Cassini's last view of SATURN'S RINGS

How our brains judge MISINFORMATION

Security pitfalls with DIGITAL MONEY

AMERICAN Scientist

November–December 2017

www.americanscientist.org

Social Jet Lag

How the clash between biological clocks and modern lifestyle harms public health

SIGMA XI

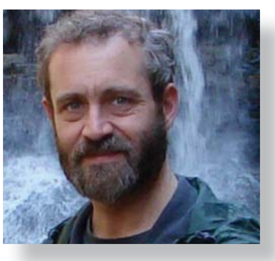
THE SCIENTIFIC RESEARCH HONOR SOCIETY

Symposium on Atmospheric Chemistry, Climate, and Health

November 10, 2017 • Raleigh Convention Center
Raleigh, North Carolina • #SigmaXimtg

IMAGE BY NASA

Featured Speakers



David Archer, PhD
The University of Chicago

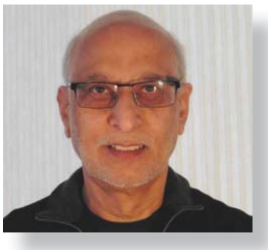
Humanity and Global Warming: Views from the Carbon Cycle

Barbara Finlayson-Pitts, PhD
University of California, Irvine

The Past is Prologue: Lessons Learned for Science-Based Policies to Address Air Quality and Climate

C. Arden Pope III, PhD
Brigham Young University

Air Pollution and Health: Scientific and Public Policy Controversies



A. R. Ravishankara, PhD
Colorado State University

The Ozone Hole: From Discovery to Recovery

Jeffrey Shaman, PhD
Columbia University

Simulation and Forecast of Infectious Disease: Environmental Determinants and Transmission Dynamics

Followed by the Student Research Conference on November 11 • meeting@sigmaxi.org

sigmaxi.org/fallsymposium-src

SIGMA XI
THE SCIENTIFIC RESEARCH HONOR SOCIETY

AMERICAN
Scientist

Volume 105 • Number 6 • November–December 2017

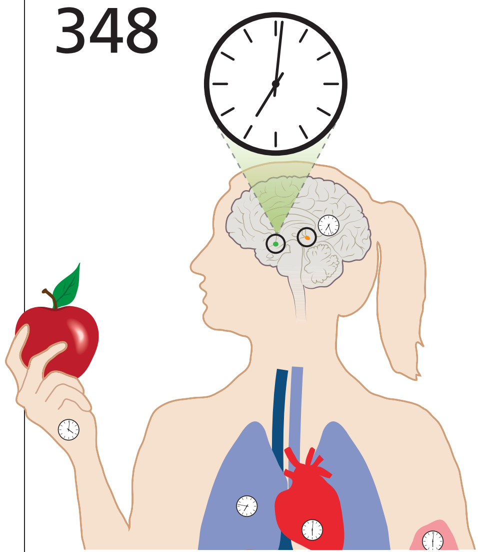
Departments

Scientists' Nightstand

From Sigma Xi

Feature Articles

348



364

The Cover

Technology has extended the times of day when people can be active and has allowed for connectivity at any time of day. A growing body of research shows that the biological clock is affected by light exposure, especially at certain times of day, which in turn can affect all sorts of biological processes necessary for health, including sleep, focus, and metabolism. More people seem to be experiencing social jet lag, when our body rhythms are out of sync with the day-night cycle. This lack of synchrony can have serious public health consequences; for example, it has been linked to sleep disorders, obesity, and mental disorders. In "Adapting Your Body Clock to a 24-Hour Society" (pages 348–355), circadian biologists Alexis Webb and Erik Herzog explain how the biological clock sets and keeps time, and what solutions are under study to help people sync up with healthy daily routines for sleeping and eating. (Cover illustration by Michael Morgenstern.)

FROM THE EDITORS

Modern Challenges

A decade ago, the robotic spacecraft Cassini had already become a huge success. It had released its Huygens probe onto Saturn's moon Titan, and it was returning spectacular imagery of the giant planet itself, as well as its other moons. At that time, we asked Carolyn Porco, leader of the imaging team for the Cassini-Huygens mission and director of the Cassini Imaging Central Laboratory for Operations, to fill in *American Scientist* readers on the mission results so far. Her resulting essay in the July–August 2007 issue showcased the early results from the first 1,000 days of the mission.

In this issue, we are returning to Cassini as its mission recently came to a preplanned end. To preclude Cassini from crashing into one of Saturn's moons as it ran out of propellant and potentially spreading terrestrial contamination there, the spacecraft was instead deliberately plunged into Saturn. But before its final encounter with the planet, Cassini spent much of the final year of its mission taking a detailed look at Saturn's rings. As Matthew S. Tiscareno explains in "Cassini and the Rings of Saturn" (pages 356–363), this concluding year has been practically a whole new mission for the spacecraft, with some spectacular results.

Space exploration uses remarkable technology, but the devices we use daily are also pretty amazing, even if we're not paying attention to how they work. This issue tackles from multiple angles topics related to technology and how it affects us. In our cover story, Alexis Webb and Erik Herzog look at the role that late-night viewing of backlit devices, such as cell phones and tablets, is having on our biological clocks, and our health, as a result. In "Adapting Your Body Clock to a 24-Hour Society" (pages 348–355), the authors look at research that might help reset our biological clocks to be more in tune with our modern style of living.

Mobile devices offer us the convenience of information at our fingertips, but are they also a conduit for making us more susceptible to falsehoods? In "The Persistence and Peril of Misinformation," Brian G. Southwell, Emily A. Thorson, and Laura Sheble review what is known about how people evaluate information and discuss how knowledge of these processes could be used to battle hoaxes and fake news.

Two articles in this issue discuss machine learning and its supporting algorithms, which are increasingly framing our lives whether we are aware of it or not. In the Spotlight section (pages 326–327), Simson L. Garfinkel reports on a recent panel discussing proprietary algorithms and who should have access to how they work. Garfinkel points out that these types of algorithms not only can control what music you hear or what news you see on social media, but also are increasingly used to screen which job applicants to interview or determine how long a jail sentence should be—and there is currently little recourse for those affected to dispute the results from the software.

One specific use of machine-learning algorithms is what's called *predictive policing,* where law enforcement agencies use software to help determine everything from where officers patrol to which members of the public might be more likely to be a victim of a crime. In the Scientists' Nightstand section (pages 376–380), an excerpt of Andrew Guthrie Ferguson's latest book, *The Rise of Big Data Policing,* digs into such programs, and why efforts in New Orleans were successful when similar programs in other cities didn't have such dramatic results.

As Garfinkel points out, going back to a world where nothing is filtered isn't really an option; we would all end up buried in spam and cat videos. Science is helping society find the balance between technology that is useful and applications that, maybe, step too far for human comfort.

—*Fenella Saunders* (@FenellaSaunders)

AMERICAN Scientist
www.americanscientist.org

VOLUME 105, NUMBER 6

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

Sigma Xi, The Scientific Research Honor Society is a society of scientists and engineers, founded in 1886 to recognize scientific achievement. A diverse organization of members and chapters, the Society fosters interaction among science, technology, and society; encourages appreciation and support of original work in science and technology; and promotes ethics and excellence in scientific and engineering research.

Printed in USA

Alone in the Universe?

To the Editors:

I found Howard A. Smith's article "Questioning Copernican Mediocrity" (July–August) interesting, especially the discussion of the radius within which we might reasonably hope to communicate with alien life forms. His article also reminded me why I've long thought that the last term in the Drake Equation, which famously estimates the number of intelligent extraterrestrial civilizations in our galaxy, is the most significant in understanding the Fermi Paradox.

Among the many factors in the equation that Frank Drake considered in 1961, the length of time that a technological civilization releases detectable signals into space now seems the weakest link. We have been releasing such radio signals for barely more than 100 years. It seems possible that within another 100 years there will be no one left on Earth sending such signals. Our exploding population, rapid exhaustion of the necessary resources to support a technological civilization, already visible destruction of a climate and environment that can sustain life on Earth, and threats of nuclear annihilation—all of these suggest that perhaps no technological civilization that evolves anywhere in the universe lasts long enough to broadcast its existence for a sufficient amount of time to be found by another short-lived technological civilization.

I presume that whatever direction evolution might take on other planets will be governed by the same principles that Charles Darwin discovered here on Earth. Like us, such planets' intelligent species (if any) will find it difficult or even impossible to overcome their tendency to overpopulate and overexploit their ecological niche, ultimately leading to their rapid extinction or reduction to a pretechnological state of living. If there is intelligent life on other planets, it is most likely to be either preindustrial or post-technological. In either case, we wouldn't be likely to find it.

If my speculations are true, Smith's plea that we treat our special planet with more respect takes on even greater significance. We may indeed be alone, but in case we are not, perhaps we should try harder to last long enough to be found.

John Cushing
Evergreen State College (Emeritus)
Bend, Oregon

Dr. Smith responds:

I thank Dr. Cushing for his comments. I think that the most uncertain factors in the Drake Equation are the ones estimating the proportion of planets harboring conscious life, because we know the least about them. Regarding the lifetime of civilizations, however, Stephen Hawking famously echoes sentiments similar to Dr. Cushing's when he speaks out against contacting aliens. He simplistically claims that because they will be much more advanced than we are and presumably will have also evolved through Darwinian processes, they will be, to his way of thinking, dangerously violent.

I am more hopeful than either Dr. Cushing or Hawking; after all, evolutionary biologists have pointed out that sentiments such as altruism and gratitude are also the products of evolution. My opinion is that if an advanced civilization is able to last more than a few millennia without destroying itself or its world, it will have learned how to master its negative impulses. Perhaps the biggest benefit we can derive from challenging our cosmic mediocrity is the added impetus to learn this lesson ourselves.

To the Editors:

In his recent feature, Howard A. Smith seems to be asserting that the fact of something being rare means that it is

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 $5.95. Back issues $7.95 per copy for 1st class mailing. U.S. subscriptions: one year $30, two years $54, three years $80. Canadian subscriptions: one year $38; other foreign subscriptions: one year $46. U.S. institutional rate: $75; Canadian $83; other foreign $91. Second-class postage paid at Durham, NC, and additional mailing offices. 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.

ONLINE @ americanscientist.org

What Makes Science Scientific
A degree in science should require students to evaluate their process of producing knowledge.
http://bit.ly/2wgNtgy

Studying the Light of the Eclipse
Solar eclipses offer rare opportunities to better understand both the Sun's corona and the Earth's atmosphere.
http://bit.ly/2fw8yQZ

Slide Show: Ecology's Colonial Legacy
This set of historical images expands beyond the visuals in "The Colonial Origins of Tropical Field Stations," in the July–August issue.
http://bit.ly/2hrI2J4

Podcast: The Moment of Totality
For thousands of years, people have been computing the timing of eclipses. We're really close. But with a little more data, we could be closer still.
http://bit.ly/2wgO4ii

Check out *AmSci Blogs*
http://www.amsci.org/blog/

Find *American Scientist* on Facebook
facebook.com/AmericanScientist

Follow us on Twitter
twitter.com/AmSciMag

Follow us on Google Plus
plus.google.com/+Americanscientist Org/about

Join us on LinkedIn
https://www.linkedin.com/company/american-scientist

Read *American Scientist* using the iPad app
Available through Apple's App Store

necessarily special. This error has been decried by numerous scientists, including Richard Feynman, who mocked the notion in his unique way: "You know, the most amazing thing happened to me tonight.... I saw a car with the license plate ARW 357. Can you imagine? Of all the millions of license plates in the state, what was the chance that I would see that particular one tonight? Amazing!"

If the universe averages one planet with intelligent life per galaxy, Earthlings would never detect other intelligent beings but would nonetheless not be special in any way. In other words, one of 100 billion examples of intelligent life in the universe is hardly special. Frankly, the notion that humans are special is still theological.

John Foster
Columbus, Ohio

Dr. Smith responds:
The "ad hoc fallacy" that Feynman illustrates is not applicable to this situation. "The state" cited in the example has millions of other license plates, and each one is indeed equally "amazing" a priori. But the universe has no other intelligent life we have seen (nor do we know whether any exists). We are an example of one. If a person responds that we are just a collection of molecules in a universe filled with lots of planets and molecules, then it is important to include the "priors" of the case before calculating the statistics, as one would in a detailed Bayesian analysis: If one knows ahead of time that the very particular license number "ARW 357" entitles one to win $1 trillion, then one must admit that all other license plates are worthless. We don't know whether the collection of intelligent molecules here on Earth is so very particular, but it might be.

We moderns are becoming accustomed to living with the realization that some things are fundamentally not knowable: quantum mechanical predictions, the outcome of deterministic but chaotic systems, and events in the universe beyond our (receding) cosmic horizon are examples.

My article asks the reader to puzzle over three consequences of the limits of knowing (and I talk more about these matters in the 2016 *Zygon* article I cite). (1) Living with uncertainty: From what we do know, for hundreds or thousands of human generations (maybe even longer) we are likely to be the only intelligent beings around, though we probably will not know for sure that we are. (2) Recognizing beliefs: Yes, we might nonetheless prefer to believe we are cosmically average—or the opposite: that we are a cosmically amazing species. But personal beliefs such as these are matters of taste, and it is worth asking ourselves what underlies that belief. Perhaps some philosophical attachment or psychological need prompts us to see ourselves as being either insignificant or significant. Still, we must try to ground our personal beliefs in the scientific facts. (3) Admitting we are rare: While we wait, perhaps forever, to find out whether there are aliens, we meanwhile ought to recognize that in the volume of the universe in which *Homo sapiens* can ever exert influence, which is perhaps thousands of light years in diameter, we are probably unusual and could be unique. Does that make us rare, that is: "a thing not found in large numbers and so of interest or value?" I would say so. The Earth is too. The word "special" Dr. Foster uses is loaded with subjectivity and I try to avoid it. Both Dr. Foster, whom I thank for his question, and I are probably average examples of *Homo sapiens*, but I daresay we are both special. It does not follow that if we are special and intelligence is rare, a theological explanation is required.

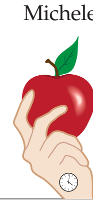
Illustration Credits

Infographic
Pages 328–329 Eleanor Lutz

Perspective
Page 345, 346 Tom Dunne

Adapting Your Body Clock to a 24-Hour Society
Pages 350, 352 Barbara Aulicino

The Persistence and Peril of Misinformation
Page 373 Michele Rosenthal



How to Write to *American Scientist*
Brief letters commenting on articles appearing in the magazine are welcomed. The editors reserve the right to edit submissions. Please include an email address if possible. Address: Letters to the Editors, P.O. Box 13975, Research Triangle Park, NC 27709 or editors@amscionline.org.

A Peek at Proprietary Algorithms

Software increasingly determines what people see online or even how long they might spend in jail, but few can access how such programs work.

Let's say you start using a music streaming service online. You pick a few songs you like, and the site creates a playlist of similar music for you based on that input, using its huge database of information classifying songs into different genres, along with data about other users' likes and dislikes. Each time the service plays a song, you rate it, and the service refines its statistical model of your preferences. This is *machine learning* at work, and it's one of computing's fundamental tools for helping people make choices in the face of information overload, no matter whether the task involves choosing songs in a catalog, classifying the overflow of posts and photographs on social media, filtering wanted email from spam, or even helping businesses sort the flood of résumés received in response to an advertisement.

To complete such tasks, software follows *algorithms,* which have been part of computing since the dawn of computer science: Ada Lovelace wrote the first one back in 1840. More recently, however, the term *algorithm* has evolved to mean "a self-contained step-by-step set of operations that computers and other 'smart' devices carry out to perform calculation, data processing, and automated reasoning tasks," according to the public policy council for the Association for Computing Machinery (ACM), the world's largest educational and scientific computing society.

Principles for Algorithmic Transparency and Accountability

1. **Awareness:** Owners, designers, builders, users, and other stakeholders of analytic systems should be aware of the possible biases involved in their design, implementation, and use and the potential harm that biases can cause to individuals and society.
2. **Access and Redress:** Regulators should encourage the adoption of mechanisms that enable questioning and redress for individuals and groups that are adversely affected by algorithmically informed decisions.
3. **Accountability:** Institutions should be held responsible for decisions made by the algorithms that they use, even if it is not feasible to explain in detail how the algorithms produce their results.
4. **Explanation:** Systems and institutions that use algorithmic decision-making are encouraged to produce explanations regarding both the procedures followed by the algorithm and the specific decisions that are made. This is particularly important in public policy contexts.
5. **Data Provenance:** A description of the way in which the training data was collected should be maintained by the builders of the algorithms, accompanied by an exploration of the potential biases induced by the human or algorithmic data-gathering process. Public scrutiny of the data provides maximum opportunity for corrections. However, concerns over privacy, protecting trade secrets, or revelation of analytics that might allow malicious actors to game the system can justify restricting access to qualified and authorized individuals.
6. **Auditability:** Models, algorithms, data, and decisions should be recorded so that they can be audited in cases where harm is suspected.
7. **Validation and Testing:** Institutions should use rigorous methods to validate their models and document those methods and results. In particular, they should routinely perform tests to assess and determine whether the model generates discriminatory harm. Institutions are encouraged to make the results of such tests public.

Source: https://www.acm.org/binaries/content/assets/public-policy/2017_usacm_statement_algorithms.pdf

Computer programs running advanced machine-learning algorithms are now everywhere, and are making decisions about people for which we often have little or no recourse. Go shopping online, and algorithms decide which products you see and what special deals you might be offered—and which are withheld. Apply to a university, and an algorithm may determine if a human even sees your admissions essays. In New York City and Santa Cruz, California, among other places, predictive policing algorithms may determine how often the police patrol a neighborhood. And if you commit a crime in Wisconsin, an algorithm might determine how much time you spend in prison.

These uses of algorithms all share several important characteristics. Like the music service scenario, they all appear to be based on advanced statistical machine-learning techniques, developed over the past two decades, which use large amounts of training data to create models that can then make predictions or perform classifications. In addition, these algorithms are proprietary: So far, the public (or regulatory officials acting on the public's behalf) has no right to inspect algorithmic implementations or the training data, even when the algorithms are used for public purposes. And finally, in many cases, the algorithm is being used by an organization with institutional power, to make decisions about people who frequently have no right to appeal if the algorithm makes a mistake.

On September 14, the ACM's U.S. Public Policy Council (USACM) held a panel discussion at the National Press Club in Washington, D.C., to discuss the effect of algorithmic decision-making on society and the technical underpinnings of algorithmic models. I moderated the panel, and I cochair the USACM's working group on Algorithmic Transparency and Accountability (ATA).

Jeanna Matthews, an associate professor of computer science at Clarkson University in New York and also a cochair of the USACM group, opened the panel's discussion by introducing the "Principles for Algorithmic Transparency and Accountability" (*see box*) that USACM issued earlier this year. Modeled on the various principles of fair information practice developed by privacy regulators worldwide over the past 40 years, these principles are designed to provide a framework for thinking about the challenges posed by the increasing use of algorithms in our society. The seven principles address many of the concerns that have been voiced about the growing use of algorithms, without placing limits on the possible beneficial uses of algorithms currently being explored.

Nicholas Diakopoulos, an assistant professor at Northwestern University, spoke about his website, Algorithm Tips (http://algorithmtips.org/), which tracks the growing use of algorithms by the U.S. government. The website aims to promote accountability and transparency, and specifically to assist investigative journalists in writing about these topics. The database contained more than 150 algorithms in September.

Dan Rubins, one of the two industry representatives on the panel, is cofounder of Legal Robot, an artificial intelligence startup that is using algorithms to analyze case law and contracts. Rubins said that his company is using the ATA principles as the framework for the transparency report that Legal Robot is publishing on its website. It is possible for companies to be open about how their technology works without giving up commercial advantage, Rubins said, but doing so requires focusing on where the company's added value happens to be. For example, he said, it can be difficult to move algorithms from one domain to another, because the training data may not be representative. One of Legal Robot's advantages, he says, is the way it has collected and labeled its training data, instead of the actual algorithms. For example, many programs that were developed to process English fail when presented with legal documents, because the vocabulary and style of language usage is so different.

Algorithmic Bias

Geoff Cohen, a vice president at Stroz Friedberg, an Aon company, has performed many computer forensic investigations of algorithms used by companies. He said that in recent years, U.S. patent law has weakened while legislation regarding trade secrets has become more powerful, and many companies have reacted by trying to keep their algorithms more secret. But inspection is possible, Cohen said, adding that inspection typically happens when an algorithmic-based company is being considered for acquisition or is the subject of a lawsuit. In the future, government regulators might instead perform such inspections.

Finally, Ansgar Koene, chair of a Standard for Algorithm Bias Considerations working group under the auspices of the Institute for Electrical and Electronics Engineers (IEEE), discussed the efforts under way to create an international standard that organizations could use to understand and eliminate unintentional algorithmic bias in their offerings. For instance, there has been increased attention to cases in which algorithms have developed "racist" classifying tendencies, some of which could be tied to underlying data sets that contained unknown biases. In other cases, algorithms have been shown to have unintended consequences, such as causing users to see on social media only news that conforms to their own political leanings, or repeatedly sending police to the same neighborhood because there were no reports of crime in others.

There is certainly a growing interest among computer scientists and academics in developing approaches for algorithmic transparency and accountability. Since 2014, for example, a workshop called Fairness, Accuracy, and Transparency in Machine Learning has explored this topic with significant mathematical rigor. Increasingly, academics are reaching out to corporations and policy makers, with the hope of establishing norms so society can benefit from algorithms while preventing individuals from being inappropriately harmed by them.

With the rise of the information economy, it would be unworkable to go back to a world in which nothing is filtered. Without algorithmic assistance, social media, Internet search engines, and even email would all be unusable. Companies are now exploring ways of using algorithms to filter not just spam, but also fake news and even hate speech. Just how far such filtering goes, and whether it is possible to turn it off, needs to be the subject of public debate. —*Simson L. Garfinkel*

Simson L. Garfinkel is an adjunct faculty member at George Mason University, where he teach digital forensics. He is the author of Database Nation: The Death of Privacy in the 21st Century *(O'Reilly, 2000). Internet: https://simson.net/*

Infographic

Living with Fire:

Surviving the forest wildfires of California

Giant Sequoia • *Sequoia-dendron giganteum*

The giant sequoia is the quintessential example of a fire-adapted plant species. Among the largest trees in the world (reaching heights of 75 meters), giant sequoias grow particularly well in the mineral soil exposed by wildfire. Giant sequoias also have thick bark (commonly 15 to 20 centimeters thick, but up to half a meter thick in mature trees), which protects them from low-severity fires. Because younger trees are very susceptible to fire, giant sequoia trees grow quickly as another adaptation to fire-prone environments.

