulfment" scenario of Figure 9 indeed shows a whale that is suddenly decelerating to a zero speed at the moment of complete buccal cavity filling—clearly an unrealistic sequence of events. But this never occurred in our simulations because they were instructed to automatically terminate whenever the cavity filled up. Such an instantaneous drop in speed was included to show that "passive engulfment" fails. Our computer model did account for the momentum transfer between whale and engulfed mass, as can be seen in the "active engulfment" graph shown in that same figure. In this case, the whale's and engulfed water's final speeds (that is, by the time the mouth closes) are indeed at about half the whale's initial speed. That is to be expected because most of the force sustained by the whale comes from the reaction to its forward push of the engulfed water.

### Water Molecules Not Repelled

To the Editors:

The Science Observer article "Sunburned Ferns?" (March–April) invokes a widespread scientific misconception. The error is the idea that water is repelled by some surfaces, in this case ginkgo leaf surfaces. In fact, water molecules and all hydrophobic substances attract one another, although weakly. The false belief in repulsion traces to the original use of the misleading descriptive term "hydrophobic." A valid exposition of the interaction of water with hydrophobic substances is found in a 2002 *Nature* article by David Chandler. Chandler states, "the term hydrophobic (water-fearing) is commonly used to describe substances that, like oil, do not mix with water. Although it may look as if water repels oil, in reality the separation of oil and water in ambient conditions is not due to repulsion ... but to particularly favorable hydrogen bonding between water molecules." Also widely in current use is the term "superhydrophobic," referring to surfaces that often are described, erroneously, as repelling water. A 2006 *Nano Letters* paper by L. Zhai et al. deals with this topic. The authors explain, "Water repellent in the context of our work simply means that water droplets placed on the surface will roll off freely at a small angle. This is how we and many others define the superhydrophobic state (contact angle greater than 150° coupled with a small rolling angle)." Not surprisingly, the forces involved—cohesion, adhesion, and gravity—are all attractive.

J. Lee Kavanau
University of California at Los Angeles

### How to Write to *American Scientist*

Brief letters commenting on articles that have appeared in the magazine are welcomed. The editors reserve the right to edit submissions. Please include a fax number or e-mail address if possible. Address: Letters to the Editors, American Scientist, P.O. Box 13975, Research Triangle Park, NC 27709 or editors@amscionline.org.

**Erratum**

In "The Ultimate Mouthful: Lunge Feeding in Rorqual Whales" (March–April), the cover image and Figure 2 should have been labeled as fin whales.

# Designing Minds

Edward A. Wasserman and Mark S. Blumberg

THE BASIC ARGUMENT of intelligent design was famously set forth in the watchmaker analogy of William Paley in 1802: The complexity and functionality of a watch imply a watchmaker; analogously, the complexity and functionality of living things also imply a designer, albeit one vastly more potent than a mere watchmaker. This argument rests on a simple analogy between the design of human artifacts and the design of natural forms. For the analogy to work, we must first accept that we design our inventions with purpose and foresight. On this point, most evolutionists and creationists agree. What distinguishes these two camps is that, when accounting for the origin of living things, proponents of intelligent design summon a divine creator, whereas evolutionists credit natural selection. Thus, evolutionists share with creationists the same understanding of *design*; they differ only in how they invoke it.

Discussions of design are prominent in the writings of evolutionists from Darwin to Dawkins. Pondering the implications of his theory of natural selection for Paley's "old argument of design in nature," Charles Darwin wrote in his autobiography that we can no longer argue that "the beautiful hinge of a bivalve shell must have been made by an intelligent being, like the hinge of a door by man.

*Edward A. Wasserman (ed-wasserman@uiowa.edu) is Dewey B. and Velma P. Stuit Professor of Experimental Psychology at the University of Iowa and coeditor with Thomas Zentall of* Comparative Cognition: Experimental Explorations of Animal Intelligence *(Oxford University Press, 2006). Mark S. Blumberg (mark-blumberg@uiowa.edu) is F. Wendell Miller Professor of Psychology at the University of Iowa, editor-in-chief of* Behavioral Neuroscience *and author most recently of* Freaks of Nature: What Anomalies Tell Us About Development and Evolution *(Oxford University Press, 2009). Both are members of the Delta Center at the University of Iowa, dedicated to the investigation of learning, development, and change.*

*How should we explain the origins of novel behaviors?*

There seems to be no more design in the variability of organic beings and in the action of natural selection, than in the course which the wind blows. Everything in nature is the result of fixed laws." A century later, Richard Dawkins pursued the issue of design and divided the world "into things that look designed (such as birds and airliners) and things that don't (rocks and mountains)." He further divided those things that look designed into "those that really are designed (submarines and tin openers) and those that aren't (sharks and hedgehogs)."

What did Dawkins mean when he wrote of things that "really are designed"? In *The Blind Watchmaker,* he provided a clear answer: "All appearances to the contrary, the only watchmaker in nature is the blind forces of physics….A true watchmaker has *foresight:* He designs his cogs and springs, and plans their interconnections, with a future *purpose* in his mind's eye" [emphasis added].