Fireweed • *Chamerion angustifolium*

After a wildfire, fireweed is one of the first plants to regrow in burned areas. Like many other plants that grow in openings created by fire, fireweed has seeds that are dispersed by wind, and a plant can produce up to 80,000 seeds a year. This method allows the plant to spread to newly burned areas, even if the parent plant is far away. Fireweed is an important post-disturbance species because it can recycle nutrients from burned soil. Indeed, it was one of the most abundant species to recolonize Mount St. Helens after its 1980 volcanic eruption. Fireweed has also been used to restore lands damaged from coal mining or oil spills.

California Black Oak • *Quercus kelloggii*

California black oak trees are quick to regrow from the stumps or snags left behind after a fire. These trees have extensive root systems that support this resprouting process. Sprouting stems can take advantage of the parent tree's well-developed root system. Even one-year-old seedlings, which are 5 to 15 centimeters tall, have long taproots that extend up to a meter below the ground. Historically, California black oak trees were an important source of food for Native American communities. Native Americans regularly burned these areas to remove other species and to promote the growth of California black oak woodlands.

California wildfires can cause devastating damage to property and human lives. But for plant species that have lived for thousands of years in the fire-prone Sierra Nevada, forest fires are an expected part of the natural environment. These species have unique adaptations to withstand fire—and some even depend on fire to survive. This infographic introduces six different California plant species, all with different adaptations to fire. (Paper plant sculptures, photography, and text by Eleanor Lutz, tabletopwhale.com. Source: USDA FEIS.)

Ponderosa Pine • *Pinus ponderosa*

Ponderosa pine has thick bark plates that protect the tree's interior and help the tree survive most fires. The very moist cores of branches also help dissipate fire heat. Mature trees can withstand high-severity fires, and even saplings can survive low-intensity fires. These trees also drop their lower branches as they grow. With less fuel near the ground, there's less chance that fire will spread from the ground up into the living tree canopy, destroying the pine needles required for photosynthesis. Finally, ponderosa pine trees have deep roots that are less susceptible to fire damage.

California Lilac • *Ceanothus cordulatus*

California lilac seeds are covered by an extremely hard seed coat. Wildfires crack open this coating, which then allows water to reach the seed and begin the sprouting process. Seedlings seldom sprout if there is no fire to begin the process of germination. Indeed, California lilac seeds can lie dormant for hundreds of years until the next wildfire, although the areas where California lilac grows naturally experience fire every one or two decades. Buried seeds can withstand fire temperatures of 105 degrees Celsius. Some forests are estimated to harbor about 2 million of these seeds per acre of soil.

Knobcone Pine • *Pinus attenuata*

Knobcone pine trees require fire to reproduce—their cones will not open without a fire temperature of about 203 degrees Celsius, which vaporizes the resin sealing the cones. In contrast to ponderosa pine, knobcone pine has a branch structure that actually promotes the spread of fire farther up the tree. The open and low branch pattern helps direct fire to the cones in the crown of the tree. Knobcone pines reach maturity and produce cones quickly (at the age of 10 to 12 years), an important characteristic in areas with frequent fire. The pine cones do not fall from the tree without fire, which ensures that a large number of accumulated seeds can be released during the next wildfire.

For an animated version of this infographic, go to at www.amsci.org.

First Person: Sally C. Seidel

Sally C. Seidel is the faculty member of the University of New Mexico's Collider Physics Group and a collaborator on the ATLAS experiment at the Large Hadron Collider housed at CERN, the European Organization for Nuclear Research, in Switzerland. The ATLAS experiment focuses on activity involving massive particles and was part of the effort to discover of the Higgs boson in 2012. Seidel's research focuses on improving the understanding of heavy quark bound states, which in turn can help to elucidate the strong force, one of the four fundamental forces of nature. Her group also works on developing particle-tracking detectors. Seidel, one of Sigma Xi's Distinguished Lecturers, spoke with editor-in-chief Fenella Saunders about her research.

How many particles do physicists think are out there?
If we are talking about the most fundamental particles, there are only about a dozen, the quarks and the leptons and so forth. Often particle physicists speak somewhat casually about particles, and we include what we would think of as bound states, combinations of quarks and so forth. In that case, there are hundreds of them.

What is the difference between fundamental particles and something like bound states, or particles that are made from other particles?
The fundamental particles, as far as we know, are typically pointlike objects. They might have mass, but sometimes they don't have any physical dimension. The particles that are like protons are physically extended, and when we collide them in colliders like the Large Hadron Collider, we can determine that they actually have structure.

How can a fundamental particle have mass, but no volume? Can you explain how that works?
I cannot. It's a mystery. That's something we're still working on.

Is there a defining line between a particle and a force? For instance, because there are certain particles that are responsible for forces, where does that defining line happen?
It is the case that some of the particles seem in nature to be constituents of physical bodies or things, like the proton, and some of them seem to have the role of transmitting forces. We are able to relate the photon, which is the particle of light, as also being the transmitter of the electromagnetic force. There are particles like that for the strong force, the weak force, and so forth.

What are the strong force and the weak force?
The strong force is the strongest of the fundamental forces, and it's the one that binds the quarks into a nucleus. It's the one that is able to overcome the repulsion that the quarks that have the same electromagnetic charge otherwise experience and which would otherwise cause the nucleus to disintegrate. The weak force is something entirely different. It's one of the forces that allows quarks and leptons to change their identity, and it's at the basis of radioactive decay.

What do you see as the purpose of a particle collider?
The purpose of a particle collider is to collect a lot of energy at a very small point, and we do that by colliding two particles that are carrying a lot of energy. Mostly these particles are carrying that energy in the form of their momentum, and some of it in the form of their mass. Typically the moment of collision creates a little microscopic fireball, which evolves according to the laws of nature. Then, we look and see what nature produces when you give it that much energy. In that sense, we are replicating a tiny Big Bang sort of event.

What particles is your group studying?
Members of my group are part of the ATLAS Experiment at the Large Hadron Collider. There are four general-purpose experiments at the LHC, and ATLAS is one of them. Within the collaboration, there are people who are studying the Higgs boson, properties of the top quark, people who are searching for new exotic particles that have properties that might help us see some unifying principle in nature. My group is interested in searching for new particles, new principles, and new properties, and we, particularly historically, have used the heavier quarks as signatures of those. The bottom quark, in particular, is the second heaviest quark. We typically look for signatures that involve a bottom quark for a number of reasons—for example, because its coupling to the Higgs boson is the strongest. Typically, it's one of the easier signatures to reconstruct.

What did you find in your study that combined two quarks?
The University of New Mexico group contributed an analysis published by the ATLAS experiment, that involved the discovery of the first excited state of the B_C meson, which is a bound state of a bottom and a charm quark.

This study described the excited state as being almost like one of the quarks was orbiting the other, like an electron in an atom. Is that accurate?
That's a reasonable model. One could think of the bottom quark as being so heavy that the motion of the two particles relative to each other in the bound state might be approximately nonrelativistic.

If you're talking about an excited state, is it analogous to the orbitals of an electron being moved up to an excited state?
Yes. Our initial motivating interest was to learn what bound states can be sustained in nature, which may give us an opportunity to understand better how the strong force works and what the binding principles of the strong force are. We're experimentalists, so we made the measurement. We're not theorists, but in principle theorists who build quantum chromodynam-

American Scientist Previous Page | Contents | Zoom in | Zoom out | Front Cover | Search Issue | Next Page

ics models might use the information of the mass of this particle combined with other bound states to infer some features of the strong interaction.

Many scientists collaborate on work at the Large Hadron Collider. What are the advantages of working in large collaborations?

There are four collaborations on the Large Hadron Collider. The ATLAS collaboration, the one that I work on, has 3,000 people. The members of the collaboration form subcollaborations so that people who have related expertise, people who naturally will support, enhance, and stimulate the work of one another, work together. People who do, for example, particle tracking, as the University of New Mexico does, form smaller groups, and then within particle tracking there are people who are interested in smaller and smaller, more and more specific aspects of that topic, until in the end the number of people who were talking to each other about a particular question may be only five people or may be only 15 or 30 people, despite the fact that there are 3,000 of us in the entire collaboration.

Someone might wake up in the morning and connect to a meeting and talk to 20 people about a technology that's in development. Then they might have a video conference with a different group of 15 or 20 people and talk about an analysis that's searching for a new particle, and then maybe later in the day talk with yet some other small group, maybe 15 or so people about some other new development, maybe a new method for data storage, data reconstruction, electronics, something like that.

In the end, the fact that there are 3,000 people is barely visible to a typical collaborator. We have very small and tightly networked communities that we work within. That seems to have evolved naturally. That's not imposed by the collaboration. It's a natural way that scientists work. It's good for the students, because they find themselves working in a community that is largely without any hierarchy of age or experience. There could be people who are graduate students who are communicating with people who are very experienced, distinguished members of the community, all talking about the same problem, same questions, and everyone growing together, moving forward together toward a solution.

What is it like to work with such an international team?

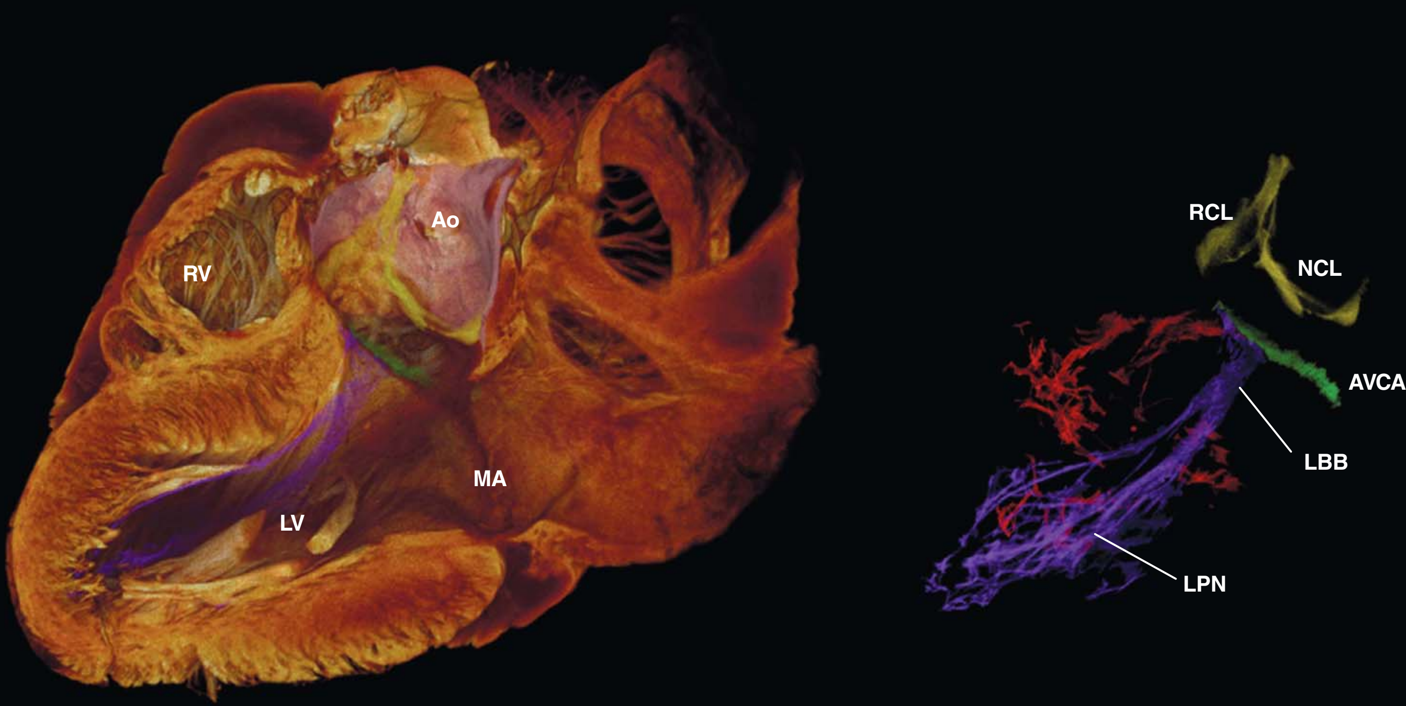
The collaboration draws upon expertise almost without consideration of national borders. A typical meeting that I might have with collaborators does not require me usually to travel any place other than as far as my laptop, where I can run some video conferencing software, and I may connect and meet with 20 people who are in 20 different locations, and I may not even know where they are, what their nationality is, or where they're physically connecting from. We're just working together on the science.

What advice would you give to students working on a PhD?

I think a student who has decided to attempt a PhD should select the single topic that they think is most important and most interesting to them. If you're working every day on something that is the most important thing to you, it's the easiest thing in the world to do. You don't feel conflicted about it, and you feel encouraged by the worthwhileness of the topic.

American Scientist Previous Page | Contents | Zoom in | Zoom out | Front Cover | Search Issue | Next Page

Sightings



Seeing the Heart's Power

The first 3D imaging of the intricate cardiac conduction system provides new detail for researchers and surgeons.

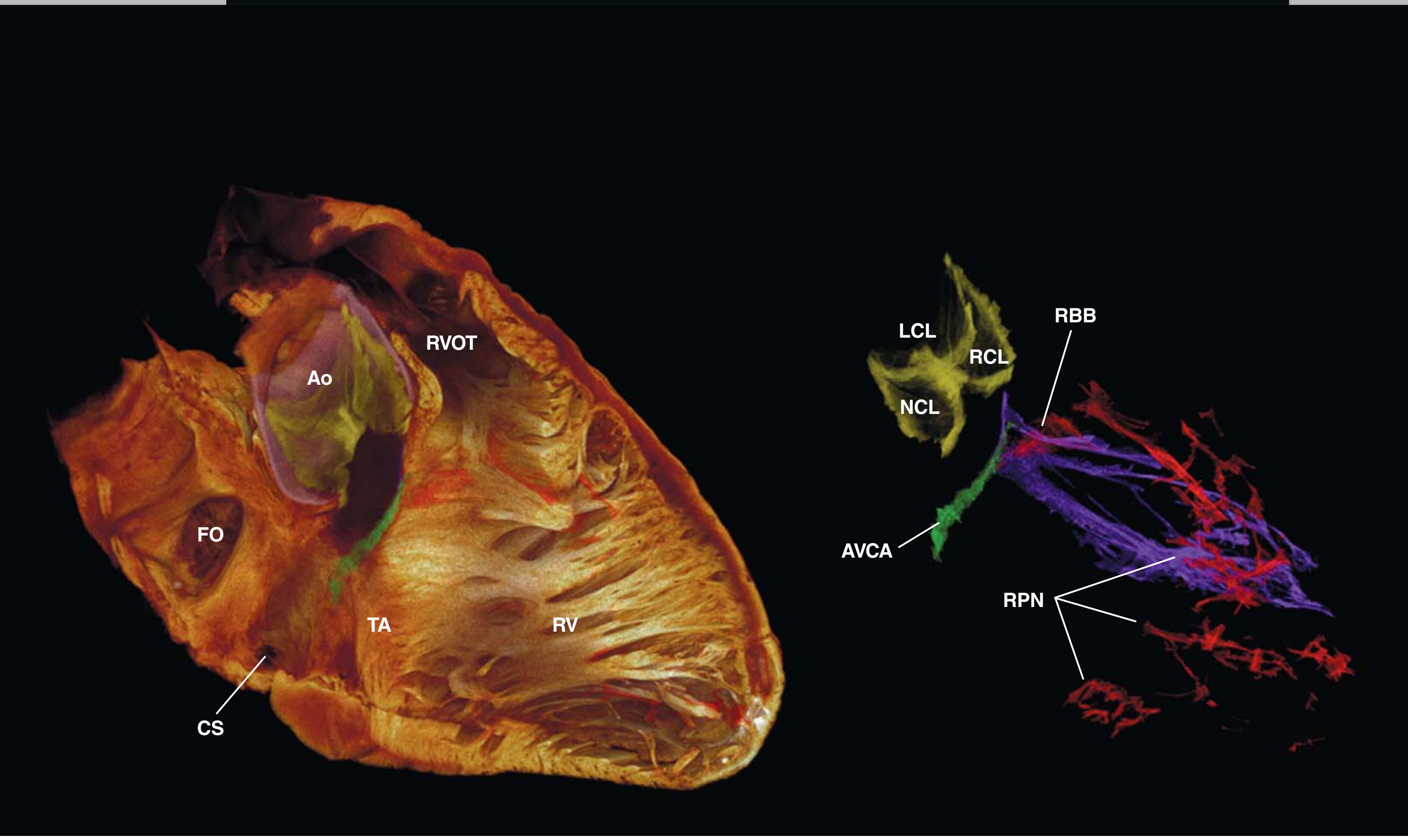
Even in the fast-paced world of digital imaging, it took seven years for researchers to develop the technique to make these three-dimensional (3D) images (*above*). Manchester University's Halina Dobrzynski, Liverpool John Moores University's Jonathan Jarvis in the United Kingdom, and a team of international researchers undertook the effort because they thought high-resolution 3D imaging of an intact heart would provide new insights into how the organ works.

"We also pretty soon realized that in the 3D image," says Jarvis, "we could tell the difference between the working myocardium—that's the cardiomyocytes that produce the pumping action—and the cardiac conduction system, which is the electrical side of the heart." The conduction system is made from specialized cardiomyocytes. Segmented conduction networks of these cells are represented by different colors and overlaid on semitransparent heart images.

Dobrzynski says the group hopes the technique will also help heart surgeons better visualize the locations of different components of the heart's conduction system. The technique is not intended to replace any tools and imaging guides. Replacing thin-slicing and tissue-staining techniques, which have been in use for more than a century with light microscopy, is also not an option, Jarvis says. That's because 3D imaging of an intact heart provides no interior access to the tissue to label particular proteins, which is the key to understanding what each population of heart cells is doing.

But results from the new technique already are improving upon previous research. For example, tracing the cardiac conduction system previously required researchers to follow the route layer-by-layer through perhaps hundreds of thin slices of cardiac tissue from the same heart. With the team's 3D images has come the understanding that "commonly accepted anatomical representations [of the cardiac conduction system] are oversimplified," the researchers write in the August 2017 issue of *Scientific Reports*.

Although the imaging technique is new, the technology behind it is not: Microcomputed tomography (micro-CT) scans have been in use for decades. Initially, micro-CT scanners could only image the human body's harder tissues, such as bone and teeth, because those tissues better absorb the x-rays that give the technology its imaging power.



Robert S. Stephenson et al.

In each pair of images, colored diagrams represent the segmented network of the cardiac conduction system, displayed both separately and overlaying a high-resolution microcomputed tomography scan. The intact heart has been sliced virtually in half and is presented semitransparently. The right half (*above*) has the thinner right ventricle (RV) for pumping blood only to the lungs. The left half (*far left*) has the thicker left ventricle (LV) for pumping blood to the rest of the body. Other parts of the cardiac conduction system and visible anatomy include the aortic root (Ao), atrioventricular conduction axis (AVCA), coronary sinus (CS), fossa ovale (FO), left/right bundle branch (LBB/RBB), hinge of left/right coronary leaflet (LCL/RCL), left/right Purkinje network (LPN/RPN), mitral annulus (MA), hinge of noncoronary leaflet (NCL), right ventricular outflow tract (RVOT), and tricuspid annulus (TA).

Imaging softer tissues has been more challenging, although many labs around the world have developed multiple techniques for doing so by injecting soft tissue with contrast agents that better absorb x-rays. That means those labs already have access to everything they need to make such 3D images of intact hearts. "I think we were successful because we knew how to handle and prepare the tissue very carefully," says Jarvis, a muscle physiologist.

Still another challenge in this kind of research is obtaining access to normal, intact heart tissue, which is usually reserved for heart-transplant patients. With collaborators at the University of Minnesota's Visible Heart Lab, the team gained access to human hearts that had been intended for use in transplantation, but which had not been transplanted for varying reasons. The team's experimentation with actual human hearts was minimal, though, because they only needed to scale up from having perfected the imaging technique using the hearts of smaller mammals, including rats. "So you do have to modify your technique a little bit," says Jarvis, such as by allowing a longer period of time for the contrast agent to diffuse into the human heart tissue. Theirs was an iodine-based agent, and that diffusion took two weeks.

That lengthy diffusion time, of course, makes this particular technique useful only with post-mortem tissue, but so too does the amount of iodine required, because iodine is toxic at such levels. Ideally, imaging techniques could guide surgeons performing real-time operations on patients, during which the cardiac conduction system is currently invisible. "It would be very unwise to say that something is impossible," Jarvis says, "because the progress in medical imaging has been so spectacular over the last 50 years—things which we thought were impossible even 10 years ago are now possible."

With this imaging technique, then, next steps include imaging older and damaged hearts to address whether, for example, the stretching or scarring of the heart muscle that often occurs in heart-disease patients is associated with strains on or interruptions to the cardiac conduction system.

In the meantime, the detail from these kinds of post-mortem scans may be used today for exploring 3D cross-sections virtually, as shown above, or for printing 3D models, both of which can aid medical professionals in visualizing any cross-section of the heart with the visible location of the cardiac conduction system. For the 3D models, Jarvis says, "You have to use a transparent material with some colored material incorporated into the print, and that is fairly challenging, but we have done it."
—*Robert Frederick*

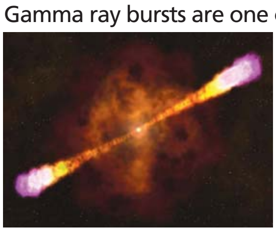
For related video and podcast, see the online version of this article at www.amsci.org.