Such uncritical acceptance of purpose and foresight in human design may well be unwise. After all, do we really know how door hinges and can openers were created? In fact, we may know less about the origins of these everyday contrivances than we know about the origins of bivalve shells, sharks and hedgehogs. By attributing the origins of animals and artifacts to different kinds of designers—one blind, the other intelligent—both Darwin and Dawkins lapse into the same kind of "designer thinking" that ensnared creationists like Paley. Such thinking rests on the familiarity and deceptive simplicity of mentalistic explanations of behavior, as when Dawkins uncritically appeals to the foresight and purpose of the watchmaker rather than entertaining possibly deeper questions about the origins of the watch. He may be giving human designers too much credit.

### Form Follows Failure

The engineer Henry Petroski has written extensively and convincingly about our often misguided characterizations of the origins of human inventions. In *The Evolution of Useful Things* (1993), Petroski argues that artifacts "do not spring fully formed from the mind of some maker but, rather, become shaped and reshaped through the (principally negative) experiences of their users…." In short, form follows *failure*, not *function*.

And what about those failures? It is all too easy to forget that the first attempts at flight featured impossible aircraft with flappable wings, man-of-war sails, and box-kite frames. Do we see the origins of today's jumbo jets in those early, comical failures? Similarly, do we appreciate the knowledge gained by bridge builders from studying the undulating destruction of the Tacoma Narrows Bridge in Washington or, more recently, the wobbling of the Millennium Bridge in London? Do we understand that even the most tragic failures—such as the Hyatt Regency walkway collapse in Kansas City or the *Challenger* space shuttle explosion—are the consequences of human tinkering on a grand scale? Beginning with the very first glimpse of a problem or an opportunity, such failures—whether large or small, tragic or comic—prompt the fine-tuning and retrofitting that, over time, have shaped even our greatest engineering achievements, from Egyptian pyramids to medieval cathedrals to suspension bridges to spacecraft.

American Journal of Primatology/Wiley and Sons

In 2003, a wild chimpanzee named JJ was observed by Shinya Yamamoto and colleagues in New Guinea using a long, rigid tool to harvest carpenter ants *(left)*. This was the first observation of ant-fishing in trees. JJ succeeded in three of 14 attempts, with three painful bites along the way. Two years later, JJ was seen using a shorter, more flexible wand to feed without being bitten *(right)*. Was this change the result of trial and error or foresight?

It is through this plodding process that today's designs—typically instantiated in the form of a detailed blueprint—embody all of the hard, painful, but often *unacknowledged* lessons of the past. Most of us are ignorant of that history, yet we glibly proclaim that the final products were intelligently designed, thereby perpetuating the myth of the creative moment. We then carry that myth forward and attribute each new artifact to individual insight, creativity and genius. But this myth cannot cheat reality; the failures just keep coming, as most recently illustrated by the massive worldwide recall of Toyota automobiles. As Petroski notes in *To Engineer Is Human* (1985), despite their mathematically precise understanding of structural materials, engineers still cannot "calculate to obviate the failure of the mind."

Because of the writings of Darwin, Dawkins and other biologists, many of us are now open to understanding the organic world in evolutionary terms—but are we equally willing to apply such evolutionary thinking to that last bastion of designer intelligence, our minds? Curiously, just as Petroski and others are painstakingly detailing the origins of human inventions, researchers are increasingly invoking unsubstantiated mental processes to explain complex human and animal behaviors.

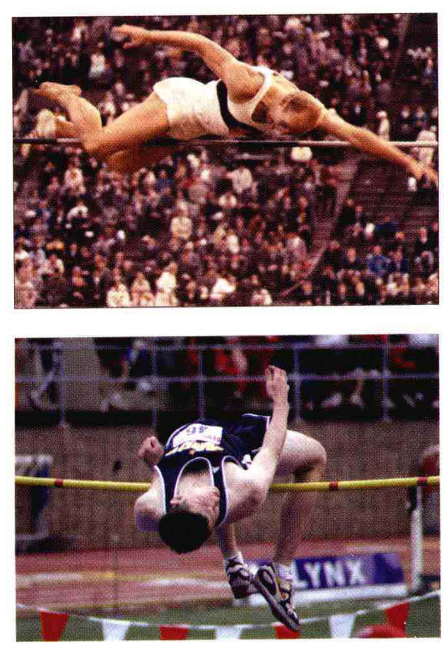
### Insight About Insight

A salient recent example can be found in a report in the *Proceedings of the National Academy of Sciences* in Spring 2009, in which crows were observed to fashion wire into hooks that were then used to retrieve out-of-reach food items. These behaviors have been interpreted by some authors as products of this species' creativity and insight. In contrast, other scientists have investigated similar "insight" problems in crows, monkeys and other animals; but by focusing on the origins of these behaviors, they have discovered the critical learning experiences, as opposed to forethought, that gave rise to them. Nonetheless, we seem to be in the midst of a resurgence of faith among some scientists that animal behavior can be explained by creativity, insight and other mentalistic concepts. For our part, we remain skeptical about the utility of such groundless explanations. Indeed, we are unconvinced that creativity and insight are proper explanations even for *human* behavior.

Of course, few people are unnerved when the cognitive prowess of crows or other animals is questioned. Things get stickier when we express similar skepticism about the human mind. Yet as with the invention of human *artifacts*, we see good reason to doubt the prevailing belief that novel human *behaviors*—what we might call behavioral inventions—are necessarily the products of a designing mind.