Briefings

In this roundup, digital features editor Katie L. Burke summarizes notable recent developments in scientific research, selected from reports compiled in the free electronic newsletter *Sigma Xi SmartBrief*. Online: https://www.smartbrief.com/sigmaxi/index.jsp

The Birth of a Black Hole

Astronomers have captured the clearest picture yet of the birth of a new black hole from a catastrophic stellar explosion. Gamma ray bursts are one of the most powerful explosions in the universe and are short-lived (sometimes lasting a few seconds). So it is challenging to react to one in time to study the comprehensive event. Using data from several ground- and space-based telescopes, an international team of astronomers assembled an unprecedentedly detailed description of a recent gamma-ray burst (called GRB 160625B) that showed up in the sky in June 2016. They were able to characterize the emission and jets of the gamma-ray burst to settle some longstanding questions about how the explosion and black-hole formation progress. Immediately after the explosion, the prompt emission comes from electrons bouncing around in a magnetic field. In the following seconds or minutes, powerful magnetic fields surrounding the black hole organize emerging matter and energy into tight jets.


NASA/Goddard Space Flight Center

Troja, E., et al. Significant and variable linear polarization during the prompt optical flash of GRB 160625B. Nature *547:425–427 (July 26)*

Plants Affect Bee Development

A new study clarifies how honeybee larvae develop into worker bees or queens, showing that plants are more important for controlling bee development than had been recognized. It's long been known that food influences honeybees' role in the colony—for example, royal jelly is fed to larvae that develop into queens—but it had not been known that worker bees' development is also influenced by an active ingredient in *bee bread,* their pollen- and honey-based food. That ingredient turns out to be a type of microRNA, a short noncoding RNA that plays a regulatory role, derived from plants. The microRNAs identified affect plant size and color. When present in bee bread fed to larvae, the microRNAs affect genes that delay reproductive development to keep worker bees sterile. Further study of environmental changes in plants that could affect microRNAs could help entomologists understand current problems with honeybee health.

Zhu, K., et al. Plant microRNAs in larval food regulate honeybee caste development. PLOS Genetics *13(8):e1006946 (August 31)*

Fish Fossil Doesn't Fit Family Tree

A fossil dating to about 360 million years ago, when fish evolved into land-dwelling animals, does not fit clearly in the current evolutionary tree, because it possesses features of two lineages thought to be distinct. The new species of lobe-finned fish, called *Hongyu chowi*, was found in northern China. The shoulder girdle and support for the gills of this 1.5-meter-long ambush predator, resemble those of early four-legged animals and their closest fish relatives, indicating that the fossil belongs to an extinct group of fish called *elpistostegids*. But other features resemble a group of fish called *rhizodonts* that were thought to have branched off from a common ancestor before the lineage that gave rise to four-legged animals emerged. Changes in the shoulder girdle and gill bones are central to the story of the evolution of weight-bearing limbs and jaws characteristic of land-dwelling vertebrates. This fish's mix of features means that either different groups of lobe-finned fish, including the rhizodonts, independently evolved this shoulder girdle and gill structure, or the rhizodonts are more closely related to four-legged animals and elpistostegids than was once thought.


Nature Publishing Group

Zhu, M., P. E. Ahlberg, W.-J. Zhao, and L.-T. Jia. A Devonian tetrapod-like fish reveals substantial parallelism in stem tetrapod evolution. Nature Ecology & Evolution *1:1470–1476 (September 4)*

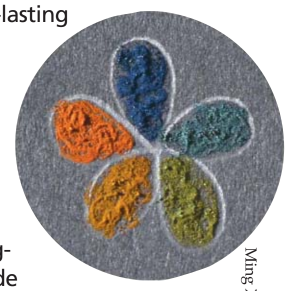
Treatment for Peanut Allergy

A clinical trial with 48 children suggests the first successful long-lasting treatment for peanut allergies. The study participants were given either a daily treatment of a probiotic combined with a peanut protein, or a placebo. The trial released preliminary results in 2013, showing that after one month of treatment 82 percent of children receiving the probiotic could eat peanuts. This follow-up study shows that four years later, 70 percent of the participants in the treatment group can still ingest peanuts without an adverse reaction. The probiotic used was a bacterium, *Lactobacillus rhamnosus*, which is thought to allay some allergic symptoms. Larger clinical studies are now needed to determine long-term safety outcomes, say the researchers at Murdoch Children's Research Institute in Australia. They are also studying whether the peanut protein is necessary in conjunction with the probiotic or whether the latter is effective alone.

Hsiao, K.-C., et al. Long-term clinical and immunological effects of probiotic and peanut oral immunotherapy after treatment cessation: 4-year follow-up of a randomised, double-blind, placebo-controlled trial. The Lancet Child & Adolescent Health *1:97–105 (August 15)*

Long-Lasting Structural Colors

Inspired by certain vivid colors in nature, such as those of some birds and butterflies, polymer scientists have figured how to make long-lasting "structural" colors. Such colors are formed through light's interaction with nanostructures rather than through its absorption by a pigment. Pigments fade when their molecules breakdown, so structural colors can be longer lasting. Oddly enough, the researchers used a pigment to form the structural color by changing the nanostructure of clumps of its molecules. The pigment melanin changes skin color in animals and appears black in its individual nanoparticles. The color was changed from olive to orange to blue by varying the spacing in clusters of the nanoparticles using a silica coating to limit how closely they pack together.


Ming Xiao, University of Akron

Xiao, M., et al. Bioinspired bright noniridescent photonic melanin supraballs. Science Advances *3(9):e1701151 (September 15)*

Computing Science

Bitcoins Maybe; Blockchains Likely

The innovative foundations of the cryptocurrency may outlive the currency itself, as its verification method finds applications everywhere.

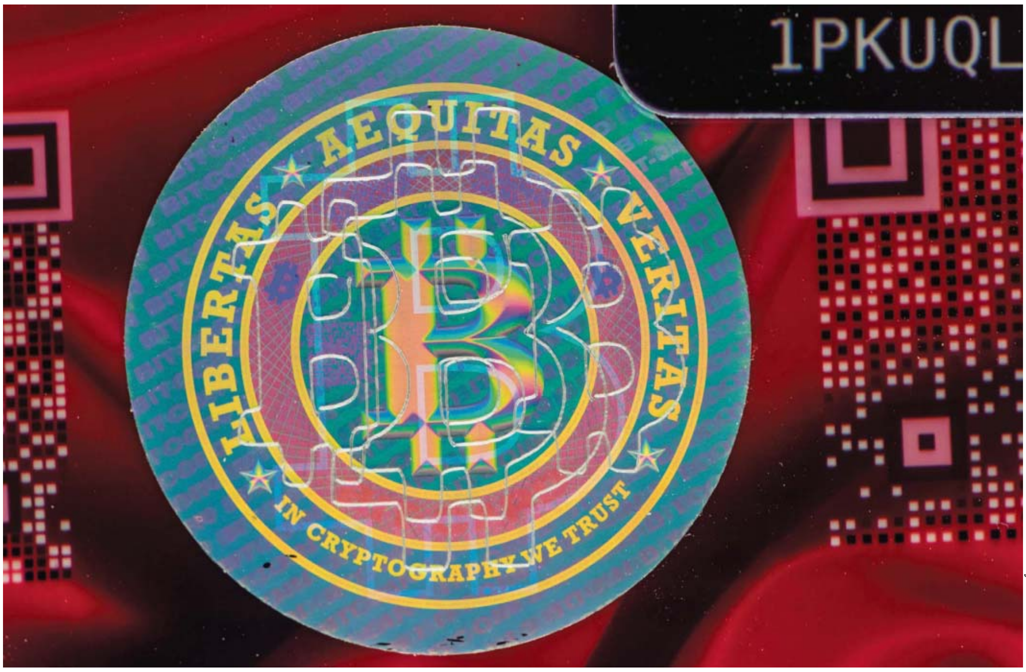
Peter J. Denning and Ted G. Lewis

On January 3, 2009, Satoshi Nakamoto began selling a new form of money and operating an associated support system called the bitcoin cryptocurrency system. The name *bitcoin* (BTC) comes from it being digital money that can be used like coinage, and the *crypto* term indicates that the bits are secured by cryptographic methods. The currency has secretive roots: It is still unknown who Nakamoto was, but it may have been Harold Thomas Finney II, a known cryptologist who created a secured transaction system similar to that used in BTC exchanges, and a developer employed by the PGP (Pretty Good Privacy) Corporation. (PGP is a protocol widely used in personal encoding keys.) Finney was the first recipient of BTCs in 2009. But we may never know for certain who Nakamoto was, because Finney died in 2014.

The term *bitcoin* conjures up an image of a real coin encoded as a bit pattern. The big problem with bit patterns is that the owner can retain a digital copy and try to make another payment with the same coin. This double-spend problem plagues all concepts of virtual money. Nakamoto wanted a system that could ensure that a sum of money can be spent exactly once—no double payments.

Banks already know how to prevent double payments: A digital transaction is completed by the bank, not the consumer. The bank has full control over the database of all accounts and can safely transfer an amount from the payer to the payee accounts. Banking is based on trust in the banking system. Nakamoto rejected the idea of a centralized database, seeking instead a system without a trusted intermediary or a single point of failure. A distributed ledger of accounts that cannot be hacked by breaking into a single database fills this bill.

Thus, Nakamoto defined bitcoins as "a chain of digital signatures." To understand what he means, consider the example of Alice wanting to pay 1 BTC to Bob. Alice can create a certificate that says "Alice pays 1 BTC to Bob, signed Alice." The certificate is signed by Alice's secret digital key, which is paired with her public digital key that anyone can use to validate her ownership. The certificate is a record of a transaction stored in the ledger of all transactions. Alice's ability to pay is determined by the entire history of all transactions that flow to or from her. When the transaction is



Erik Tham / Alamy Stock Photo

Bitcoin users purchase paper certificates or plastic cards for the convenience of storing their encryption keys for transactions. Some of the key-storing cards are elaborate and employ antitampering technologies such as holograms, as in the close-up image of this card. Storage devices in the form of actual coinlike tokens are also available.

Peter J. Denning is Distinguished Professor of Computer Science and Director of the Cebrowski Institute for information innovation at the Naval Postgraduate School in Monterey, California, and editor of ACM Ubiquity. *The author's views expressed here are not necessarily those of his employer or the U.S. Federal Government. Ted G. Lewis is a cofounder of the Center for Homeland Defense and Security at the Naval Postgraduate School. Email for Denning: pjd@nps.edu*

The value of bitcoins in U.S. dollars over the past year shows extreme volatility.

completed, the ledger will show her net worth decreasing by 1 BTC and Bob's increasing by 1 BTC. In other words, cryptographic coins are the value generated by transactions recorded in a ledger. We scan the ledger to find out anyone's net worth at any moment of time.

The idea of calculating value based on a history of transactions is used in many places. A familiar example is the title to a property. It is a certificate that says, "As of this date, Alice is the sole owner of the property described here; signed, notary." When Bob wants to buy the property, the escrow company hires a title search firm to review all the records of previous sales and liens back to the beginning to validate that Alice is indeed the legitimate sole owner. The trail of transactions on the property is called a *chain*. Although the current title is an integral part of the chain, the validity of that title depends on the entire history. Nakamoto's definition of value is parallel to this scenario if the title company is replaced by a computer system and the transactions are all electronic fund transfers.

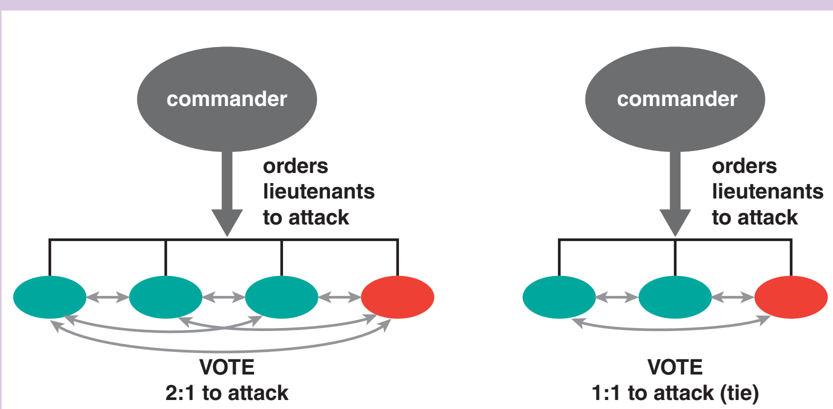
The records of all bitcoin transactions are compiled into blocks of about 4 megabytes in size. The linked list of all blocks going back to the beginning in 2009 is called a *blockchain*. Although the blockchain could be stored in a central database, Nakamoto wanted to distribute it among many computing nodes in the network to remove the risk of single point of failure. Thus, there are many copies of the blockchain in a distributed network of nodes, but the entire chain is protected from tampering by a complex arrangement of links made from digital signatures and *hashes*, mathematical functions that scramble all the bits of a file into a fixed-length code. A form of majority voting is used to decide which blocks are valid and can be added to the chain. Changing the contents of a copy in any node without being detected is next to impossible.

The bitcoin system comes with user interface software called a *wallet*. A user logs into a wallet and specifies transactions. The wallet handles all the details, such as representing the transactions as signed certificates, broadcasting them into the network, and receiving transactions from the network that would affect the funds in the wallet. The wallet converts currencies such as the U.S.

Does Majority Consensus Work?

With any distributed ledger, there is a serious problem in deciding which copies of new blocks are valid and which should be discarded. The problem is complicated by the fact that some of the miners may be dishonest and are seeking to insert fraudulent blocks that pay them a lot of money. In 2017 there were nearly 5,000 miners in the bitcoin blockchain. Nakamoto's Point-of-Work (POW) is a clever way of using the computing power of the majority to overcome a dishonest minority. In addition, mining requires very fast computers that consume lots of electricity. Miners only get paid when they win the POW race. What happens if one miner becomes so powerful that it starves out all others? The mining community shrinks to a handful that can afford very expensive, fast, energy-hungry computing machines.

In 1982 Lamport, Shostak, and Pease proposed a solution to a very similar problem, called the Byzantine Generals Problem. A general in charge of an army wants to distribute attack or retreat orders to N lieutenants. M lieutenants are disloyal and will not relay the orders properly. The loyal lieutenants can use a majority vote to decide if the messages they received are legitimate. Lamport and his colleagues proved that the vote can be trusted only if the honest majority is at least $2N/3$, and N must be at least $3M+1$. Under those conditions the minority votes cannot prevail. The figures below illustrate a simple case. The honest lieutenants (*green*) relay the order correctly; the dishonest ones (*red*) lie. The honest lieutenants go with the majority vote of the confirmations. In the left figure, each green node receives votes 2:1 favoring attack, and act as the commander ordered. In the right figure, each green node receives votes 1:1 and cannot reach an affirmative decision to attack.



dollar into bitcoins through a system of brokers called BTC exchanges.

Signatures and Hashes

In the world of online transactions, two concepts are absolutely fundamental: the digital signature and the hash. Using a bitcoin wallet, a user can obtain a matched pair of encryption keys (one public and one private, each consisting of patterns typically 256 bits long), the basis of a public key cryptosystem. In the blockchain, message secrecy is unimportant, but authenticity is. Alice's transaction "Pay 1 BTC to Bob" is encoded as "Pay 1 BTC to public-key-of-Bob," signed with her private key, and submitted to the blockchain. Only Bob can "cash" the transaction—in other words, use his secret key to unlock the transaction. The digital signatures guarantee the validity of the transactions recorded in the ledger.

But the signatures do not protect against double spending, which could happen if Alice inserted two identical "pay Bob" transactions into the blockchain. The blockchain protocols would not accept Alice's duplicate transaction, but if Bob can cash them before the protocols reject the duplicate, Bob gets paid twice and Alice pays once. Nakamoto solved this problem by a complex series of cryptographic techniques based on what are called *hashing functions.* A *hash* is the output of a special function that takes all the bits of a file, scrambles them, and condenses the scramble into a fixed number of bits, typically 256. A well-designed hash function will so completely scramble the original file that altering a single input bit causes most bits in the output to change. The hash function is not invertible: Given the output, the only way to find the input that generated it is an exhaustive search of all possible inputs, which would take more time than the remaining life of the universe. To guarantee that no one can tamper with any transactions in the blockchain, each is entered into the chain along with its hash. It is easy to verify that a transaction is valid simply by computing its hash and comparing with the hash stored with the transaction.

The hash of a block of transactions in the chain can be computed by combining the hash of the previous block with the hashes of all transactions in the current block and a *nonce* (a random number used once). This method of linking blocks is called a *ratchet* because any change to a block requires recomputing all the block hashes all the way to the end of the chain.

Proof of Work

Using an algorithm called a *Proof-of-Work* (POW), the time to compute a new hash of a block is made to be time-consuming and expensive, on purpose and for good reason. If the time needed to make a hash is approximately the same as the time it takes for the network to add a new block to the chain, it becomes impossible for a hacker to catch up and substitute the recomputed chain for the original.

The nonce plays a pivotal role in POW. The nonce is not just any number, but a random number that causes the hash of the block to begin with a certain number of leading zero bits. For example, if the number-of-zeros parameter is 60, the block hash must begin with 60 zero bits. The only way this format can be met is to find a nonce that, when combined with all the other hashes of the block, yields a block hash beginning with 60 zero bits. Because hash functions are not invertible, this process can only be done by repeatedly trying a nonce that is one larger than the previous nonce.

The number-of-zeroes parameter in bitcoin is frequently adjusted so that POW takes an average of about 10 minutes. That means that a hacker expects to spend about 10 minutes of computing time to find a nonce that generates a valid hash for a fraudulent block. In turn that means that the hacker has a very small chance of overtaking whoever is linking a new valid block to the chain. Nakamoto set up the system to make POW progressively harder (with more leading zeroes) so that new nonces are progressively harder to find. Finding new nonces becomes impossible when the total number of bitcoins is 21 million.

The value of cryptographic coins is calculated based on a history of transactions recorded in a ledger, similar to verifying the title to a property through a review of all its previous sales and liens.

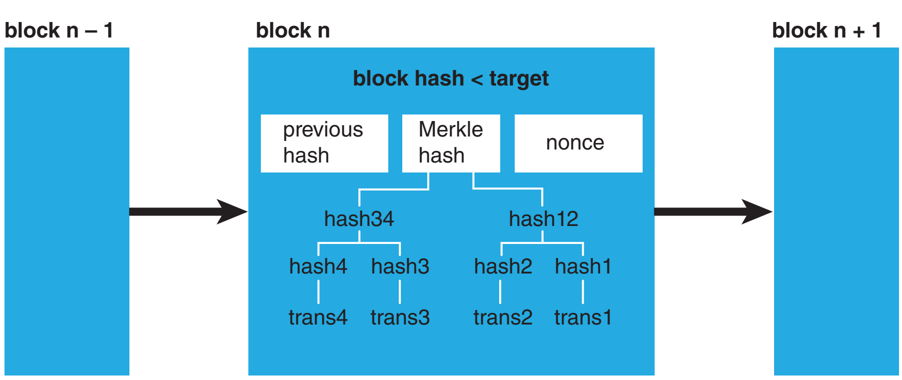
Miners

A miner is a special node that computes the POW for new blocks proposed for the chain and builds the blockchain by performing POW and adding blocks. It works as follows:

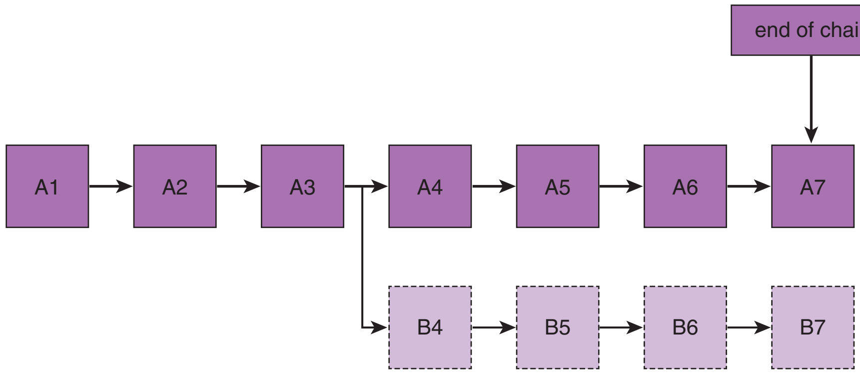
New transactions are broadcast to all nodes.

Each node collects new transactions into a block.

Each miner node does the POW to find a hash for the new block.



A blockchain is a linked list of blocks, each of which contains transactions, a previous block's *hash* (a unique string of bits that is generated based on the bits in the file), a Merkle hash of the transactions (which is a way to repeatedly hash pairs of hashes until only a single, total hash remains), and a *nonce* or random number produced as a byproduct of Proof-of-Work (POW).



A normal blockchain of valid transactions consists of blocks A1 to A7. A hacker tries to insert a fraudulent block B4. Because of the use of a Proof-of-Work (POW) algorithm, the hacker has to fork the chain by copying all the blocks from A4 to A7, and recomputing their hashes. The POW makes the time to recompute a block's hash the same as the time to add the next new block to the chain. The hacker can never catch up and is unable to force the forked chain to substitute for the correct chain.

Different miners working in parallel will find different nonces and hashes.

When a miner node completes a POW, it broadcasts the block and its hash to all nodes.

A node validates the new block by verifying the signatures and hashes of all the transactions and then adds the block to its local copy of the blockchain.

After accepting a block, a miner node begins working on creating the next block in the chain, using the hash of the accepted block as the previous hash. All the other miners drop their attempts to validate a block and start over to mine the next block.

The first miner to find a hash for the new block gets a reward of a few bitcoins.

Miners are constantly looking for ways to amass so much computing power that they can always win the race to add the next block to the chain, and thereby reap the reward. Mining is such a big business that there are now special-purpose chips for computing hashes. This environment encourages the formation of a cartel of miners, who together control 51 percent or more of the computing power in the network, to collaborate on producing new blocks and share the rewards. With at least 51 percent of the computing power, the group can complete POW faster than any nodes in the network. This scenario is called a *51-percent attack* because it can also be used to populate the chain with fraudulent blocks.

To prevent such fraud, Nakamoto's system depends heavily on self-interest rather than trust. He claims, "If a greedy attacker is able to assemble more CPU power than all the honest nodes, he would have to choose between using it to defraud people by stealing back his payments, or using it to generate new coins. He ought to find it more profitable to play by the rules, such rules that favor him with more new coins than everyone else combined, than to undermine the system and the validity of his own wealth." But Nakamoto overlooked the possibility of a 51-percent cartel.

Blockchain verification is intended to be computationally expensive, which means it is also energy intensive.

Blockchain Spinoffs

Although blockchains were invented for the bitcoin currency, many see their potential for other purposes, such as dealing with public and organizational records in a decentralized network. There are proposals to move the Internet's DNS (Domain Name System) to a blockchain, which would be much faster as well as resilient against outages. Despite the enthusiasm, there are also a number of important issues that give pause and will need resolution before blockchains will be widely adopted. These issues include performance, trust, volatility of cryptocurrencies, reliability, and overall world energy cost from operating blockchains.

Performance

By design, updates in the Nakamoto blockchain are computationally intensive. The result is that the overall throughput of the bitcoin network is about 7 transactions per second, and it takes about 10 minutes to close a transaction after it is submitted. In contrast, credit card companies and banks today can handle thousands of transactions every second, and give almost instantaneous response times to close transactions. On August 1, 2017, the BTC blockchain experienced a hard fork when the block size was changed from 1 megabyte to 4 megabytes, and a lighter version of BTC called *bitcoin cash* was created to improve transaction processing speed.

In addition to massive computational power, the blockchain requires massive storage. In 2017, the blockchain occupied 100 gigabytes, and it took many days to download and verify a new copy on a personal computer. This blockchain clearly cannot scale up to the size that would be needed to support a cryptocurrency with billions of users.

There are at least a dozen alternative blockchain architectures that all aim to reduce the computational work to verify a new block and add it to the chain. The most promising is Ethereum, which uses a *Proof-of-Stake* (POS) rather than Proof-of-Work: In POS, nodes with larger amounts of currency get greater weight in a vote determining when a new block will be added to the chain. But these alternative currencies appear to be more susceptible than bitcoin to a 51-percent attack. It will take some time to test out all these alternatives to find out which ones are reliable, scalable, and hackproof.

Another alternative is being explored by the Hyperledger.org project, which aims to produce an open architecture for business blockchains. They have defined a set of layers, each providing a critical function. One of the most critical to performance is the consensus layer, which adds a proposed new block to the chain once a consensus of the nodes is achieved. They are experimenting with alternatives to POW that are much less computationally intensive, such as using a lottery to select a node that then provides the proposed new block or accepting the proposed block only if a majority vote for it. These new options are possible when assumptions about trust are relaxed. In Nakamoto's network, no one trusts anyone; the system is designed to achieve consensus on additions to the blockchain when each node is anonymous and considers the other nodes as untrustworthy. In the new system, people within an organization may reveal themselves and have a greater basic level of trust.

Trust

Miners of the blockchain network are becoming highly specialized professionals relying on expensive chips specifically designed for POW. Because many ordinary users cannot afford to be miners, much of the computing power is concentrating in a relatively small minority of the network. After a recent $79 million heist of Ethereum coins, the miners proposed to collaborate so that they could go back and revise the blockchain to delete the transactions performing the heist. Although this act would deprive the criminals of their bounty, it also would undermine trust in the network. What else might a coalition of miners do?

Volatility

The exchange rates for cryptocurrencies can be quite volatile, which encourages speculators to buy up bitcoins and hoard them. The saturation limit of 21 million BTC adds to the volatility because of the scarcity of the coins. There is no central government bank monitoring the currency and keeping its value stable. For this reason, the Chinese government has banned speculation in ICOs (Initial Coin Offers) that exchange cryptocurrency for investment in startups.

The volatility of bitcoin values is likely to be correlated with social and political uncertainty, especially as confidence in government declines. At one point, rumors that China held most bitcoins in circulation led to price spikes because of fears that China might devalue its currency. The exponential increase in bitcoin exchange rates during the 2016–2017 seasons may be a reaction to political changes in Europe and the United States.

Economist William Luther of Kenyon College in Ohio claims that bitcoin has failed to catch on because of the steep switching cost in moving away from government-backed currencies. Bitcoin adopters need to have their own copy of the ledger (taking up 100 megabytes in 2017) and enough computing power to do the hashes and signatures rapidly. And even if they do make the switch, the volatility could easily rob them of their wealth.

Energy Cost

We think the biggest problem of all is in the POW idea itself. It is intended to be computationally expensive, even when the world's most powerful computers and special chips are deployed. This computation load means that POW is also energy intensive. Various estimates are that the current electrical demands of world data centers consume about 7 percent of the world's electricity. Massive use of blockchains would significantly boost that total. Do we want a system of currencies that requires so much of the planet's energy to power? Or that becomes ever more susceptible to total collapse if someone finds a way to shut down the power grids?

Coining a New System

It seems most likely that bitcoins will remain limited in use until there are affirmative answers to questions about their security and resilience, but such solutions are not obvious right now. On the other hand, the idea of blockchain ratcheting is already spreading to related technologies and could grow further, assuming some of its issues can be resolved. For example, the Signal Protocol used by social networks such as Instagram and Facebook employs ratchets inspired by blockchain technology. It's another case where the adoption of a technology has taken an unexpected turn, and determining which new developments will stand the test of time can be difficult to predict.

Bibliography

Cocco, L., and M. Marchesi. Modeling and Simulation of the economics of mining in the bitcoin market. https://arxiv.org/pdf/1605.01354v1.pdf.

Franco, P. 2015. *Understanding Bitcoin: Cryptography, Engineering, and Economics.* Chichester, UK: John Wiley & Sons.

Lamport, L., R. Shostak, and M. Pease. 1982. The Byzantine Generals Problem. *ACM Transactions on Programming Languages and Systems* 4:382–401. doi:10.1145/357172.357176.

Luther, W. J. 2013. Cryptocurrencies, network effects, and switching costs. Mercatus Center Working Paper No. 13–17. https://ssrn.com/abstract=2295134.

Engineering

Problematic Pedestrian Bridges

Crowds walking, running, dancing, and even standing still on these spans often prove to be the ultimate test of a structure's soundness.

Henry Petroski

Pedestrian bridges are ubiquitous, but most are unremarkable and anonymous. They span interstate highways to carry hiking and biking trails over heavy traffic, thus keeping walkers, runners, and cyclists safely above the fray. Some pedestrian bridges rise to prominence, however, often because of controversies or problems associated with their location, design, construction, or performance.

This was the case with the now-infamous London Millennium Bridge connecting across the River Thames the tourist areas of St. Paul's Cathedral and Tate Modern art museum, housed in the repurposed Bankside Power Station. As originally built and opened in 2000, the bridge was the result of an imposed collaboration among an engineer, an architect, and an artist. Within three days of its opening, the low-slung suspension bridge was closed because it swayed too much under the crowds of people who walked across it in synchrony with its movement.

The problem was traced to neglecting to consider that as people walk, they naturally exert on the pavement sidewise forces, in addition to the more commonly recognized vertical ones associated with their feet pushing downward. The downward pushes are reacted to, according to Newton's Third Law, by an upward force from the pavement. And such vertical forces are accompanied by friction forces that prevent a foot from slipping and sliding. The component of friction force pointing straight ahead is what propels the walker forward; the effect of its reaction on a bridge deck is usually negligible. The sideways component of the friction force exerted on the bridge deck, however, can move it left and right.

Engineers generally check the natural frequency of a structural design against the typical frequency of the vertical tramp of human footsteps, which can range from one to four cycles per second. Unfortunately, in the case of the London Millennium Bridge, engineers neglected to check the structure's natural frequency against the frequency of sidewise motion, which is typically about half that of footfalls, because pushing off sideways with the right (or left) foot occurs only half as frequently as the two feet alternately coming down on the pavement. Ignoring this lower-frequency excitation of human footsteps set the unusually low-slung suspended bridge deck design into a swaying motion and made it unsuitable for its purpose.

Elephants and Human Stampedes

There is indeed a long history of bridges moving too much beneath the feet of pedestrians and marchers. Just the year before the London Millennium Bridge problem came to light, the new passerelle Solférino (now called passerelle Léopold-Sédar-Senghora) in Paris behaved poorly. In each case, the structure had to be modified to alter its natural frequency that coincided with that of pedestrian footfalls. In addition, *tuned mass dampers* (devices that ameliorate vibrations, often by means of countervailing movements) were added to the bridge in London to keep in check any motion that developed.

For whatever reason, even highly experienced engineers sometimes fail to take into account all the possible dynamic effects of people walking, running, marching, and dancing on bridges. Perhaps pedestrian bridges are not considered as seriously as the heavier, wider, and usually much longer and grander structures that carry railroad and vehicle traffic.

Yet even crowds of people standing still have proven time and again to represent the ultimate test of a bridge structure. In Missouri, the Kansas City Hyatt Regency Hotel's elevated walkways, which were essentially bridges spanning the building's atrium, collapsed in 1981 under the sheer weight of people, some evidently moving to the music of a band playing for a tea dance. The tragedy claimed 114 lives and injured many more. The structural engineers responsible for the design were found negligent in approving a suggested change in a support detail without giving it sufficient thought.

To celebrate the 50th anniversary of the opening day of the Golden Gate Bridge in San Francisco in 1987, the structure was closed to vehicle traffic, and pedestrians were allowed to crowd onto virtually every square foot of the bridge's roadway and sidewalks. The resulting shoulder-to-shoulder mass of celebrants represented the largest load the bridge had experienced in its half-century of operation, a condition evidently not anticipated by its design engineers. The center of the bridge's roadway, which under normal circumstances

Henry Petroski is the Aleksandar S. Vesic Professor of Civil Engineering and a professor of history at Duke University. His most recent book is The Road Taken: The History and Future of America's Infrastructure *(Bloomsbury, 2016). Address: Box 90287, Durham, NC 27708.*

has a distinct camber to it, flattened out noticeably, and engineers observing the phenomenon expressed genuine concern. Thankfully, the Golden Gate Bridge survived its trial by pedestrian, and subsequent anniversaries have not been the free-for-all of the 50th. It is reasonable to expect that limitations on the crowd size will be imposed when the 100th anniversary is celebrated.

The most recent debacle involving a bridge and the people designing and using it occurred in New York City, in the shadow of the Brooklyn Bridge. Ironically, that historic 130-year-old structure carries one of the most celebrated and successful pedestrian walkways in the world. The walkway of the Brooklyn Bridge was thoughtfully designed to rise above the traffic on either side, thus giving amblers a clear view of the surroundings. Depending upon which way one transits the bridge, there are spectacular and virtually unobstructed views of New York Harbor, the Statue of Liberty, the Manhattan skyline, and other massive New York City bridges, including the Manhattan to the north and the Verrazano-Narrows to the south. This latter bridge is the site for the beginning leg of the annual New York City Marathon, whose route crosses several bridges to touch each of the city's five boroughs.

Pedestrians or runners have never caused any of these great suspension bridges to move in an unsafe manner, to my knowledge, but an unfortunate incident occurred just a week after the Brooklyn Bridge opened in 1883. A rumor that the new bridge was about to collapse, with about 20,000 people crowded onto its walkway, touched off a panic that resulted in a stampede, leading to the death of a dozen people and the injury of three times that many. The bridge was never in any danger of collapse, but some people remained unconvinced until the circus came to town the following year and the great showman P. T. Barnum led a parade of 21 elephants across the structure. It had long been believed that by some sixth sense even a single elephant would balk at setting foot on an unsafe bridge. Indeed, an elephant or two was often called upon to test a new bridge to prove its safety.

Brian Jeffery Beggerly/Wikimedia Commons

The London Millenium Bridge, which opened in 2000, was the first new bridge to cross the Thames River in more than a century, and was London's first dedicated pedestrian bridge. Within a few days, it closed because it swayed too much from the sideways load of a crowd's footsteps. After repairs, the bridge no longer wobbled and was reopened in 2002.

Aesthetics and Views

Even before the Brooklyn Bridge was completed and tested, the waterfront adjacent to the famous structure was dense with piers, to and from which goods from all around the world were loaded onto and off of ships. This commerce was in part what made the East River, over which the great bridge spans, so busy, and it was precisely to alleviate some of the congestion on the water that the bridge was supported financially, politically, and popularly. The Brooklyn waterfront began to change dramatically in the 1950s with the introduction of containerized shipping, which marked the development of larger ships requiring larger piers. In time, much of the dockside commerce was driven across the harbor into Newark Bay. All cargo ship operations in the vicinity of the Brooklyn Bridge ended in 1983, the centennial of the structure's opening.

With the piers obsolete and abandoned, there remained the question of what to do with the properties. The idea for a Brooklyn Bridge Park was pushed in earnest when the Port Authority of New York and New Jersey announced plans to sell the piers for commercial development, a move that could have resulted in high-rises blocking views from preexisting buildings of the river, its bridges, and lower Manhattan. Activists prevailed, and in 2002 an agreement was reached to create a Brooklyn Bridge Park on 85 acres, some reaching out into the East River and incorporating the remains of the obsolete Brooklyn piers. Parts of the parkland would be landfill built from soil excavated from the construction site of the new One World Trade Center, also known as Freedom Tower, located south of the site of the twin towers that terrorists destroyed in 2001.

@Painet Inc./Alamy Stock Photo

During the 50th anniversary of the Golden Gate Bridge in 1987, the bridge was closed to vehicles, and pedestrians were permitted to pack the span, adding far more weight on the bridge than would any vehicle traffic. Engineers expressed concern, but fortunately the structure held.

Construction of the park began in 2008, and it was opened in stages. As attractive as the park was to people living in the neighborhood, they were cut off from easy access to it by the heavily trafficked Brooklyn-Queens Expressway, which separated a landlocked section known as Squibb Park from Brooklyn Bridge Park proper. The original design for a pedestrian bridge connecting the waterfront park with the residential area of Brooklyn fell to HNTB, a world-class infrastructure firm based in Kansas City, Missouri. HNTB is an employee-owned corporation that traces its roots to the founding in 1914 of the firm Harrington, Howard, and Ash, which ironically specialized in the design of moveable bridges. The movement of those bridges was deliberate, however, because they were typically built with low clearances over waterways in heavily industrialized parts of cities. Various types of intentionally moveable bridges swing, rise, or lift out of the way to allow ships to pass along the rivers, canals, and navigations they span. Many engineers of the late 19th and early 20th centuries cut their professional teeth designing such structures.

In 1941, the firm of Harrington, Howard, and Ash changed its name to Howard, Needles, Tammen, and Bergendoff, after the surviving founding partner and his new partners. The mouthful corporate name was formally changed to the simpler HNTB in 1993, and in 2000 the firm became employee-owned. Today, HNTB's most prominent engineer is probably Theodore P. Zoli III, usually called simply Ted Zoli. He came to prominence beyond the engineering community when in 2009 he was the recipient of one of the genius grants from the John D. and Catherine T. MacArthur Foundation, which identifies him as "a structural engineer who is leading the design of elegant and enduring bridges around the world." Besides being a MacArthur Fellow, Zoli is a visiting lecturer at Princeton University, his alma mater, and an adjunct professor at Columbia University. He was the principal designer of the Squibb Park Bridge.

The bridge had become so weighed down with hundreds of thousands of locks that its fragile-looking ironwork was in danger of buckling.

Zoli's vision for the Squibb Park Bridge was at the same time pedestrian and exotic. In plan, it resembles, depending on your point of view, an elongated M or W or a distorted Greek Σ. Any way you look from it, the zigzag of its walkway gives the person walking down to Brooklyn Bridge Park changing views, from lower Manhattan across the river to the Brooklyn Bridge just upriver. The feel underfoot of the 120-meter-long bridge was intended by design to evoke the sensation of walking across the kind of bouncy bridge often found in state parks.

But the Squibb Park structure is not a conventional suspension bridge. It has no network of vertical suspenders dropping from curving steel cables to support its walkway of unconventional black-locust wood, a rot-resistant hardwood that is expected to last a long time in the vicinity of the estuarial East River.

But it is not only what was above the bridge's walkway that is remarkable. The structure's concrete piers are not simple cylinders or prisms, as are those of so many highway bridges. The piers of the Squibb Park structure rise out of the ground like tree trunks, and branch out to suggest perennially leafless limbs. And the wooden trusses spanning between the arboreal supports are supplemented by steel cables, which are the mechanism that at the same time give the bridge both its strength and its limited stiffness, or bounce.

The bridge opened to critical acclaim in the spring of 2013. After a little more than a year, however, it was abruptly closed. The walkway was not only bouncing more than desired, it was also twisting in ways that were considered potentially unsafe. Zoli and other HNTB engineers sought to fix the structure by "retensioning" it to bring it back into alignment, and by adding cross bracing and otherwise retrofitting it to maintain that alignment. An impatient Brooklyn citizenry was told the bridge would reopen in a matter of months, which soon dragged on past a year.

When HNTB did not seem able to bring the repairs to a satisfactory conclusion, the Brooklyn Bridge Park Corporation, the operator of the park, announced that it was suing the engineering company for the $3 million cost of the bridge and that a different firm would take over responsibility for developing a plan to fix the structure. The company chosen to do the retrofitting was an international engineering firm called Arup. Incidentally, this is the same firm that was embarrassed to be responsible for the engineering of the Thames-crossing London Millenium Bridge, nicknamed "Wobbly" by the British. But Arup did eventually fix that structure.

Among the fixes Arup applied to the Squibb Park Bridge were improvements to the connections between the timber truss and steel cables under the walkway. In addition, tuned mass dampers, in the form of heavy metal plates attached by springs to the underside of the main structure, were installed to reduce the bounce. Movement of the walkway was not meant to be completely absent underfoot, but it would be more muted and less alarming to walkers with a low tolerance for motion of an unexpected kind and amplitude. The modifications were not expected to alter much of the appearance or experience of using the bridge, but it was believed to be much more safe when it reopened in the spring of this year.

Lee Snider/Alamy Stock Photo

The Squibb Park Bridge in Brooklyn, New York, opened in 2013 but was closed within a year because it was deemed to bounce and twist too much for pedestrian comfort. After false starts on repairs, the bridge reopened 32 months later with retrofittings to damp out the effects of footsteps.

Too Much Love

Just as a large assemblage of people on a bridge can cause it problems, a bridge can also receive too much love. Around 2010 in Paris, a fad gained popularity: A pair of lovers would affix a metal padlock to the ironwork of a bridge across the Seine and throw the key into the river as a symbol of their permanent attachment to each other. By 2015 the Pont des Arts—the Seine crossing located just downriver from the Pont Neuf at the foot of the Île de la Cité—had accumulated so many heavy locks that entire sections of balustrades and the meshwork below had to be replaced lest they be pulled by gravity onto the deck of a passing tour boat. But as soon as a section of meshwork was replaced, it again began to attract a new collection of locks, many etched with the names or initials of the lovers.

In 2015 it was estimated that the Pont des Arts had become so weighed down with hundreds of thousands of locks that its fragile-looking ironwork was in danger of buckling. By one estimate, the collection of love tokens amounted to an extra 45-ton load on the structure, about the same weight as the entire amount of black-locust wood planking in the Squibb Park Bridge. To keep the Pont des Arts from being pulled down into the Seine, in the spring of 2015 city workers removed entire sections of lock-laden bridge grills and replaced them with side panels decorated with designs too wide to allow a common lock to be affixed to them. Authorities had no plans about what to do with the estimated 700,000 keys resting on the bottom of the Seine.

When the French tradition moved across the Atlantic, padlocks began to be left along the walkway of the Brooklyn Bridge. But after a wire supporting a light fixture snapped under the weight of too many locks, the city moved quickly. Workers using heavy wire cutters tried removing locks one at a time, but the Sisyphean task soon led to the posting of signs with a humorous spin on the bridge. They read: NO LOCKS… YES LOX, the first phrase followed by the image of a padlock in a red circle with a slash through it. The second was illustrated with a New York bagel loaded with smoked salmon and a schmear of cream cheese. That the offense was no laughing matter was emphasized by the punch line: FINE $100. Still, padlocks can often be found hanging from the same light structures that hold the signs. Humans, after all, aren't known for considering how their small actions can affect the greater good or even their own potential safety.

Selected Bibliography

Foderaro, L. W. 2016. Repairs coming for bridge that bounced too much. *New York Times* (July 14).

Foderaro, L. W. 2017. Rhythmically but safely: That's the way the bridge bounces. *New York Times* (April 18).

Nir, S. M. 2016. Dear John… Token of love discouraged at landmark. *New York Times* (November 15).

Rubin, A. J., and A. Breeden. 2015. Showing love for historic bridge, Paris removes tons of padlocks. *New York Times* (June 2).

Webster, N., and D. Shirley. 2016. *A History of Brooklyn Bridge Park: How a Community Reclaimed and Transformed New York City's Waterfront*. New York: Columbia University Press.

American Scientist Previous Page | Contents | Zoom in | Zoom out | Front Cover | Search Issue | Next Page

Perspective

The Forgotten Mystery of Inertia

A century after Ernst Mach and Albert Einstein cast doubt on absolute space, we still don't know how a gyroscope stays pointed in a fixed direction.

Tony Rothman

In days of yore, at a World Science Fiction Convention in Boston, a Harvard graduate student polished his reputation as a brilliant mad scientist by roaming the convention halls, brandishing what at first glance appeared to be a rather peculiar steel bowling ball. Portholes perforated its surface, providing a glimpse of electronic hardware inside; tangled wires sprouted from the same holes, and a gear train surrounded the mysterious object's equator.

"What's that?" I asked him.

"It's the gyro platform for an intercontinental ballistic missile," he replied. "If you put it on a Titan rocket, it will fly to Kiev."

"How do you know?"

"It's an inertial guidance system, stupid. It knows where Kiev is."

"I know how inertial guidance systems work, but how do you know it knows where Kiev is?"

"Oh, that. It was stamped on the box."

This sorcerer's apprentice had discovered that for $900 you could buy a surplus intercontinental ballistic missile, 10 years before the electronics were declassified. His Titan was delivered on two railway cars, "Kiev Titan Missile" stamped on the crates. He junked the body, donated the engines to an art museum, and saved the electronics for his research. A tall tale? Sounds like one, but the gyro platform was there for all to see.