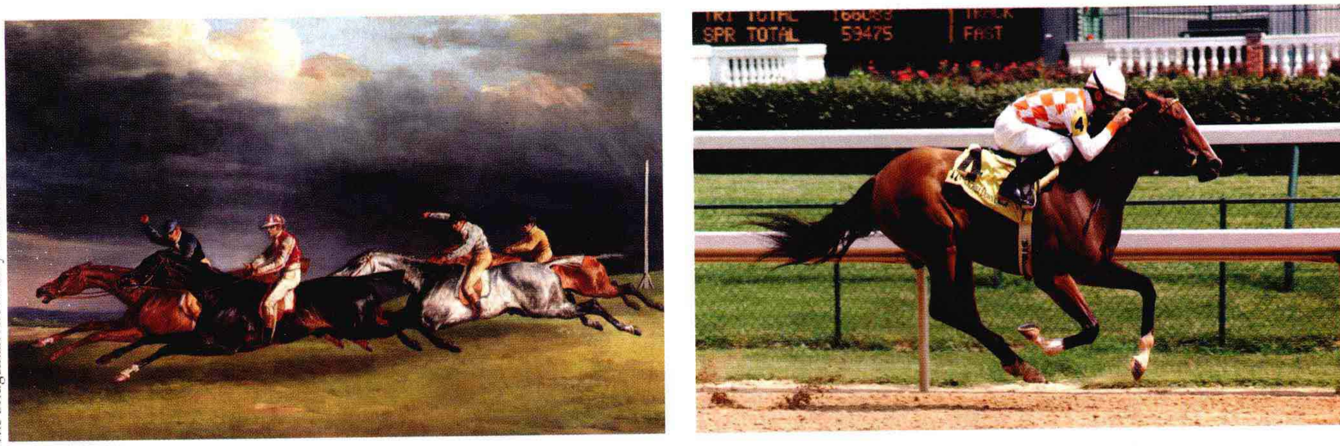
### Successful Flop

A celebrated case of human behavioral invention lends credence to our view. Dick Fosbury revolutionized the high jump with a world-record bound of 7 feet, 4 ¼ inches, which earned him a gold medal at the 1968 Olympics. Some might suspect that his innovation—the so-called Fosbury Flop—was designed with purpose and foresight in a single creative moment. In fact, it unfolded over considerable time, beginning in high school when Fosbury used the outmoded "scissors" jump. Urged by his coach to adopt the more sophisticated "straddle," his lanky body failed to comply with his coach's wishes. When Fosbury reverted to the "scissors," he began to lift his hips to reach higher altitude, thereby forcing back his head and shoulders. In this way, the flop evolved, not from design, but from a protracted trial-and-error process that combined repeated




Bettman/CORBIS

Until the Fosbury Flop *(bottom)* revolutionized high jumping, athletes used the "scissors" or "straddle" *(top)*. By Dick Fosbury's own account, the flop evolved without forethought through a trial-and-error process. After Fosbury won the gold medal at the 1968 Olympics in Mexico City, his flop quickly dominated the sport. (Photograph at bottom by Matthew L Romano/U.S. Navy.)




The Bridgeman Art Library International

**In this painting by Theodore Gericault *(left)*, jockeys in the 1821 Epsom Derby ride in a leisurely posture. A modern thoroughbred jockey at Churchill Downs *(right)* rides poised in the monkey crouch, a behavioral innovation that first appeared in the late 1800s but whose biomechanical benefits have only recently been demonstrated. (Photo at right by Jeff Kubina/Wikimedia Commons.)**

effort with the biomechanics of Fosbury's gangly physique. Here is how Fosbury himself described this process: "I began to lift my hips up and my shoulders went back in reaction to that. At the end of the competition, I had improved my best by 6", from 5' 4" to 5' 10" and even placed third! The next two years in high school, with my curved approach, I began to lead with my shoulder and eventually was going over head first like today's Floppers."

Another example of human behavioral invention from the sporting world—this one from thoroughbred racing—further supports our view. A recent report in *Science* carefully explained how the monkey crouch—the currently dominant racing style, in which the jockey rides poised above the saddle leaning forward—promotes faster racing times. At the expense of a much more strenuous ride for the jockey than the earlier, upright style, the monkey crouch confers measurable biomechanical benefits for the horse. No one has yet suggested that the monkey crouch was designed with purpose and foresight to maximize biomechanical efficiency. So how did it arise?

Some authors have credited two American jockeys with bringing the monkey crouch to England in the late 1800s. An English rider, Harding Cox, may actually have adopted this riding style a bit earlier. Critical to our present considerations, Cox suggested in his memoir the possible benefits that the monkey crouch conferred: "When hunting, I rode very short, and leant well forward in my seat. When racing, I found that by so doing I avoided, to a certain extent, *wind pressure*, which … is very obvious to the rider. By accentuating this position, I discovered that my mount had the advantage of *freer hind leverage*. Perhaps that is why I managed to win on animals that had been looked upon as 'impossibles,' 'back numbers,' rogues and jades."

Although the authors of the *Science* report emphasize the biomechanical benefit to the horse of having the jockey rise from the saddle, and they deemphasize the role of decreased wind resistance, Cox's account provides a key insight into this innovation's true origins. Specifically, decreased wind resistance may have initially encouraged Cox's forward adjustment, which allowed his later accentuation of the posture into the fully realized monkey crouch. Like a scaffold that provides a temporary structure for the construction of a building, Cox's response to wind pressure may have scaffolded his behavioral transition to a novel riding style—one that transformed modern thoroughbred racing.

Inventive behaviors are commonly attributed to creativity, insight or genius, but a far simpler explanation may do. For the Fosbury Flop and the monkey crouch, an elegant and plausible way to understand the origins of novel behaviors can be found in the law of effect, which emerged a century ago from the animal-behavior studies of psychologist Edward Thorndike. The law of effect states that *successful behavioral variations are retained and unsuccessful variations are not*. Importantly, this positively Darwinian process exists entirely outside the realm of purpose or foresight. If everything in nature is the result of fixed laws, as Darwin himself proposed, then would he not also have marveled at the explanatory power of the law of effect—which was not discovered until several decades after his death—and its compelling parallels with natural selection?

Our prime point here is the importance of the search for *origins*. Darwin has taught us that the search for the origin of species reveals the action of natural mechanisms that do not require guidance from a creative, intelligent designer. Similarly, Petroski has taught us to look beyond the romance of the iconoclastic inventor and the drama of the creative moment to appreciate the real origins of human artifacts. Petroski's insight should free evolutionists from their continuing dispute with creationists over where to draw the line between things that *really are* designed and things that *only appear* to be designed. Belief in the existence of that false line only serves to obscure the powerful *selectionist* processes that are at work in producing so many of the world's creations—both organic and synthetic.

Beyond the concerns of Darwin and Petroski, we see additional fertile ground for reshaping how we think about the origins of behavioral innovations. We have focused here on the Fosbury Flop and the monkey crouch, but we could also have discussed the role of serendipity in scientific discovery or the developmental path by which each of us learns to crawl, walk and run. From our first days of life, we are all inventors who discover by trial and error how our growing bodies work and move. As with organic evolution, the development of behavior is indeed a creative process, but it is one that unfolds without purposeful design.

# The Bootstrap

Cosma Shalizi

*Statisticians can reuse their data to quantify the uncertainty of complex models*

Statistics is the branch of applied mathematics that studies ways of drawing inferences from limited and imperfect data. We may want to know how a neuron in a rat's brain responds when one of its whiskers gets tweaked, or how many rats live in Manhattan, or how high the water will get under the Brooklyn Bridge, or the typical course of daily temperatures in the city over the year. We have some data on all of these things, but we know that our data are incomplete, and experience tells us that repeating our experiments or observations, even taking great care to replicate the conditions, gives more or less different answers every time. It is foolish to treat any inference from only the data in hand as certain.

If all data sources were totally capricious, there'd be nothing to do beyond piously qualifying every conclusion with "but we could be wrong about this." A mathematical science of statistics is possible because, although repeating an experiment gives different results, some types of results are more common than others; their relative frequencies are reasonably stable. We can thus model the data-generating mechanism through probability distributions and stochastic processes—random series with some indeterminacy about how the events might evolve over time, although some paths may be more likely than others. When and why we can use stochastic models are very deep questions, but ones for another time. But if we *can* use them in a problem, quantities such as these are represented as "parameters" of the stochastic models. In other words, they are functions of the underlying probability distribution. Parameters can be single numbers, such as the total rat population; vectors; or even whole curves, such as the expected time-course of temperature over the year. Statistical inference comes down to estimating those parameters, or testing hypotheses about them.

These estimates and other inferences are functions of the data values, which means that they inherit variability from the underlying stochastic process. If we "reran the tape" (as Stephen Jay Gould used to say) of an event that happened, we would get different data with a certain characteristic distribution, and applying a fixed procedure would yield different inferences, again with a certain distribution. Statisticians want to use this distribution to quantify the uncertainty of the inferences. For instance, by how much would our estimate of a parameter vary, typically, from one replication of the experiment to another—say, to be precise, what is the root-mean-square (the square root of the mean average of the squares) deviation of the estimate from its average value, or the *standard error?* Or we could ask, "What are all the parameter values that *could* have produced this data with at least some specified probability?" In other words, what are all the parameter values under which our data are not low-probability outliers? This gives us the *confidence region* for the parameter—rather than a *point estimate,* a promise that either the true parameter point lies in that region, or something very unlikely under any circumstances happened—or that our stochastic model is wrong.

To get standard errors or confidence intervals, we need to know the distribution of our estimates around the true parameters. These *sampling distributions* follow from the distribution of the data, because our estimates are functions of the data. Mathematically the problem is well defined, but actually *computing* anything is another story. Estimates are typically complicated functions of the data, and mathematically convenient distributions all may be poor approximations of the data source. Saying anything in closed form about the distribution of estimates can be simply hopeless. The two classical responses of statisticians have been to focus on tractable special cases, and to appeal to asymptotic analysis, a method that approximates the limits of functions.

### Origin Myths

If you've taken an elementary statistics course, you were probably drilled in the special cases. From one end of the possible set of solutions, we can limit the kinds of estimator we use to those with a simple mathematical form—say, mean averages and other linear functions of the data. From the other, we can assume that the probability distributions featured in the stochastic model take one of a few forms for which exact calculation *is* possible, either analytically or via tables of special functions. Most such distributions have origin myths: The Gaussian bell curve arises from averaging many independent variables of equal size (say, the many genes that contribute to height in humans); the Poisson distribution comes from counting how many of a large number of independent and individually improbable events have occurred (say, radium nuclei decaying in a given second), and so on. Squeezed from both ends, the sampling distribution of estimators and other functions of the data becomes exactly calculable in terms of the aforementioned special functions.