I didn't understand it, nor did my mad interlocutor, and anyone who claims otherwise is not being entirely honest. Despite a gyroscope's utter simplicity—it is, after all, nothing more than a wheel on an axle—it remains the most fascinating and mysterious device ever created. Spin up the wheel, set it on a pedestal, and there it stays, pointed at…?

That is the question. At what, exactly, is the gyroscope pointed? According to the law of inertia, objects tend to continue doing what they've been doing: If at rest, they remain at rest; if moving, they continue moving at the same speed in the same direction. The gyroscope also bends to inertia's will, but in confounding ways. Touch it, and the gyro opposes you by veering in unexpected directions. If it is spinning extremely rapidly, the gyroscope remains rigidly locked in the direction it has been set, its sights set on... Kiev—hence the term *inertial* guidance systems. If a rocket veers off the gyro's fixed course, a sensor detects the error, and a servomechanism realigns the missile with the gyroscope axis.

But this is just the explanation stamped on the box. What tells the gyro it is set on the Great Gate of Kiev? Isaac Newton would argue that the gyro is pointed in a fixed direction relative to "absolute space," what physicists term an *inertial reference frame*—indeed, the ultimate inertial reference frame. Think of it as an invisible reference grid somehow etched into the fabric of the universe. But if absolute space is a highly abstract concept, the gyro's behavior is very tangible. Set up a gyro on a lecture table at a university (or a coffee table at home), and as the day progresses it appears to rotate with respect to walls…or is it vice versa? Are the rotations relative, or are they absolute?

In his magnum opus, *Principia Mathematica*, Newton proposed a thought experiment to prove that rotation takes place with respect to absolute space. He imagined a bucket partially filled with water and hanging from a rope, which an experimenter has twisted up. When the experimenter releases the bucket, the rope untwists, and the bucket begins to spin. At first the water remains flat, but as the pail speeds up and drags along the water, its surface eventually becomes concave due to the centrifugal force of the rotation. At that stage, the water and vessel are rotating together, and there is no relative motion between them. Yet somehow the water "knows" to create a concave surface.

Newton insisted that the concavity must be due to the water's rotation with respect to something else—absolute space. Rotation is absolute, not relative. That answer stood largely unchallenged for two centuries, until the Austrian physicist Ernst Mach flatly declared Newton to be wrong.

Relative Revolutions

In his 1883 book *Science of Mechanics*, Mach wrote that Newton's thought experiment "simply informs us, that the relative rotation of the water with respect to the sides of the vessel produces no noticeable centrifugal forces, but that such forces are produced by its relative rotation with respect to the mass of the Earth and the other celestial bodies." Mach continued, that "no one is competent to say how the experiment would turn out if the sides of the vessel increased in thickness and mass till they were ultimately several leagues thick." He dismissed absolute space as an "arbitrary fiction of our imagination."

Tony Rothman teaches in the applied physics department at New York University. His latest book is The Course of Fortune *(iBooks, 2015), a historical novel about the Great Siege of Malta in 1565. Email: tonyrothman@gmail.com*

American Scientist Previous Page | Contents | Zoom in | Zoom out | Front Cover | Search Issue | Next Page

Albert Einstein's general theory of relativity, inspired by Ernst Mach's rejection of absolute space, predicted that spinning black holes or other massive objects would drag around the surrounding reference frame of space; such "frame dragging" should deflect a gyroscope placed there. The effect has been verified, but its broader implications for inertia remain unclear.

Mach never gave a precise formulation of what became known as Mach's Principle. Nevertheless, the essential idea is simple enough. According to Mach, Newton's conception of absolute space lacks all meaning. Inertia—that tendency of massive objects to move at constant velocity—must depend on other bodies, because motion itself must be measured relative to other bodies. Rotations and accelerations along straight paths take place with respect to the reference frame of the distant stars and galaxies. The centrifugal forces that throw you to the side of an automobile as it rounds a corner arise because you are accelerating with respect to the distant matter in the universe.

Reframed as a question, no proposition goes straighter to the fundamentals than Mach's Principle: Would you feel centrifugal forces in an empty universe? Does the law of inertia mean anything in an empty universe? Mach would give a resounding "no" to both questions: Inertia is not a property intrinsic to an object but depends upon all the mass in the universe. Mach's suggestion proved heretical, not least because Newton's laws, with their assumption of absolute space, work so exquisitely. It furthermore went against our human intuition that our experience of the world is fundamental, not contingent.

Mach's arguments had a profound effect on Albert Einstein, who devised the general theory of relativity largely to abolish the idea of absolute motion, and indeed it was Einstein who, in 1918, coined the phrase "Mach's Principle." The basis of Einstein's general theory was his observation that the gravitational force the Earth exerts on you is canceled out in a freely falling elevator. When accelerating downward in an elevator whose cable has snapped, you feel weightless. By the same token, when accelerating upward in an elevator, you feel heavier than usual, as if the gravitational pull of the Earth has suddenly increased.

With "the happiest thought of his life," Einstein realized that within the elevator's confines, it is impossible to distinguish acceleration from gravity. To explain the origin of accelerations, then, he would need to create a theory of gravitation. Furthermore, because free fall abolishes gravity in an elevator, the origin of inertia cannot be found in the interaction with nearby bodies such as Earth. Inertia's origin must depend on the distant matter in the universe, as Mach had insisted.

Einstein hoped that, within the framework of general relativity, the distribution of matter of the universe would fully determine the inertia of material objects. Most modern discussions of Mach's Principle (when the topic comes up at all) center on whether general relativity succeeded in this goal. As simple as the question sounds, there is as yet no definitive answer.

While still formulating general relativity, Einstein calculated that the gravitational field of a shell of matter rotating around a gyroscope (picture a gyro inside a spinning, hollow Earth) would displace the gyro's axis relative to the fixed stars. The gravitational field literally drags around the gyro's "inertial compass." In 1918, Austrian physicist Hans Thirring, with Einstein's input, published a similar calculation based on the completed relativity theory. Josef Lense, an Austrian mathematician, later provided relevant astronomical observations. Inevitably, this concept of *frame dragging* has come to be known as the Lense-Thirring effect.

In the paltry gravitational field of the rotating Earth, the predicted amount of frame dragging is enough to displace the axis of an orbiting gyro by only 0.042 seconds of arc annually. That is roughly the angular size of a quarter held aloft by the Statue of Liberty in New York as viewed by William Penn atop Philadelphia's city hall. Despite the challenge of measuring such a minuscule effect, frame dragging has been detected by the Laser Geometric Environmental Observation Survey (LAGEOS) and Gravity Probe B satellites, although the results were too uncertain to confirm the exact prediction of Einstein's theory.

In 1963, New Zealander Roy Kerr, then at the University of Texas, discovered the general-relativistic description of rotating black holes. Soon physicists recognized that frame dragging can become far more pronounced around extreme astronomical bodies. Several teams have claimed to find observational evidence of frame dragging in the disks around supermassive black holes, although the results are indirect and imprecise.

In Isaac Newton's thought experiment, water forms a flat surface in a still bucket but a concave surface in one that is spinning. He asserted that this behavior is determined relative to absolute space.

The Cosmic Connection

There is no longer a question in any physicist's mind that general relativity predicts Machian effects on gyroscopes, and so far, the data seem to support Einstein's predictions. The question of how the distant reaches of the universe give marching orders to a gyroscope spinning on my desk is an altogether trickier issue.

The first triumph of general relativity was its exact prediction of the orbital precession of Mercury's perihelion: the small but puzzling 43 seconds of arc per century by which the point of the planet's closest approach to the Sun shifts with time, a movement that Newton's laws could not easily explain. Einstein described the gravitational field of the Sun in terms of curved space and showed that the curvature caused Mercury's perihelion to shift by precisely the right amount.

As physicist Erwin Schrödinger wrote, however, in a striking 1925 article on Mach's Principle, "…every naïve person has to ask: With respect to *what*, according to the *theory*, does the orbital ellipse perform this precession, which according to *experience* takes place with respect to the average system of the fixed stars?" In calculating the effects of curved space around the Sun, Einstein needed to assume that at large distances from the Sun, spacetime becomes flat—and absolute. In other words, he had to impose by fiat "boundary conditions" at infinity to complete his solution. General relativity, in and of itself, did not entirely determine the precession of Mercury's orbit.

Mathematician Kurt Gödel brought the point home in 1949 when he published a cosmological model designed to show that Mach and Einstein were incompatible. Gödel assumed the same uniform distribution of matter as Einstein did in his first cosmological model of 1917, but Gödel's solution exhibited fundamentally different behavior. That discrepancy directly contradicted Einstein's premise that the universe's matter distribution should uniquely determine gyroscopic behavior.

The key difference was that Gödel's universe rotated, meaning that distant galaxies rotate with respect to a gyroscope sitting on my desk, and that anyone, anywhere in the universe would observe the same behavior. (It does *not* mean that the universe is rotating around some central axis.) For true followers of Mach, a gyro should track the bulk matter of the cosmos, and so it should remain stationary with respect to distant galaxies. Since Gödel, researchers have found other rotating models of the universe, all of which similarly contradict Mach's premise. Such models can be declared unphysical, however, because they flagrantly contradict observations of the real universe. Nevertheless, as theoretical solutions they demonstrate the difficulties that come with defining inertia purely in relation to other objects.

The litany of counterexamples eventually convinced Einstein to abandon Mach's Principle. In 1990, however, Harry King, then at the University of Texas, proved that a closed universe—one that is destined to stop expanding and eventually recollapse—can exhibit no rotation. Any rotation of galaxies as in Gödel's model is cancelled out by gravitational waves moving in the opposite direction. If Gödel's result was a setback for Mach, then King's result was a definite victory. Unfortunately, today's cosmologists believe that the real universe is open—destined to expand forever—demonstrating again how difficult it is to connect a theoretical understanding of inertia to the actual cosmos.

More recently, Christoph Schmid of the Swiss Federal Institute of Technology claims to have vindicated both Einstein and Mach. Schmid has concluded that vorticity added to a realistic cosmological model would indeed drag gyroscope axes. But according to Schmid's calculations, the influence of matter diminishes exponentially beyond a critical radius, related to the distance that light has traveled since the Big Bang, and the exact shape of space at much larger distances becomes irrelevant. In this way, he bypasses the necessity to impose boundary conditions at infinity, the problem that bedeviled Einstein. The matter distribution of the universe in and of itself determines the behavior of gyroscopes. Schmid claims that general relativity then perfectly embodies Mach's Principle; on the other hand, some of his calculations are performed in a closed universe, which would seem to run afoul of King's proof.

In independent analyses, Cambridge University's Donald Lynden-Bell and his collaborators have accepted certain aspects of Schmid's results. Modern studies of the cosmic microwave background radiation, however, largely rule out any large-scale rotations in the universe. Many cosmologists believe that the universe went through an early inflationary period, in which the size of the universe increased exponentially. Such inflation would likely have damped out any rotation, leaving a still cosmos in which gyroscopes would naturally be at rest relative to distant matter, rendering Mach's Principle redundant—at least as far as rotating objects such as gyros are concerned. A nonrotating universe still does not necessarily explain the forces that push you back into the seat of a car when you accelerate along a straight line.

To most contemporary physicists, Mach's Principle is not merely redundant, it is forgotten entirely—as unknown to them as it may have been to the madcap graduate student flaunting his missile guidance system. Yet the mystery is never far away. To this day, we do not truly understand how the distant universe gives gyroscopes their marching orders.

The backwater of Mach's Principle holds many lessons. One is that the teaching of undergraduate physics has become divorced from the practice of physics. From arbitrary conventions that are presented as natural laws, to the insistence on precise answers to unrealistic problems, to the banishment of concepts such as centrifugal force, university physics has evolved to inhibit creativity rather than encourage it. How can one hold high the tenants of relativity—that physics can be performed in any reference frame—while simultaneously declaring that Newton's laws hold only in inertial frames? How can one begin to discuss Mach's Principle if one denies the very existence of centrifugal forces? Aye, centrifugal forces disappear in inertial frames; gravity disappears in free-falling elevators. Does that mean that gravity is a fictitious force?

Does the law of inertia mean anything in an empty universe? Would you feel centrifugal forces? Mach would say no.

Part of the difficulty in tackling Mach's Principle lies in its similarity to the utterances of the Delphic Oracle. Julian Barbour and Herbert Pfister, organizers of a 1993 conference on Mach's Principle in Tübingin, listed no fewer than 21 different interpretations. An exit poll at the conference's end revealed that only three attendees believed that general relativity perfectly embodied Mach's Principle, whereas 21 did not. Fourteen viewed it as "very Machian," and seven did not consider it Machian at all. An illustrious colleague recently remarked to me that he found Mach irrelevant for doing science: "Science for me was always a bunch of tools, not a branch of philosophy."

The greater obstacle is one of fashion. Today's cosmologists constantly wade into shallow philosophical waters if they align with tastes of the current scientific Pradas and Versaces. They worry about creating Big Bang models whose input parameters occur "naturally" rather than needing to be "fine-tuned" by hand. They consider the "cosmological constant problem"—why the "dark energy" driving the universe's expansion is some 125 orders of magnitude less than what you'd "expect"—to be the outstanding dilemma of their field. Physics or philosophy? Each year hundreds of papers are published on the string theory landscape and on the universe of universes—the multiverse. Each year dozens of conferences take place on particle physics or string theory. The only conference on Mach's Principle took place in 1993.

But replace the quaint-sounding words "Mach's Principle" with "Why does a gyroscope point in a direction fixed relative to distant quasars?" and one stands face to face with one of the most striking questions nature presents to us. It is more of a question than "How does the Higgs boson impart mass to other subatomic particles?" It is hardly less a question than "Why does time move forward when the basic laws of physics do not?" It is perhaps more fruitful than the vastly more popular question, "What lies behind quantum mechanics?" And Mach's Principle outshines virtually all other riddles in one other category: It is pure romance.

We have a working theory of gravity, one that has been tested more precisely than any other theory ever devised. It accounts for the expansion of the universe; it describes the behavior of black holes; and it has successfully predicted the existence of gravitational waves. It should tell us why gyroscopes point at the stars.

Bibliography

Barbour, J., and H. Pfister, eds. 1995. *Mach's Principle: From Newton's Bucket to Quantum Gravity.* Boston, MA: Birkhauser.

Bondi, H. 2010. *Cosmology.* Mineola, NY: Dover Books on Physics.

Sciama, D. W. 1953. On the origin of inertia. *Monthly Notices of the Royal Astronomical Society* 113:34–42.

American Scientist Previous Page | Contents | Zoom in | Zoom out | Front Cover | Search Issue | Next Page

Adapting Your Body Clock to a 24-Hour Society

Can new research and technology help people address the disconnect between the body's sleep-wake cycle and the rhythms of modern life?

Alexis Webb and Erik Herzog

"Aunt Lexie!!!" came the ear-piercing squeal from my niece as she ran toward my legs. I (Webb) smiled weakly, exhausted. My husband and I had been awake for nearly 20 hours, rising at the equivalent of the night before, local time, to travel to America from our home in London. How do you explain jet lag to a six-year-old? She was having none of it—our visit should mean that she had two of her favorite playmates at her disposal. But after little rest over the course of a day spent on planes and sitting in airports, my husband and I were desperate for some peace and quiet. Convincing a child that your body still thinks it's on the other side of the Atlantic is nearly as difficult as convincing your body that it's not. Unlike our phones and other wired technologies, our bodies can't read the new time off the local network tower once we have arrived in a new location and adjust instantly.

Perhaps when my niece is my age, her activity tracker or some future device will be able to tell her brain the timing of her new environment, and jet lag will be a thing of the past. But so far, our access to technology has mostly been exacerbating the problem, not alleviating it.

Having studied biological rhythms and timing for decades, the two of us (Webb and Herzog) are usually the go-to people for friends and colleagues seeking advice on how to beat the transatlantic travel blues. Our basic tenets include: Spend time outside, preferably in direct sunlight, when you arrive. Eat a meal on local time. Try to stay awake until your normal bedtime where you are, not where you arrived from. For most people, flying west, which prolongs your wake-sleep rhythm relative to normal, feels easier than flying east. Although you may find yourself unable to sleep in the early morning hours and flagging at the end of the day, flying west tends to be kinder to the body. It's not always so easy to follow these rules, however, and it gets harder still as our daily routines keep changing in ways that can sacrifice well-being.

The feeling of jet lag, of being just a bit off and out of sorts, arises because of a discrepancy between internal time and the environment. In the brain sits a biological clock that coordinates the regular daily timing of nearly all of the body's processes. This clock is set or entrained to local time through light input from the eyes. The change in time experienced when people travel across time zones cannot be processed instantly by the clock in the brain; there is no watch hand to wind to a new time. The internal clocks must re-adjust to the new environment through experience of the new day-night cycle.

These days, people don't need to get on a plane to experience conditions that are at odds with internal rhythms. Modern 24-hour lifestyles have similar effects because society no longer relies on the day-night differences in external conditions that human bodies evolved to track and set processes to. The ever-present access to food, technology, and, most notably, electric light means that people are regularly exposed to stimuli that are not coordinated with the body's internal time. Couple this development with the constraints of the standard working day, which may necessitate waking before preferred morning rise time, and the clash of body time with environmental time is continual. Researchers have described this phenomenon as *social jet lag*—going through daily life at odds with the circadian rhythm—which can have profound consequences for health.

The good news is that advances in research and technology are providing better ways to track and train the body clock to cope with modern living. At the same time, scientists are improving their understanding of possible targets for drugs, making better use of data from wearable devices, and bringing together citizen scientists to examine real-world behavior, which can be vastly different from what is observed in the controlled conditions of the laboratory. New technologies that are changing our lifestyles and challenging our clocks will also be able to provide exquisite details about our biology, leading to solutions never realized before.

Entrainment in a Modern World

The modern world increasingly makes possible or even demands nighttime activity. Thanks to social media and other modes of connectivity, people from anywhere in the world can see what our planet looks like from the

Alexis Webb has a PhD in neuroscience and spent her early career researching the role of biological clocks in behavior and development in both the United States and Europe. She is a senior research funding manager for the world's largest cancer research charity and is based in London. Erik Herzog is a chronobiologist studying the molecules, cells, and circuits that underlie daily rhythms in physiology and behavior. He is professor and associate chair of biology at Washington University in St. Louis, where he is also the codirector of the Graduate Program in Neuroscience and the director of the St. Louis Neuroscience Pipeline Program, a National Institutes of Health–funded initiative to increase diversity in the neurosciences. He serves as the President-elect of the Society for Research on Biological Rhythms. Email for Webb: lex.b.webb@gmail.com

Edwin Eemsberg/Alamy Stock Photo

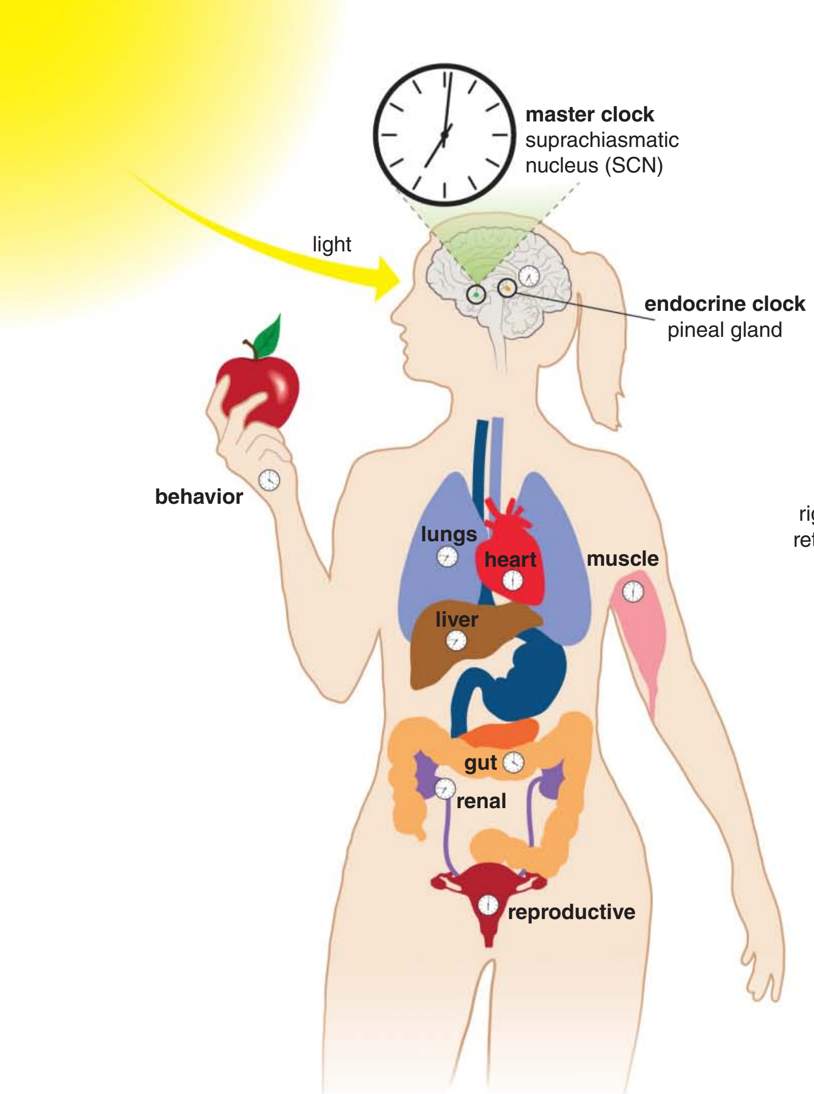
People around the world, including crowds under ornate lights in Taksim Square in Istanbul, Turkey, are becoming more active at night with the rise of connectivity and technologies that began with the electric era nearly 100 years ago. The rhythms of modern life are often out of sync with the day-night cycle that sets many of our bodily routines, contributing to sleep problems, obesity, and other disease risks. Alleviating this social jet lag is at the forefront of circadian biology research.