That these origin myths invoke various limits is no accident. The great re-

*Cosma Shalizi received his Ph.D. in physics from the University of Wisconsin–Madison in 2001. He is an assistant professor of statistics at Carnegie Mellon University and an external professor at the Santa Fe Institute. Address: 132 Baker Hall, Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, PA 15213. Internet: http://www.bactra.org*

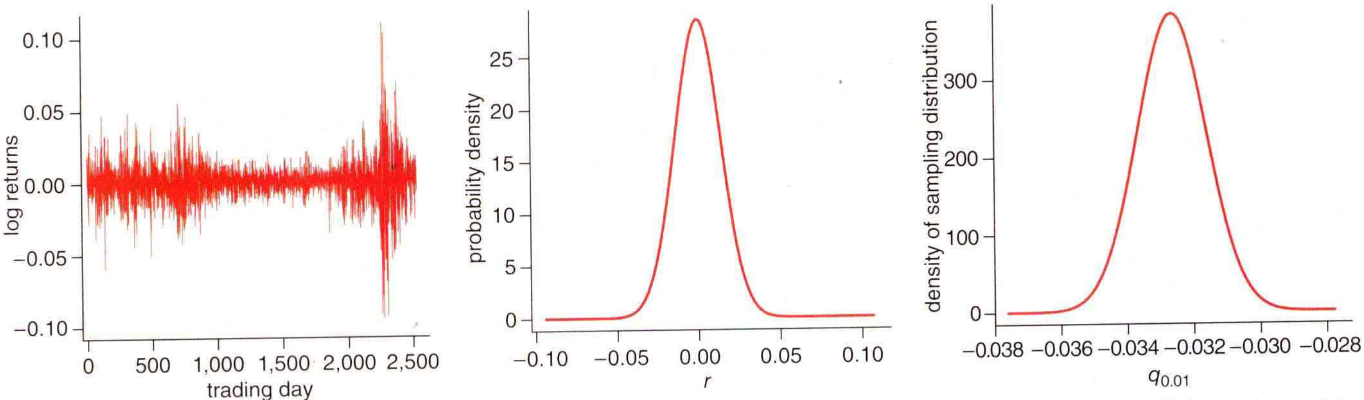



**Figure 1. A series of log returns from the Standard and Poor's 500 stock index from October 1, 1999, to October 20, 2009 *(left)*, can be used to illustrate a classical approach to probability. A financial model that assumes the series are sequences of independent, identically distributed Gaussian random variables yields the distribution function shown at center. A theoretical sampling distribution that models the smallest 1 percent of daily returns (denoted as $q_{0.01}$) shows a value of –0.0326 ± 0.00104 *(right)*, but we need a way to determine the uncertainty of this estimate.**

sults of probability theory—the laws of large numbers, the ergodic theorem, the central limit theorem and so on—describe limits in which *all* stochastic processes in broad classes of models display the same asymptotic behavior. The central limit theorem (CLT), for instance, says that if we average more and more independent random quantities with a common distribution, and if that common distribution is not too pathological, then the distribution of their means approaches a Gaussian. (The non-Gaussian parts of the distribution wash away under averaging, but the average of two Gaussians is another Gaussian.) Typically, as in the CLT, the limits involve taking more and more data from the source, so statisticians use the theorems to find the asymptotic, large-sample distributions of their estimates. We have been especially devoted to rewriting our estimates as averages of independent quantities, so that we can use the CLT to get Gaussian asymptotics. Refinements to such results would consider, say, the rate at which the error of the asymptotic Gaussian approximation shrinks as the sample sizes grow.

To illustrate the classical approach and the modern alternatives, I'll introduce some data: The daily closing prices of the Standard and Poor's 500 stock index from October 1, 1999, to October 20, 2009. (I use these data because they happen to be publicly available and familiar to many readers, not to impart any kind of financial advice.) Professional investors care more about changes in prices than their level, specifically the *log returns*, the log of the price today divided by the price yesterday. For this time period of 2,529 trading days, there are 2,528 such values *(see Figure 1)*. The "efficient market hypothesis" from financial theory says the returns can't be predicted from any public information, including their own past values. In fact, many financial models assume such series are sequences of independent, identically distributed (IID) Gaussian random variables. Fitting such a model yields the distribution function in the center graph of Figure 1.

An investor might want to know, for instance, how bad the returns could be. The lowest conceivable log return is negative infinity (with all the stocks in the index losing all value), but most investors worry less about an

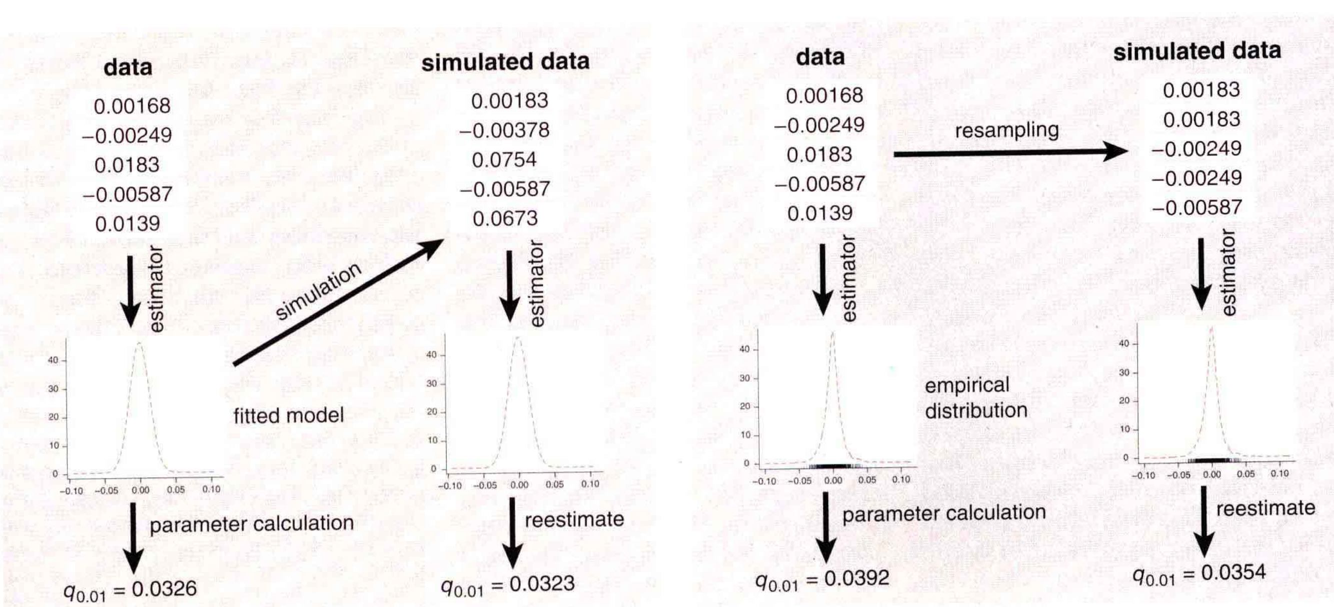



Figure 2. A schematic for model-based bootstrapping *(left)* shows that simulated values are generated from the fitted model, and then they are treated like the original data, yielding a new parameter estimate. Alternately, in nonparametric bootstrapping, a schematic *(right)* shows that new data are simulated by resampling from the original data (allowing repeated values), then parameters are calculated directly from the empirical distribution.

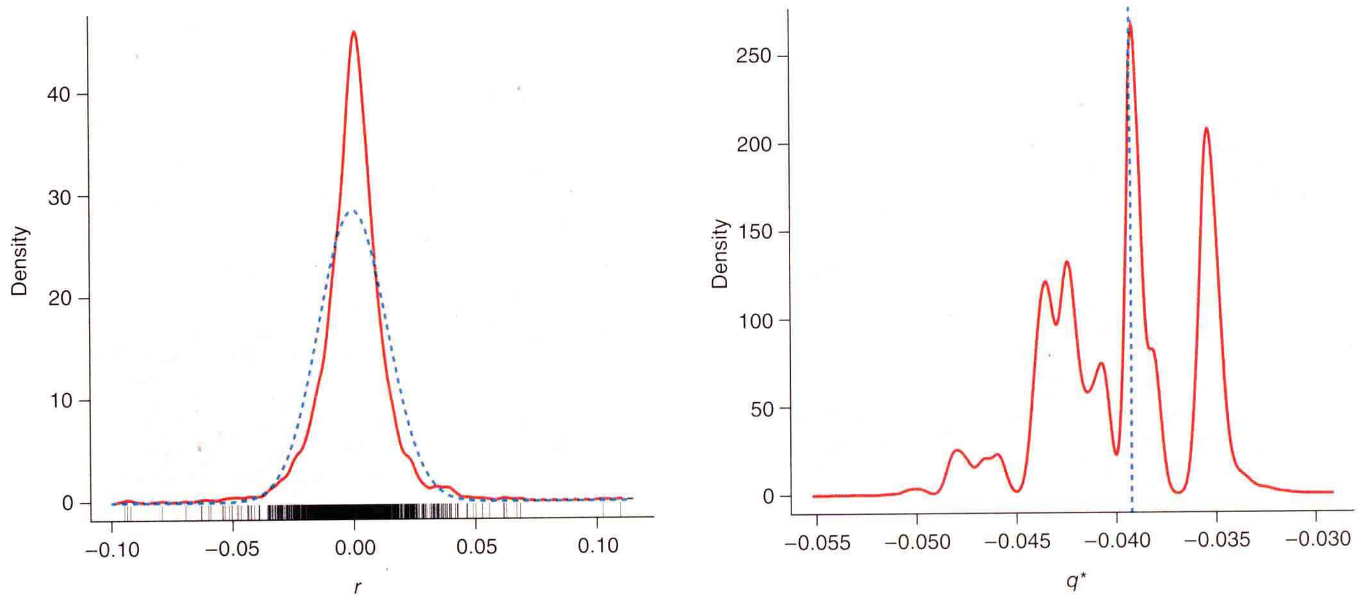



**Figure 3. An empirical distribution** ***(left, in red, smoothed for visual clarity)*** **of the log returns from a stock-market index is more peaked and has substantially more large-magnitude returns than a Gaussian fit** ***(blue).*** **The black marks on the horizontal axis show all the observed values. The distribution of $q_{0.01}$ based on 100,000 nonparametric replications is very non-Gaussian** ***(right, in red).*** **The empirical estimate is marked by the blue dashed line.**