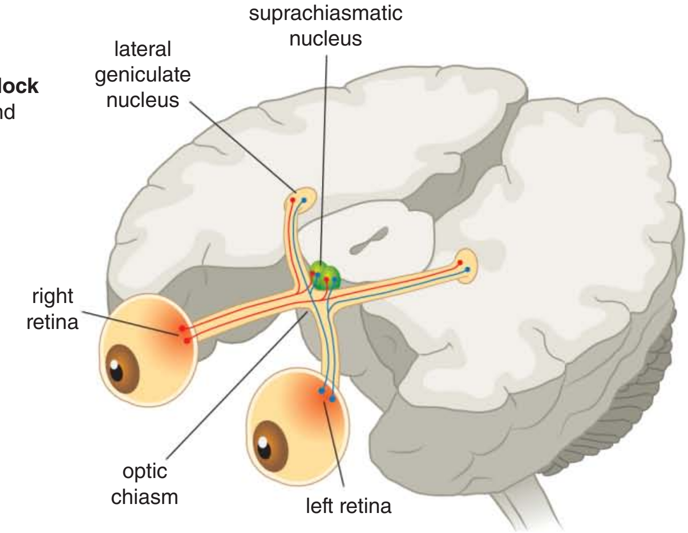
International Space Station at any time and see that the surface of the Earth looks much different than it did 100 years ago. Aside from the changes to the landmasses, oceans, and ice sheets, the area now covered by electric lights, which resemble millions of fireflies shining out into darkness, is immense. Although large parts of Africa, South America, and Asia still face night in relative darkness, North America and Europe appear as glowing lattices of cities and towns.

In contrast to the circumstances of previous generations, the end of the day no longer means the end of work and activity. Thanks to the inventions of Nikola Tesla and Thomas Edison, humans can remain active at night. One of those lights below is a hospital in Manchester, where nurses and doctors are working their rotating overnight shift. Another is an all-night fast-food restaurant in Atlanta, where employees serve burgers and greasy fries to patrons into the wee hours. More than 20 percent of the workforce in industrialized countries has irregular, rotating, or unstable shifts. Not only are people working at all hours, they have constant access to technology, information, and, well, stuff. Since the advent of Amazon.com in the 1990s e-commerce has boomed; more than half of all purchases are now made online, where we have 24-7 access to shopping. And thanks to this change toward a retail-focused economy, more people are working jobs with antisocial hours. Modern society no longer requires citizens to rest and reset. The world is nonstop, and the consequences are beginning to show.

Evolution has not caught up to these changes in day-to-day life. Humans, as well as nearly every other organism on the planet, are adapted to the 24-hour rhythm of light and darkness that arises from the spinning of Earth on its axis. Fighting against this finely tuned clock, which is responsible for coordinating essential daily processes, such as metabolism and hormone release, has enormous implications for human health.

Nurses working the night shift at a hospital have an 8 percent greater chance of developing breast cancer, and that chance continues to rise for every year that an individual works a rotating night shift. Nighttime fast-food employees and their hungry customers are fighting against the body's metabolic rhythms by eating food around the clock, including at times when the body expects to be at rest and fasting. This kind of metabolic jet lag is linked to obesity, which brings with it a whole host of additional health problems. On the whole, society is working more and sleeping less, increasing the risk for depression and anxiety issues. Such disobedience to the circadian clock, a behavioral problem arising from changes in modern living, is adding to the growing financial strain on health services worldwide. In the United States alone, an estimated $14 billion per year is spent on health care for sleep disorders, according to the U.S. Institute of Medicine Committee on Sleep Medicine and Research.





A master clock in the brain, the *suprachiasmatic nucleus* (SCN), helps keep the schedule for clocks in tissues throughout the body (*left*). The strongest signal for setting this master clock is light. Specialized nerve cells for light detection in the retina connect the eyes to the SCN, which is a pair of bilateral nuclei sitting on top of the crossing of the optic nerves at the base of the brain within the hypothalamus.

The word circadian, from the Latin *circa diem*, meaning "about a day," describes the body's near 24-hour internal clock that is set to the local environment. This clock is intrinsic; it will persist even in constant conditions that do not change, such as in a cave or deep in the ocean. Modern working environments are quite different from conditions in a cave, but in either setting the body must use information from the sensory system to tune the intrinsic circadian clock to the environment—put simply, to tell whether it is day or night.

Specialized cells in the retina, called *intrinsically sensitive retinal ganglion cells*, are primarily used for light detection rather than for image formation. These nerve cells connect the eyes to an area at the base of the brain within the hypothalamus called the *suprachiasmatic nucleus* (SCN). Scientists have studied this structure for decades—first observing that animals with lesions in this area lose all behavioral rhythms. Pioneering transplantation studies led to the conclusion that the SCN structure is both necessary for generating daily rhythms and sufficient for providing them.

On the whole, society is working more and sleeping less, increasing the risk for depression and anxiety issues.

The rhythm provided by the SCN is not exactly 24 hours in length. Depending on the organism, it can be slightly longer or slightly shorter than 24 hours, with humans falling into the former category. Light detection through the eyes sets the internal rhythm to the day-night cycle. The size of the adjustment needed is flexible; depending on the latitude and the time of year, humans experience different lengths of day and night. Therefore, our internal clock must change its relationship with the light cycle to be awake and asleep at the optimal times. This offset is otherwise known as the *phase relationship* between the body clock and the environment, signifying how internal time must speed up or slow down to fall in line with the daily experience of light and dark. An added complication for many people is that the times they need to be awake and asleep change throughout the week and may be at odds with what is happening in the environment.

Changing the body clock to the right phase relationship underlies the process of entrainment. This process is at the heart of improving circadian fitness and is likely the reason that more and more people will struggle with their health. Sunlight is the most powerful entraining signal for the body clock. But in the modern era of electricity and nonstop access to information, it's the light received via a computer screen after dinner or a smartphone screen before bed that counteracts the day-night cycle provided by the Sun. Too many constant inputs are in conflict with the daily biological cycle we evolved for.

Historically, there have been two conceptual models that explain the entrainment that takes place to synchronize the internal and external rhythms: the discrete model and the continuous one. To understand the difference, imagine two cyclists going around a circular track—they are cycling at the same speed, but there is a distance between them. This distance represents the phase difference between cyclist 1 (the environmental clock) and cyclist 2 (the internal clock). In the discrete model, a light stimulus (or other entraining signal) given at a certain point on the route around the track allows cyclist 2 to immediately overcome the distance and jump to a new point on the track, the same point as cyclist 1, so that the internal clock is in sync with the actual environment.

Essentially, according to the discrete model, light can immediately shift the internal clock to the phase of the

Biological Clocks: There's an App for That

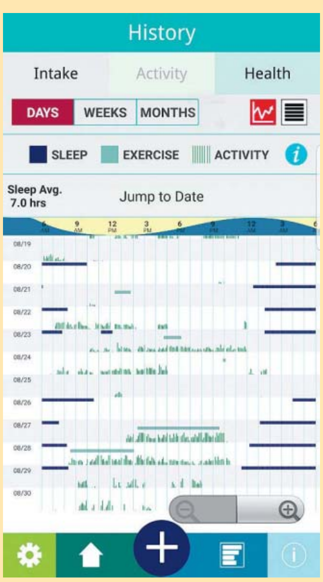
Apps and smartphone settings are helping people learn about their daily rhythms and how this information can improve their sleep and entrainment. These technologies also can be used for crowdsourcing data to study the connections between daily activities, biological clocks, and disease risks. Several groundbreaking examples are detailed here.

My Circadian Clock
Use this app to track your sleep-wake cycle, exercise, and meals throughout the day. Feedogram data help you track when you eat throughout the day; you may be surprised at how frequently across the day and night you are consuming food. Activity data can help inform when you should be restricting your meals according to your circadian cycle. By eating during a window that corresponds to when your metabolism is ramped up (during the day when your clock knows you're active), you may find it easier to maintain a healthy weight. Members of Satchidananda Panda's laboratory, who are based at University of California, San Diego, and are creators of the app, found that over the course of 16 weeks, overweight, healthy adults lost weight, improved energy levels, and slept better when eating in the restricted time window prescribed by the app. Available for Android and iOS.
https://mycircadianclock.org

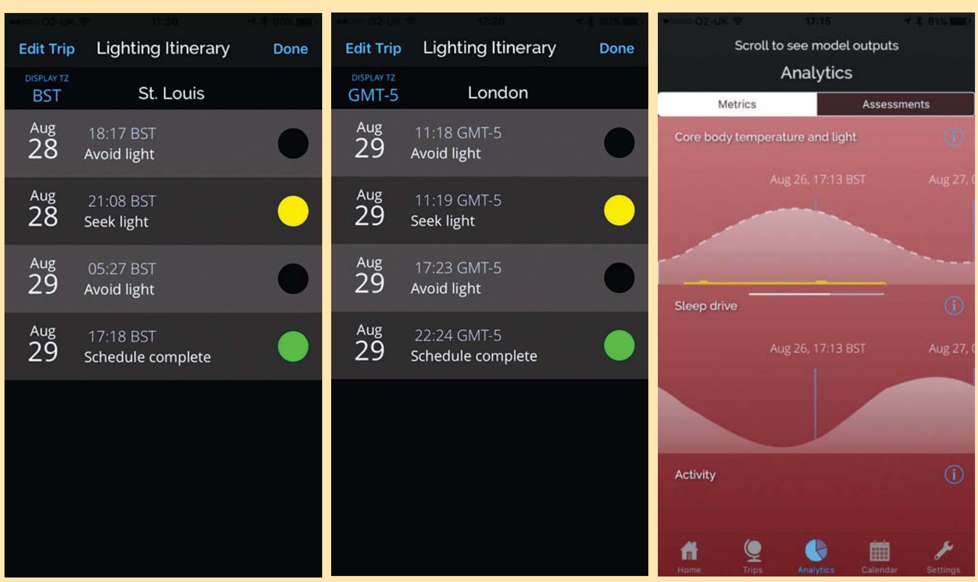
Entrain
If you are traveling across the country or around the world, this app will help shift your clock to local time faster. Enter information about your trip: where you are traveling from, what's your destination, when do you leave. The app then collects data on your sleep and activity, as well as light exposure, which is fed into a mathematical model of the circadian clock. The model uses phase response curves based on your circadian rhythm to make predictions on when you should see light before and during your trip to best enhance your entrainment. Creators of the app in Daniel Forger's laboratory, which is based at University of Michigan, found that bedtime, not wake time, predicts how long people sleep, and that those people who experience outside light during the day go to bed earlier and sleep more. Available for Android and iOS.
http://entrain.math.lsa.umich.edu

Night Shift and f.lux
[*not related to data collection*]
The Night Shift setting is part of the iPhone operating system, designed to minimize the impact of the blue-light-emitting screens on delaying the circadian clock and making it harder to fall asleep at night. Using a set time, your screen will automatically filter out blue light and enhance warmer colored light from your phone in the evenings. Similar modifications are available on some models of Android phones. A program called f.lux will also automatically adjust the brightness and color of your laptop screen based on time of day and lighting within the room:
https://justgetflux.com

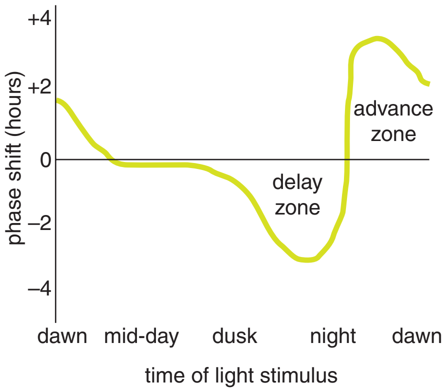
MY CIRCADIAN CLOCK



ENTRAIN



American Scientist Previous Page | Contents | Zoom in | Zoom out | Front Cover | Search Issue | Next Page



Bright light exposure after dusk has the ability to delay the internal clock, shifting one's activity later, as if the day were longer. Light exposure before dawn has the ability to advance the clock, shifting activity earlier to start a new day. This phase-response curve, a plot of the change in the internal clock relative to the external environment based on the time of light exposure, shows that the biggest phase shifts occur at two points in the night: just after dusk and just before dawn.

environmental light-dark cycle. The continuous model argues that an entraining stimulus would cause cyclist 2 to increase or decrease her speed over the subsequent cycles until she is at the same point on the track as cyclist 1. This process happens over multiple loops as the internal clock is continuously adjusting to fall in line with the environmental cycle. Although aspects of both discrete and continuous entrainment are still favorably regarded in the literature today, these two models of entrainment have failed to unify the field. Recently, an alternative theory has been proposed, which postulates that changes in how the circadian system integrates light signals at different times of day allows it to tune to the environment.

In both the continuous and discrete models, depending on what time the light stimulus arrives, the extent of the light shift varies. For example, light exposure during the day has little effect on the intrinsic clock. This idea seems intuitive—if the purpose of light is to shift the clock, then either limiting exposure to the light, or in the case of the circadian system, limiting when light has a shifting effect, would be beneficial. Plotting the change in the phase of the internal clock relative to the external environment based on time of light exposure, otherwise known as the *phase-response curve*, shows that the biggest shifts occur at two points: just after dusk and just before dawn.

Bright light exposure after dusk has the ability to delay the clock, shifting activity later, as if one were experiencing a longer day. Light exposure before dawn has the ability to advance the clock, shifting activity earlier to start a new day. Night-shift workers face light exposure that can both advance and delay their clocks; rotating-shift workers struggle between "regular" and shifted schedules, never successfully entraining to either. Models simulating phase-response curves based on data from the underlying circadian rhythm can now help frequent travelers predict when they should be exposed to light to best entrain their body clocks to their current local time (*see sidebar on preceding page*).

Although most of the world won't be attempting to intentionally shift circadian rhythms in this way, it is the transitions in light at dawn and dusk that help entrain the slightly longer internal rhythm to the 24-hour day-night cycle of the environment. Studies have shown that two light stimuli, one when diurnal activity starts and the other when it stops, provide enough time information to the circadian system to set it to the environmental rhythm.

How much light is necessary to shift and set the body clock? Although bright sunlight with an intensity measured at 100,000 lux is most efficient, recent studies by researchers at Harvard Medical School have shown that even the light emitted from the LED (light-emitting diode) screens of smartphones and tablets, when seen before bed, is enough to delay the circadian clock and disrupt sleep. Indeed, the light-sensing retinal ganglion cells that project to the SCN are particularly sensitive to the blue light from these screens.

Avoiding phone use at night to prevent body clocks from shifting is not the only option; technology has been developed that can purposefully block the blue light from being emitted (*see sidebar on preceding page*). On an overcast day, the eyes still receive around 2,000 lux. But the illumination level of indoor lighting is often much lower, and can be as low as 50 lux. Limited bright light exposure for elderly patients in care homes could be one reason for the disrupted sleep and activity rhythms observed in those populations.

	illuminance (lux)	surfaces illuminated by
colorless vision	0.0001	moonless, overcast night sky (starlight)
limited color vision	0.001 and 10	moonlight to early twilight
	0.002	moonless, clear night sky with airglow
	0.05 to 0.36	full moon on a clear night
	3.4	dark limit of civil twilight under a clear sky
color vision	10 to 50	public areas with dark surroundings
	30	tablet/smartphone screen
	50	family living room lights (Australia, 1998)
	80	office building hallway/toilet lighting
	100	very dark overcast day
	320 to 500	office lighting
	400	sunrise or sunset on a clear day
	500 to 1,000	computer screen
	1,000	overcast day; typical TV studio lighting
	10,000	therapeutic light box
pupil constriction and phase shifting	10,000 to 25,000	full daylight (not direct sun)
	32,000 to 120,000	direct sunlight

Entrainment is stronger when people are exposed to brighter light. Sunlight is ideal, because it is bright enough to cause pupil constriction and phase shifting. Most office lighting is not bright enough to entrain the body's clock. Light with a brightness of at least 10 lux is detected by both rods and cones in the eye. Light with an illuminance below 0.0001 lux allows for detection with rods only (colorless vision). The rods and some cones can detect dim light (0.001–10 lux), allowing for limited color vision.

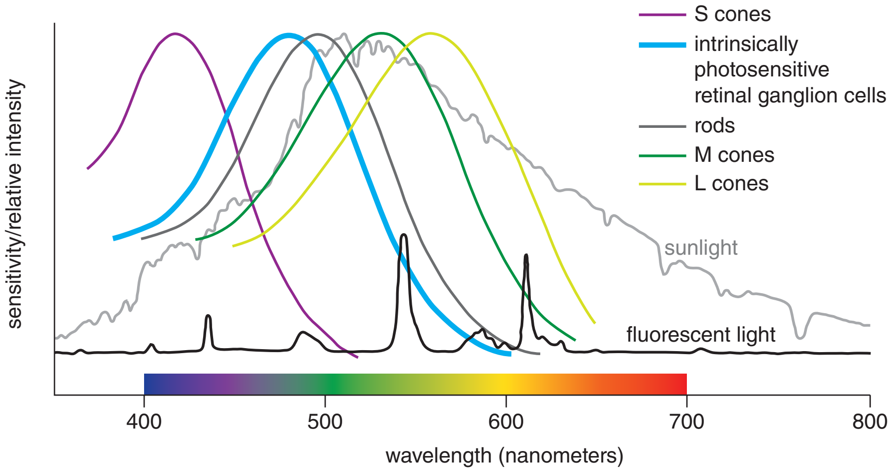
Can We Entrain Faster?

Being in tune with the environment is a boon for organisms, not only for "the early bird that gets the worm," but also for nocturnal animals who time their activity cycle to minimize predation and improve survival. As mentioned in the previous section, the strongest entraining signal for the circadian clock is light. Unlike the brains of birds, human brains cannot experience light directly through the skull and must rely instead on a specialized part of the visual system. Indeed, research over the past 20 years in humans (and mice) lacking parts of the image detection apparatus of the eye—the rods and cones—showed that they still maintain daily rhythms; that research contributed to the discovery of those specialized light-detecting ganglion cells crucial for entrainment.

Many people, such as parents of young children and business travelers, would surely benefit from an internal clock that can be set to local time faster. The nurse and the fast-food worker who are forced to work against their daily rhythms would be able to improve their health with a clock that was easier to shift and reset. Can circadian biology provide any insights?

Bad news first: There are lots of gimmicks that are not based on sound science. Probably one of the most infamous published studies in the field claimed that light exposure to the back of the knees improved entrainment following a long-haul flight—a conclusion based on a faulty correlation. There are no photoreceptors located behind the kneecap. Plenty of other proposed time-shifting quick-fixes exist that haven't been published in scientific journals, but they still are sometimes promoted in the pages of in-flight magazines or newspapers. Many of these treatments, such as therapy with melatonin or benzodiazapines, are designed to promote sleep, but they won't necessarily aid in entraining your circadian rhythm to a new external time. In addition to exposure to light, one of the other rhythmic signals that may help realign the clock faster is shifting meals to local time at the destination as soon as possible.

More recent research has examined the effects of chemicals, such as kinase inhibitors or DNA damage regulators, on the behavior of the SCN, specifically to find out whether exposure to them can speed up the adjustment to a new time. A pill that could help beat



Bluer light has a stronger effect on entrainment than other wavelengths of light. The specialized cells that detect light and send the signal to the SCN, called *intrinsically sensitive retinal ganglion cells*, contain the photopigment melanospin, which is especially sensitive to the blue part of the light spectrum peaking around 460 nanometers. Rods and cones, on the other hand, can pick up the full color spectrum and play a larger role in image formation, not light detection. People who have blindness caused by the degradation of rods and cones can still detect enough light to have a weakly entrained circadian rhythm. (Figure adapted from Hatori and Panda, 2010.)

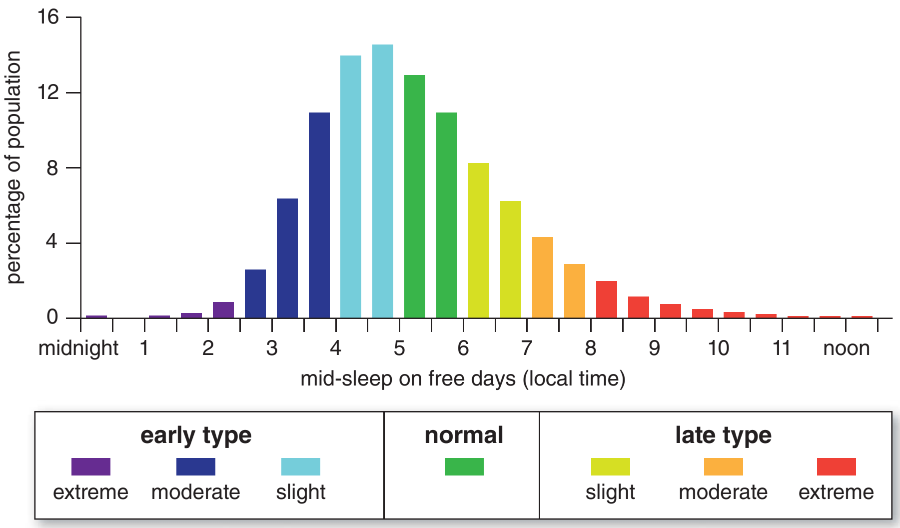
jet lag is likely years away, but studies suggest that certain drugs may be able to influence the timing of brain cells or their ability to respond to entraining signals. Human SCNs are comprised of tens of thousands of neurons, electrically active cells that can send timing information to the rest of the brain and body. These neurons are themselves rhythmic; their firing is cyclic over a 24-hour period and it is light input (and the neurochemicals released by those light-receiving cells) that influences the timing of activity.

In addition to maintaining their timing relative to the outside world, SCN neurons are also set in time with one another. Imagine a symphony of musicians. They may not play all at once, but an individual section must keep time relative to the others, with the conductor setting the overall pace of the performance. Light input communicates external timing to the conductor in the SCN, who then relays this information to the rest of the neurons, and in turn, to the rest of the body. Many studies have tried to identify the roles of different areas and cells types in the SCN and to elucidate how the symphony comes together.

When neurons in the SCN fire, a neuropeptide, *vasoactive intestinal polypeptide* or VIP, is released. Under normal conditions, VIP plays an important role in keeping the cells of the SCN synchronized to the same beat. Researchers in one of our (Herzog's) labs have shown that high concentrations of VIP, however, can throw the SCN symphony into disarray. This lack of synchrony in the rhythms may be beneficial when it comes to speeding up entrainment.

In addition to light, one of the other rhythmic signals that may help realign the clock faster is shifting meals to local time at the destination as soon as possible.

To explain how this role might work, we collaborated with computational scientists, who built a mathematical model of the circadian rhythm in an



The Munich Chronotype Questionnaire places individuals on a frequency distribution based on the midpoint at which they prefer to be asleep on an ideal day (when they are not sleep deprived and wake without an alarm). Most people's preferred midpoint of sleep is between 5 AM and 6 AM. This survey has been important in collecting anecdotal evidence about people's sleep habits. Now, with activity trackers collecting information through smartphones and wearables, the scale of raw data available on people's daily behaviors is enormous, with the potential to transform what is known about disease risks and solutions related to out-of-sync biological clocks. (Figure adapted from Roenneberg and Merrow, 2007.)

SCN neuron. Our model predicted how desynchrony could make entrainment happen faster. In the model, neurons responded to high doses of VIP with a behavior described as "phase tumbling." Essentially, lots of VIP produces a transient phase decoupling across the cells, shifting them in time relative to one another. This dispersion likely occurs due to molecular variability or noise in the levels of gene activity and protein amounts in individual cells—evidence our lab has also observed. In the model, the shifted cells are also quicker to readjust and entrain to new times, suggesting that high VIP levels might help alleviate jet lag.

To test this idea, researchers in our lab recently measured the rhythmic activity of mice provided with running wheels. They simulated jet lag for the mice by shifting the light schedule to either advance or delay the environmental signal by six hours relative to the previous schedule. Researchers examined when the mice would be active on their running wheels following the shift. Mice that received injections of high concentrations of VIP in the area around their SCN were able to entrain faster to the new schedule compared with control mice that received injections of saline instead. Because VIP is released from the neurons that receive input from light-sensing cells in the eyes, a good dose of bright sunlight may help trigger phase tumbling in our SCN and hasten readjustment to a new time.

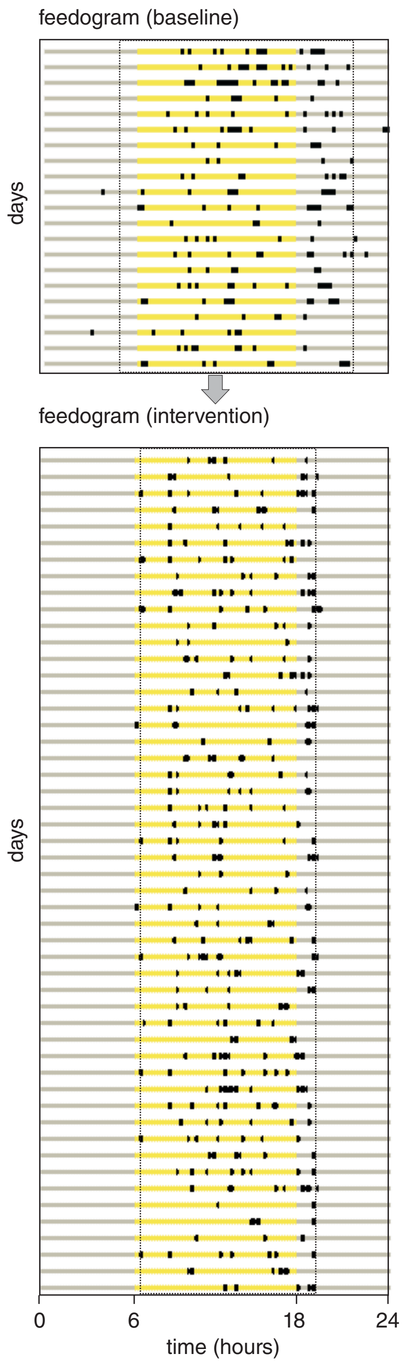
Researchers have also been examining whether the introduction of other compounds could improve entrainment. Screens of libraries of small molecules, including primarily synthetic compounds that can bind to genes and proteins that could produce or lead to enhanced shifts, is a high-throughput way of identifying candidates. A future for improving entrainment may rely on a combinatorial approach. By using light together with drugs, timed meals, and other stimuli, humans may be able to better tune and enhance the phase relationship between their internal and external timing.

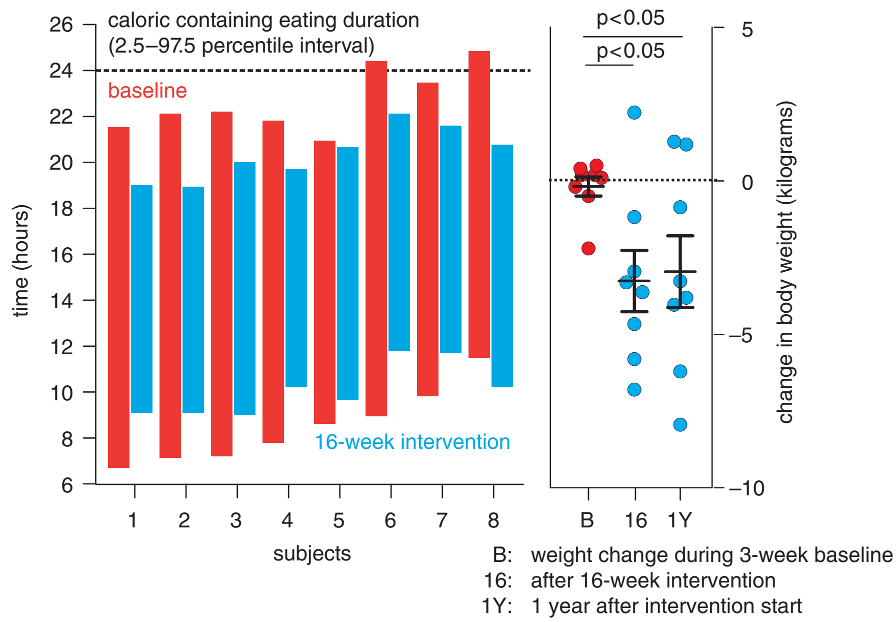
Technology to Improve Entrainment

The past 10 years have seen an explosion in constant connectivity by way of mobile devices and other technologies. In addition to listicles and cat pictures, the internet now contains vast amounts of data generated by the population, including information about sleep and wake patterns, food intake, and location. This era of quantified self, collected through activity trackers and apps, means that there are more data than ever before about people's daily lives. Some fast-food workers, for example, might have a Fitbit that measures their steps, exercise, sleep, and heart rates. They could use this wealth of information to make better and more informed decisions about their daily behavior. Activity trackers encourage wearers to think about how much and what they do on a daily basis—and also about *when* they are doing each of their routine activities. Those who want to learn more about their circadian behavior now have access to apps and websites offering more information. Scientists also are utilizing the internet to connect with other members of the public who would like to contribute to research. These citizen scientists can use mobile technology to generate and share data with researchers, essentially turning the world into a giant laboratory.

Developers of research-based apps have empowered people to become part of the scientific process of understanding their biology. Using an app to track behavior can provide a window into daily life in the real world, without the constraints of a controlled environment in the lab. These sorts of data are particularly useful to observe the changes seen in 21st-century lifestyles compared with lifestyles of 50 or even 100 years ago. Large-scale surveys such as the Munich Chronotype Questionnaire, whose questions place individuals across a distribution based on the midpoint of when they'd prefer to be asleep, have been important in collecting anecdotal evidence about the when, where, and how of people's sleep habits. But with activity trackers, the scale of raw data that we have about people's daily behavior from around the world is enormous. Apps that track sleep and wake times, food intake, and other aspects of biological rhythms, such as My Circadian Clock and Entrain, have already provided new information about obesity and recovery from jet lag (*see sidebar on page 349*). Now and in the near future, researchers could use such apps in combination with other personal data, such as medical records and genetic predispositions, to paint a more detailed picture of how the circadian clock contributes to overall human health as well as disease risk.

The challenge now is figuring out how to spread the word to people about the effects of a healthy circadian rhythm on conditions such as obesity, cancer, and mental health, and how best to encourage more people to examine the data they generate to





Shubhroz Gill and Satchidananda Panda created the app My Circadian Clock to study the timing of food intake and its effects on weight loss among healthy individuals. They first monitored the times that study subjects ate food without any intervention and found that people tended to eat frequently and erratically throughout the day and consumed more daily calories in the afternoon and evening (*top left*). When the app coached the same individuals on the time window in which to eat, so that they stopped any food intake earlier in the evening after peak circadian activity (*bottom left*), subjects overall lost weight (*above*).

discover how their rhythms could be improved. Until better solutions for circadian dysfunction exist, those of us prone to late-night Netflix streaming or scrolling through our Instragram feed before we turn out the light can take steps to be healthier by acknowledging our circadian rhythms and trying to minimize disruption to our clocks.

A couple of straightforward changes can make a marked difference in alleviating entrainment problems: removing all technology from the bedroom and ensuring that some time is spent outdoors every day. An instantly adjusting body clock, aided by wearable technology and apps, may someday be possible. But achieving it will require more research to understand the ways in which biology and the rhythms of modern life can work together in harmony.

Bibliography

An, S., et al. 2013. A neuropeptide speeds circadian entrainment by reducing intercellular synchrony. *Proceedings of the National Academy of Sciences of the U.S.A.* 110(46):E4355–E4361. doi:10.1073/pnas.1307088110.

Chang, A.-M., D. Aeschbach, J. F. Duffy, and C. A. Czeisler. 2015. Evening use of light-emitting eReaders negatively affects sleep, circadian timing, and next-morning alertness. *Proceedings of the National Academy of Sciences of the U.S.A.* 112(4):1232–1237.

Chen, Z., S.-H. Yoo, and J. S. Takahashi. 2013. Small molecule modifiers of circadian clocks. *Cell Molecular Life Sciences* 70(16):2985–2998.

Dunlap, J. C., J. J. Loros, and P. J. DeCoursey. 2009. *Chronobiology: Biological Timekeeping.* Sunderland, MA: Sinauer Associates.

Foster, R., and L. Kreitzman. 2017. *Circadian Rhythms: A Very Short Introduction.* Oxford, UK: Oxford University Press.

Gill, S., and S. Panda. 2015. A smartphone app reveals erratic diurnal eating patterns in humans that can be modulated for health benefits. *Cell Metabolism* 22(5):789–798.

Hatori, M., and S. Panda. 2010. The emerging roles of melanopsin in behavioral adaptation to light. *Trends in Molecular Medicine* 16:435–446.

Herzog, E. D. 2007. Neurons and networks in daily rhythms. *Nature Reviews Neuroscience* 8:790–802.

Lucas, R. J., et al. 2013. Measuring and using light in the melanopsin age. *Trends in Neurosciences* doi:10.1016/j.tins.2013.10.004.

Roenneberg, T. 2017. *Internal Time: Chronotypes, Social Jet Lag, and Why You're So Tired.* Cambridge, MA: Harvard University Press.

Roenneberg, T., and M. Merrow. 2007. Entrainment of the human circadian clock. *Cold Spring Harbor Symposia on Quantitative Biology* 72:293–299.

Schibler, U., et al. 2015. Clock-talk: Interactions between central and peripheral circadian oscillators in mammals. *Cold Spring Harbor Symposia on Quantitative Biology* 80:223–232.

Walch, O. J., A. Cochran, and D. B. Forger. 2016. A global quantification of "normal" sleep schedules using smartphone data. *Science Advances* 2(5):e1501705. doi:10.1126/sciadv.1501705.

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

www.amsci.org/magazine/issues/2017/november-december

Cassini and the Rings of Saturn

As one of its final scientific feats, the accomplished robotic spacecraft returned unprecedentedly detailed data on the material encircling the giant planet.

Matthew S. Tiscareno

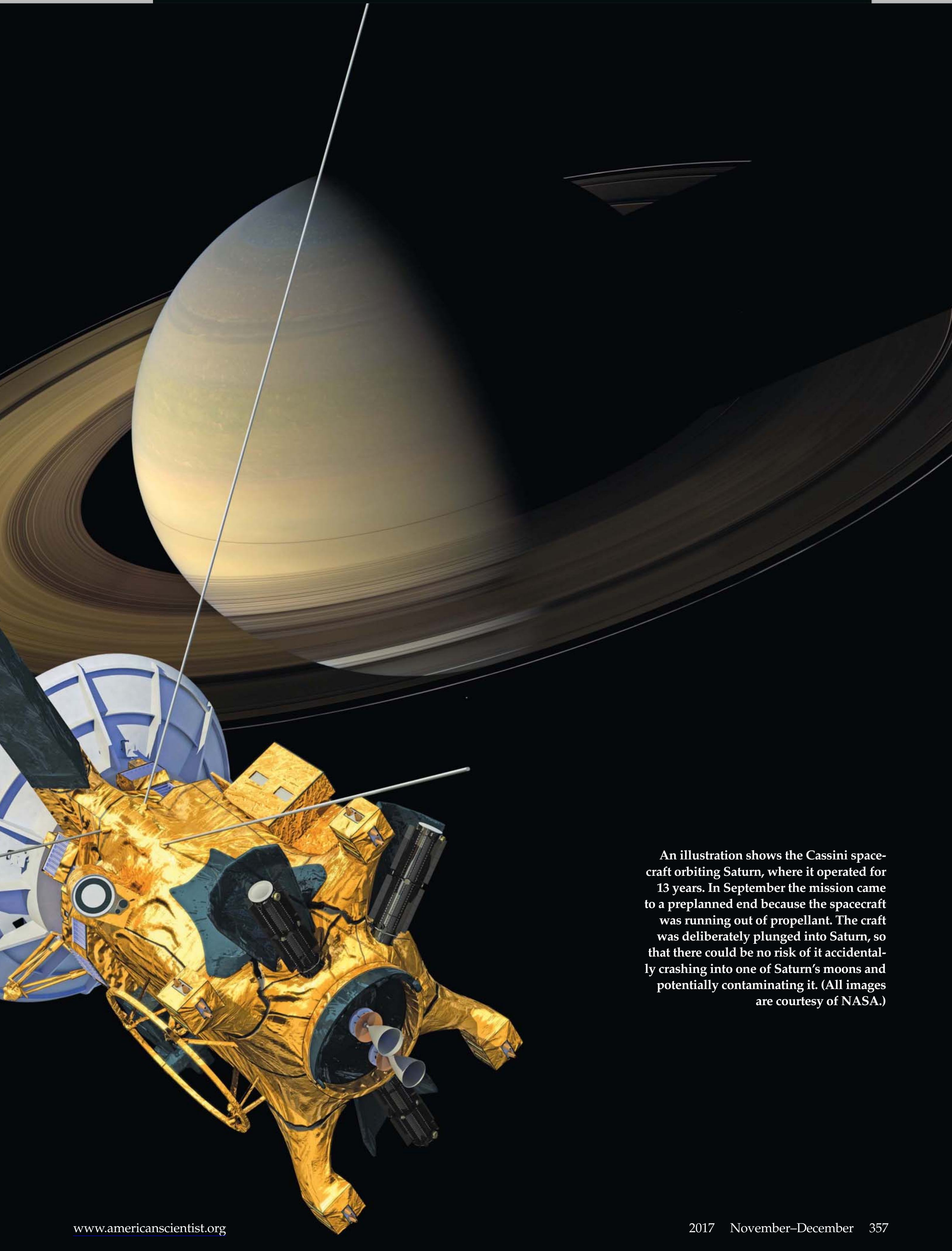
The rings of Saturn are delicate, enigmatic, beautiful, and useful. NASA's Cassini spacecraft has been studying them (and the rest of the Saturn system) over the course of a spectacularly successful 13-year mission, but its concluding year has been the most spectacular of all—practically a whole new mission. In December 2016, Cassini commenced weekly plunges through the ring plane just off the outer edge of the main rings (activity termed the "Ring-Grazing Orbits"), and transitioned in April 2017 to weekly plunges between the rings and the planet's cloud-tops (the "Grand Finale" orbits). The mission concluded in September 2017 with a final descent into the planet's atmosphere, to preclude possible terrestrial contamination of Saturn's moons.

The science goals of the Ring-Grazing Orbits and the Grand Finale orbits included direct sampling of particles from Saturn's rings and atmosphere, detailed measurements of Saturn's gravity and magnetic field to probe the planet's interior, and unprecedentedly close-range imaging of Saturn and its rings with the spacecraft's main camera and similar instruments. These observations followed up on discoveries made through the prime and extended Cassini missions, continued monitoring seasonal or evolving phenomena, and took advantage of Cassini's unique proximity to Saturn in these closing stages of the mission.

Saturn's planetary ring system is made up of countless chunks of ice, each in its own orbit around the planet Saturn. The chunks range from marble-size to house-size. The ring system is confined to the plane of Saturn's equator and is arguably the flattest structure known to humanity, with an end-to-end dimension equivalent to circling the Earth seven times, but a vertical height about that of a house.

Saturn's rings are useful to scientists because they provide physical and chemical clues regarding the formation and history of the entire Saturn system, they serve as detectors and amplifiers for planetary phenomena around them, and they help us understand more generally how disk systems operate, providing clues about other kinds of disks, such as baby solar systems. A few of the most compelling science questions regarding Saturn's rings are: How do ring particles interact with one another, with moons embedded within them, and with moons farther away? Are ring particles made of ice that is fluffy or dense, pristine or sooty? What are their shapes and sizes? What structures do we see in the dusty parts of the rings, and what can we continue to learn from them?

Matthew S. Tiscareno is a senior research scientist at the SETI Institute in Mountain View, California, and a participating scientist and imaging team associate for the Cassini spacecraft. He received his PhD in planetary science from the University of Arizona in 2004. His research focuses on the orbital and rotational motions of rings, satellites, and planets. He is the coeditor of Planetary Ring Systems *(Cambridge University Press, 2018). Email: matt@seti.org*



An illustration shows the Cassini spacecraft orbiting Saturn, where it operated for 13 years. In September the mission came to a preplanned end because the spacecraft was running out of propellant. The craft was deliberately plunged into Saturn, so that there could be no risk of it accidentally crashing into one of Saturn's moons and potentially contaminating it. (All images are courtesy of NASA.)

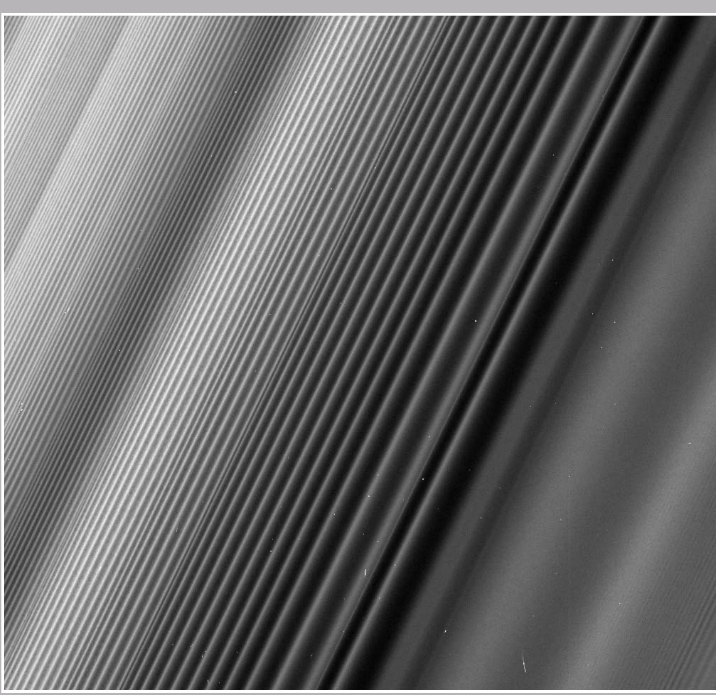
The structure of Saturn's ring system is dominated by concentric bands and tightly wound spirals. Very few of these bands are empty gaps, so it is better to think of the entire system as a disk, rather than as "countless rings." Much of the structure is not well understood by scientists, but the most common type of understood structure are waves that propagate through the rings at locations where ring particle orbits *resonate* (or essentially "hum in tune") with the orbit of a moon orbiting beyond the rings. Cassini scientists discovered in recent years that a few of these waves are excited by structures within the planet Saturn.

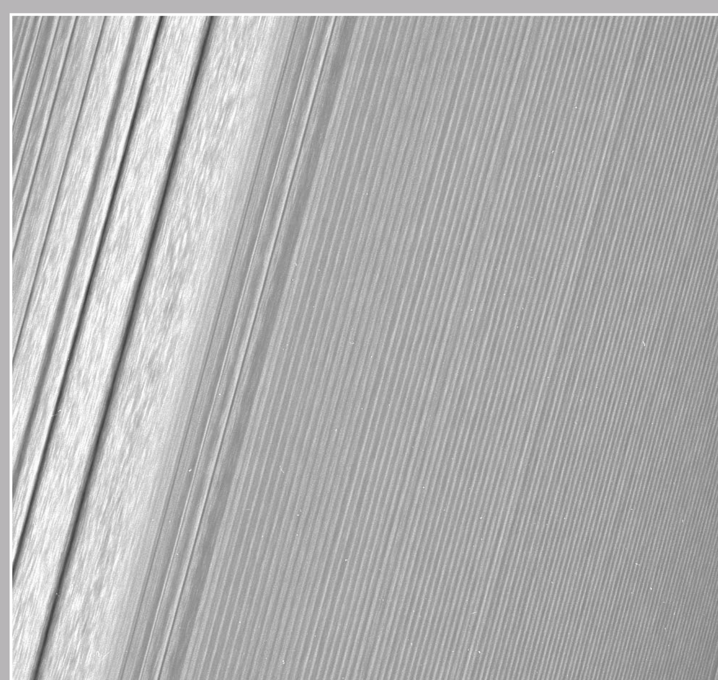
Moons that orbit within the rings deflect nearby ring particles with their gravity and thus create a disturbance around themselves. The moons Pan and Daphnis are large enough that this disturbance becomes a sharp-edged empty gap extending around the entire circumference of the rings. Extrapolating from this archetype, scientists expected each of the dozen other sharp-edged gaps in the rings to also host a moon apiece; however, careful searches by Cassini have shown that this is not the case, so these gaps must have a more complex origin. Moons in a smaller size class than Pan and Daphnis create a propeller-shaped disturbance that is filled back in by the rings before it can extend circumferentially.