apocalyptic end of American capitalism than about large-but-still-typical losses—say, how bad are the smallest 1 percent of daily returns? Call this number $q_{0.01}$; if we know it, we know that we will do better about 99 percent of the time, and we can see whether we can handle occasional losses of that magnitude. (There are about 250 trading days in a year, so we should expect two or three days at least that bad in a year.) From the fitted distribution, we can calculate that $q_{0.01} = -0.0326$, or, undoing the logarithm, a 3.21 percent loss. How uncertain is this point estimate? The Gaussian assumption lets us calculate the asymptotic sampling distribution of $q_{0.01}$, which turns out to be another Gaussian *(see the right graph in Figure 1)*, implying a standard error of $\pm 0.00104$. The 95 percent confidence interval is (–0.0347, –0.0306): Either the real $q_{0.01}$ is in that range, or our data set is one big fluke (at 1-in-20 odds), or the IID-Gaussian model is wrong.

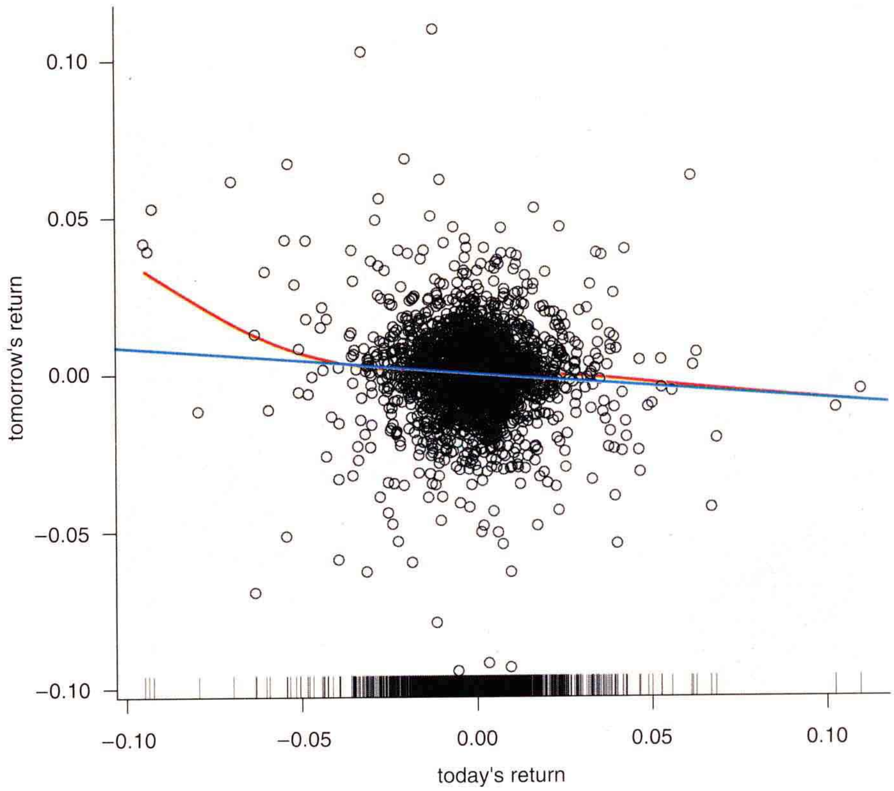



**Figure 4. A scatter plot of black circles shows log returns from a stock-market index on successive days. The best-fit line** ***(blue)*** **is a linear function that minimizes the mean-squared prediction error. Its negative slope indicates that days with below-average returns tend to be followed by days with above-average returns, and vice versa. The red line shows an optimization procedure, called** ***spline smoothing,*** **that will become more or less curved depending on looser or tighter constraints.**

### Fitting Models

From its origins in the 19th century through about the 1960s, statistics was split between developing general ideas about how to draw and evaluate statistical inferences, and working out the properties of inferential procedures in tractable special cases (like the one we just went through) or under asymptotic approximations. This yoked a very broad and abstract theory of inference to very narrow and concrete practical formulas, an uneasy combination often preserved in basic statistics classes.

The arrival of (comparatively) cheap and fast computers made it feasible for scientists and statisticians to record lots of data and to fit models to them. Sometimes the models were conventional ones, including the special-case assumptions, which often enough turned out to be detectably, and consequentially, wrong. At other times, scientists wanted more complicated or flexible models, some of which had been proposed long before but now moved from being theoretical curiosities to stuff that could run overnight. In principle, asymptotics might handle either kind of problem, but convergence to the limit could be unacceptably slow, especially for more complex models.

By the 1970s statistics faced the problem of quantifying the uncertainty of inferences without using either implausibly helpful assumptions or asymptotics; all of the solutions turned out to demand *even more* computation. Perhaps the most successful was a proposal by Stanford University statistician Bradley Efron, in a now-famous 1977 paper, to combine estimation with simulation. Over the last three decades, Efron's "bootstrap" has spread into all areas of statistics, sprouting endless elaborations; here I'll stick to its most basic forms.

Remember that the key to dealing with uncertainty in parameters is the sampling distribution of estimators. Knowing what distribution we'd get for our estimates on repeating the experiment would give us quantities, such as standard errors. Efron's insight was that we can *simulate* replication. After all, we have already fitted a model to the data, which is a guess at the mechanism that generated the data. Running that mechanism generates simulated data that, by hypothesis, have nearly the same distribution as the real data. Feeding the simulated data through our estimator gives us one draw from the sampling distribution; repeating this many times yields the sampling distribution as a whole. Because the method gives itself its own uncertainty, Efron called this "bootstrapping"; unlike Baron von Münchhausen's plan for getting himself out of a swamp by pulling himself out by his bootstraps, it works.

Let's see how this works with the stock-index returns. Figure 2 shows the overall process: Fit a model to data, use the model to calculate the parameter, then get the sampling distribution by generating new, synthetic data from the model and repeating the estimation on the simulation output. The first time I recalculate $q_{0.01}$ from a simulation, I get -0.0323. Replicated 100,000 times, I get a standard error of 0.00104, and a 95 percent confidence interval of (–0.0347, –0.0306), matching the theoretical calculations to three significant digits. This close agreement shows that I simulated properly! But the point of the bootstrap is that it doesn't rely on the Gaussian assumption, just on our ability to simulate.

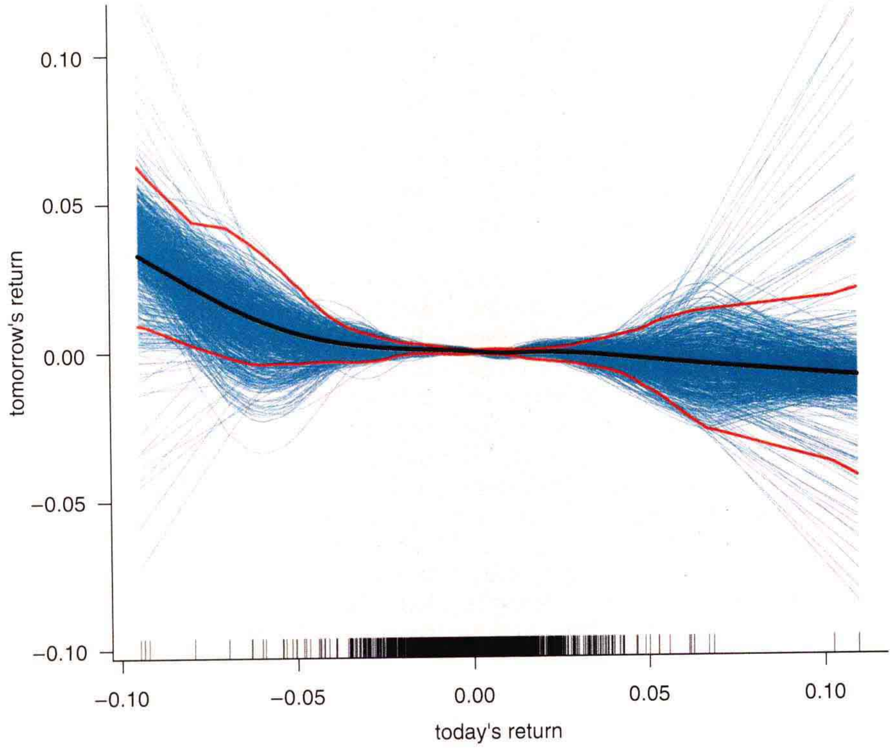



**Figure 5. The same spline fit from the previous figure *(black line)* is combined with 800 splines fit to bootstrapped resamples of the data *(blue curves)* and the resulting 95 percent confidence limits for the true regression curve *(red lines)*.**

#### Bootstrapping

The bootstrap approximates the sampling distribution, with three sources of approximation error. First there's *simulation error*, using finitely many replications to stand for the full sampling distribution. Clever simulation design can shrink this, but brute force—just using enough replications—can also make it arbitrarily small. Second, there's *statistical error:* The sampling distribution of the bootstrap reestimates under our fitted model is not exactly the same as the sampling distribution of estimates under the true data-generating process. The sampling distribution changes with the parameters, and our initial fit is not completely accurate. But it often turns out that distribution of estimates *around* the truth is more nearly invariant than the distribution of estimates themselves, so subtracting the initial estimate from the bootstrapped values helps reduce the statistical error; there are many subtler tricks to the same end. The final source of error in bootstrapping is *specification error:* The data source doesn't exactly follow our model at all. Simulating the model then never quite matches the actual sampling distribution.

Here Efron had a second brilliant idea, which is to address specification error by replacing simulation from the model with resampling from the data. After all, our initial collection of data gives us a lot of information about the relative probabilities of different values, and in certain senses this "empirical distribution" is actually the least prejudiced estimate possible of the underlying distribution—anything else imposes biases or preconceptions, which are possibly accurate but also potentially misleading. We could estimate $q_{0.01}$ directly from the empirical distribution, without the mediation of the Gaussian model. Efron's "nonparametric bootstrap" treats the original data set as a complete population and draws a new, simulated sample from it, picking each observation with equal probability (allowing repeated values) and then re-running the estimation (as shown in Figure 2).

This new method matters here because the Gaussian model is inaccurate; the true distribution is more sharply peaked around zero and has substantially more large-magnitude returns, in both directions, than the Gaussian *(see the left graph in Figure 3)*. For the empirical distribution, $q_{0.01} = -0.0392$. This may seem close to our previous point estimate of –0.0326, but it's well beyond the confidence interval, and under the Gaussian model we should see values that negative only 0.25 percent of the