The interior and surface properties of ring particles are poorly known. Are they more like ice cubes or snowballs? Are their surfaces fluffy or slushy or frosty? The answer may be all of the above. Images from the Ring-Grazing Orbits and Grand Finale have shed new light on the collective interactions of ring particles. A speckled strawlike texture occurs within waves at locations where the particles had been recently compressed but that pressure has now been released, perhaps as a result of temporary clumps falling apart. More surprisingly, sharp-edged bands of similarly speckled texture occur at many locations within the rings, as well as streaky texture and other types of texture. Each of these textures likely indicates a different mode of mutual interaction among the ring particles, and/or different ring particle properties.

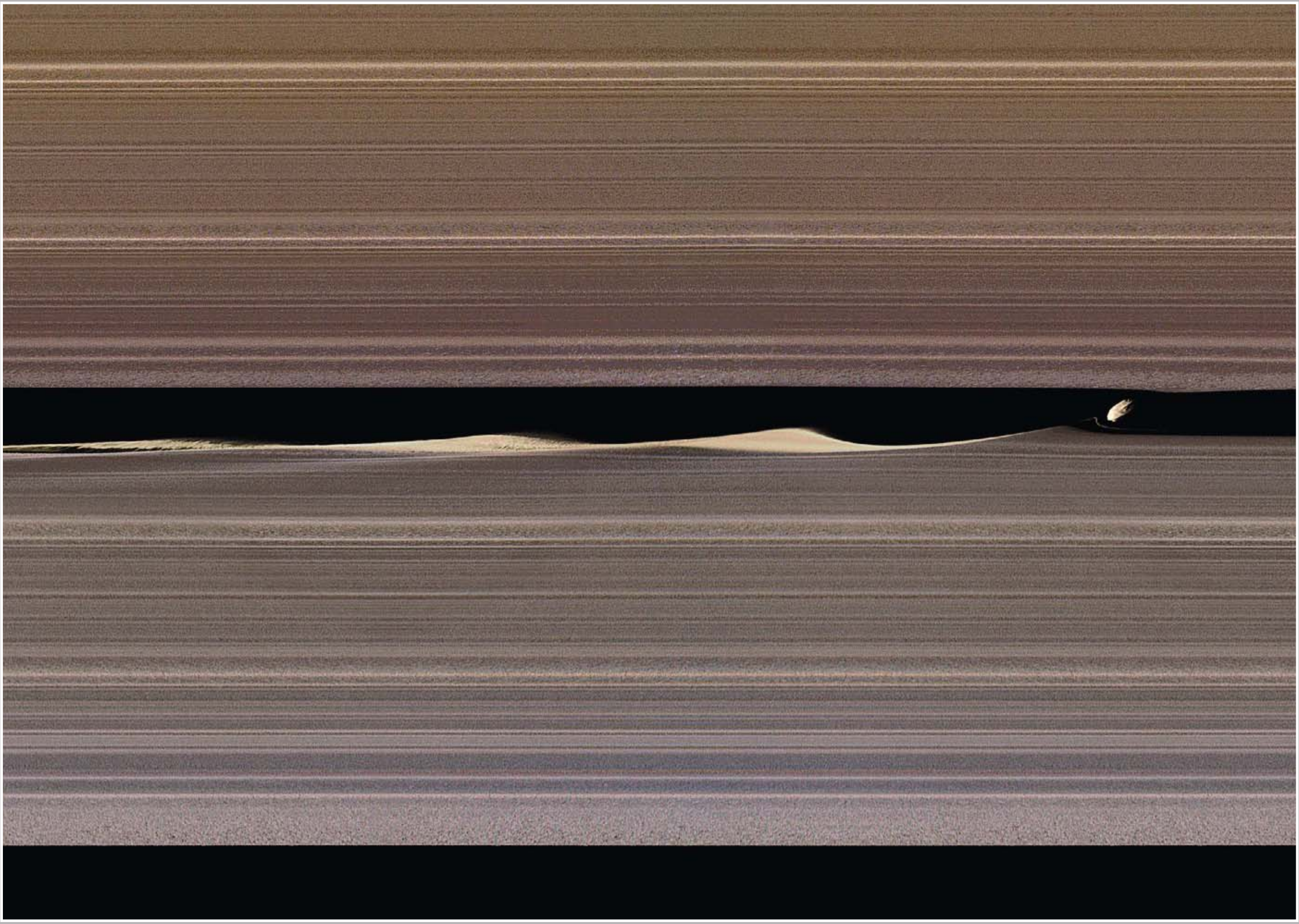
The main rings are largely free of dust (that is, of icy particles that are smaller than the period at the end of this sentence), because dust tends to collect as a layer on the surfaces of larger ring particles. In the few places where dust does coexist with larger particles, it indicates the presence of vigorous ongoing activity that is keeping the dust in motion. In locations that are far from the main rings, only dust is present, and the dust then can be sculpted into shapes that reflect the forces surrounding it.

Perhaps the most fundamental mystery surrounding Saturn's ring system is its age. Certain aspects of the dynamical interactions between the rings and moons can only be rewound for about 100 million years, indicating either that a significant reworking of the ring system's structure occurred within that time, or that the entire ring system is only that old. An important clue in this mystery is the rate at which the rings are being polluted by "soot" falling in from the Solar System. If that infalling rate is low, or if the rings have a relatively large mass of water ice with which to mix the pollution, then it would be reasonable to conclude that the rings are as old as Saturn and still sport the relatively pure water ice we see them to have. However, if the infalling rate of pollution is high and the mass of the rings is low, then the rings are likely no older than 100 million years. Cassini's dust detectors were working on a definitive measurement of the pollution rate, and the mass of the rings will be measured directly from the gravitational pull of the rings on Cassini during the Grand Finale orbits. After analysis of the final Cassini data has been completed, these questions should find more concrete answers.

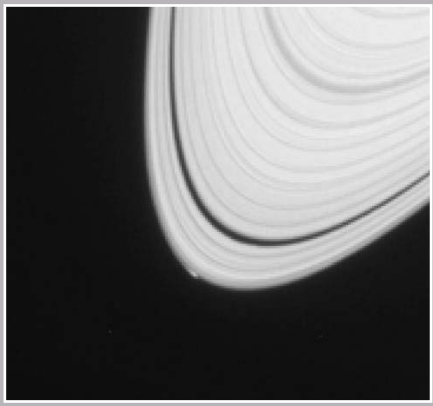




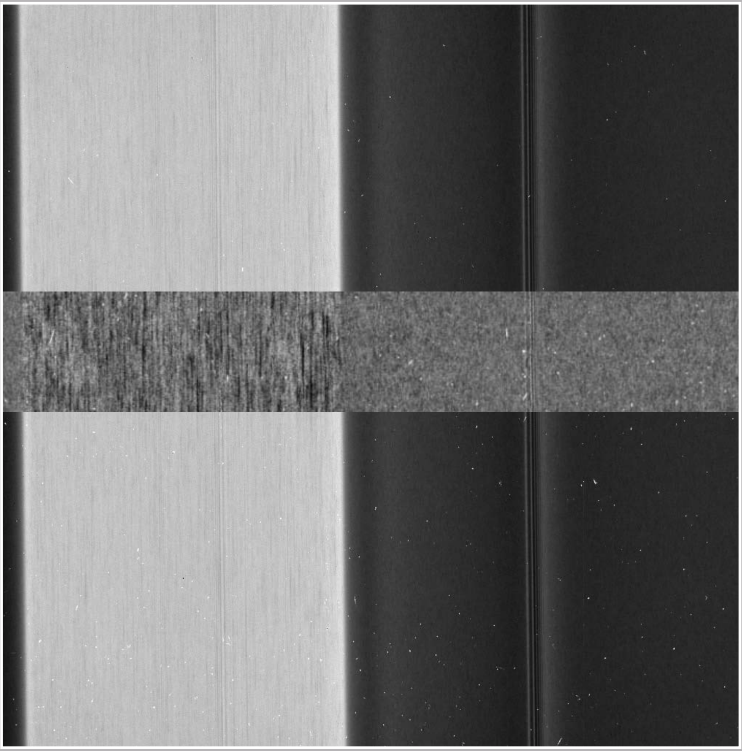
The *spiral density wave* at left occurs at the location where ring particles orbit Saturn twice for every time that the moon Janus orbits once (called a *2-to-1 resonance*). Because Janus trades orbits with its partner moon Epimetheus every four years, this wave sports several "glitches" that trace the wave's history going back several decades. The spiral density wave at right is also caused by the gravitational pull of the moon Janus, but at a location where ring particles orbit Saturn four times for every three Janus orbits (a *4-to-3 resonance*). The peaks of the wave are dark, and the broad, bright troughs contain a speckly, strawlike structure, likely due to transient clumps of ring particles that are now falling apart. The corduroy-like structure at center and right is a result of a recent pass of the moon Pan.



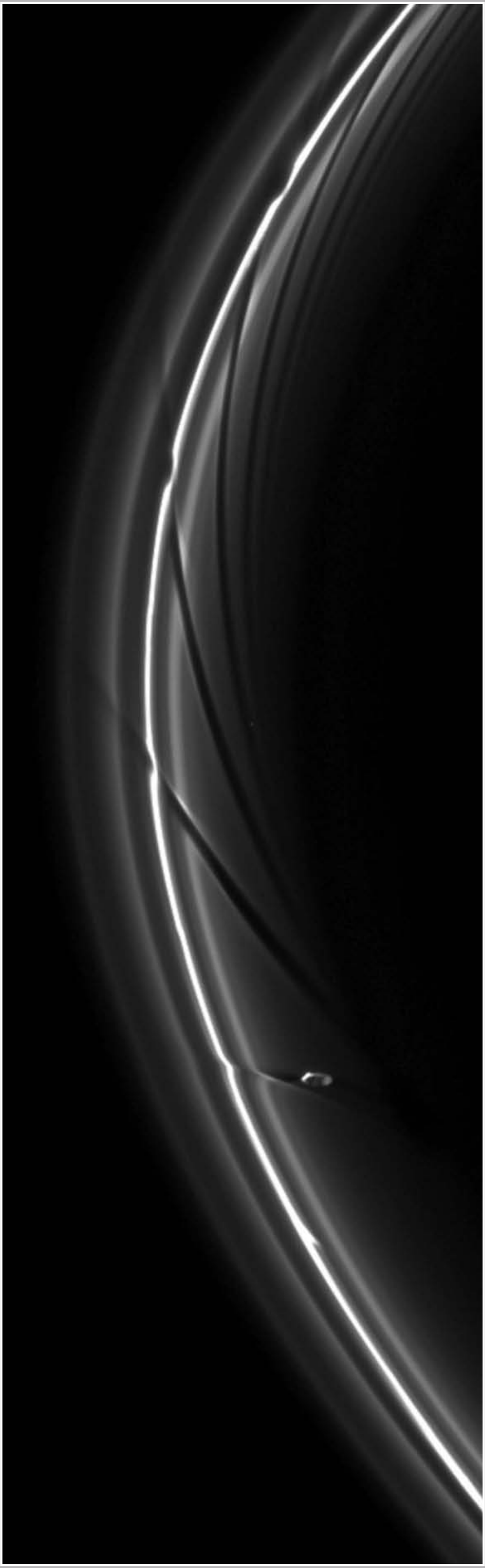
The moon Daphnis is seen orbiting within its sharp-edged gap. Ring particles that have recently passed the moon are perturbed into wavelike structures in the gap edge, but Daphnis's influence is waning by the third wavecrest back, which is disintegrating into a collection of small clumps.



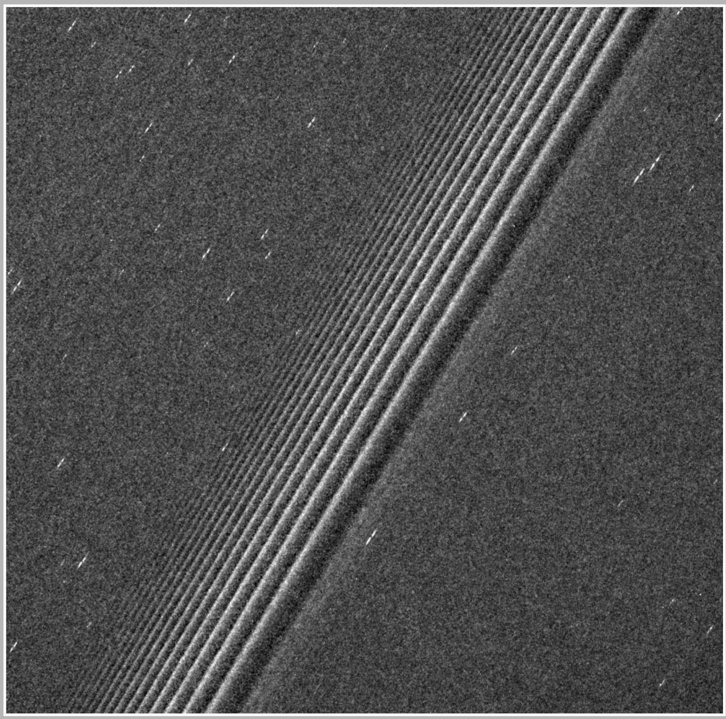
The quasi-moon "Peggy" appears at the outer edge of the main rings, looking as though it might spiral away from the rings and become a freely orbiting moon. Subsequent Cassini images have shown Peggy's story to be more complicated, because it has broken into several pieces and has remained within the rings.



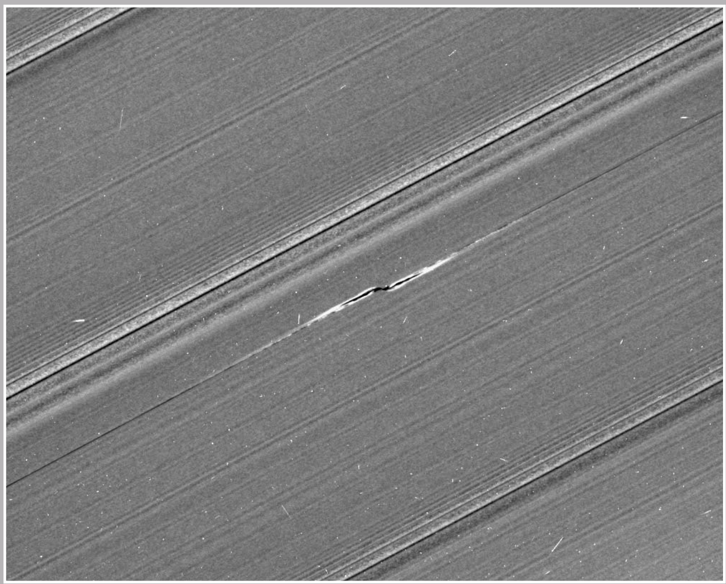
At left, a sharp-edged bright band (called a plateau) has an unknown cause. The central strip, from which the first-order brightness has been removed so that local structure becomes more visible, shows that the plateau has a streaky texture whereas the surrounding region has a speckly texture. The different textures must indicate different modes of particle properties and/or interactions. Also visible to the right in this image is a spiral density wave driven by structure inside the planet Saturn.

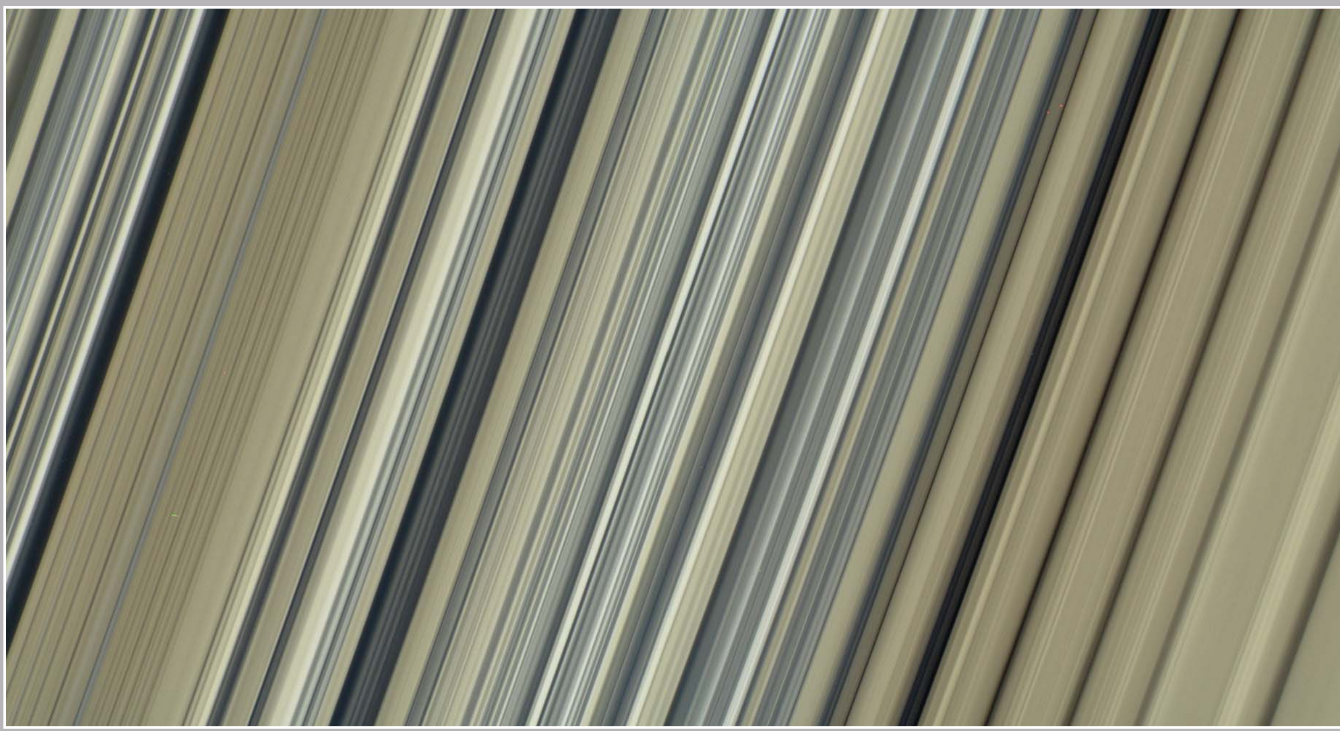


The moon Prometheus dips into the dusty F ring and sculpts its material. The resulting channels flatten out as they move downstream from Prometheus. Also visible in this image are several strands of dust comprising the F ring, with some bright knots that indicate embedded moons.

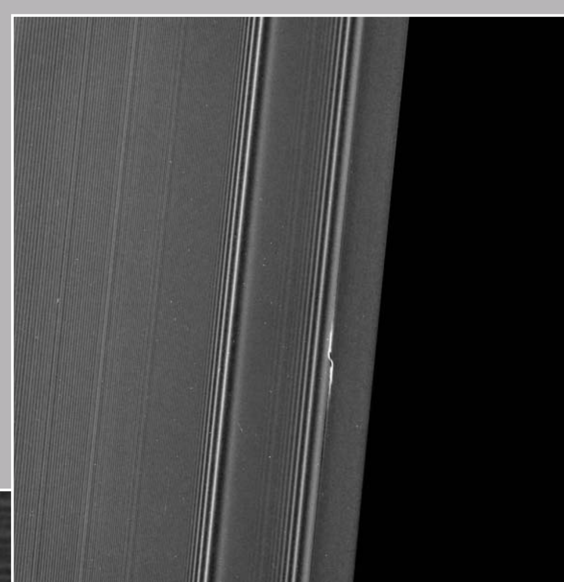


Swarms of small *propellers* of debris (*above*) surround a density wave (with a 9-to-8 resonance with the moon Prometheus). Each propeller contains a 100-meter moonlet at its core, which creates the disturbance around itself.

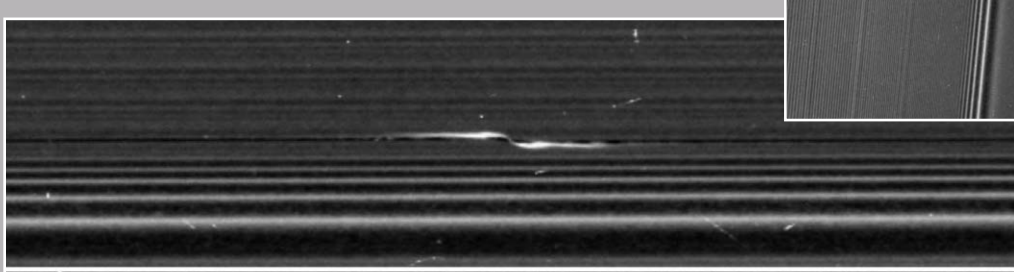


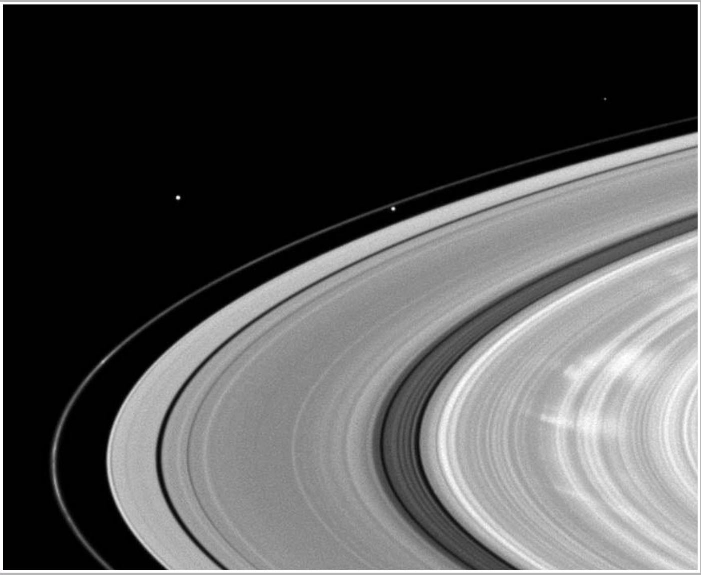


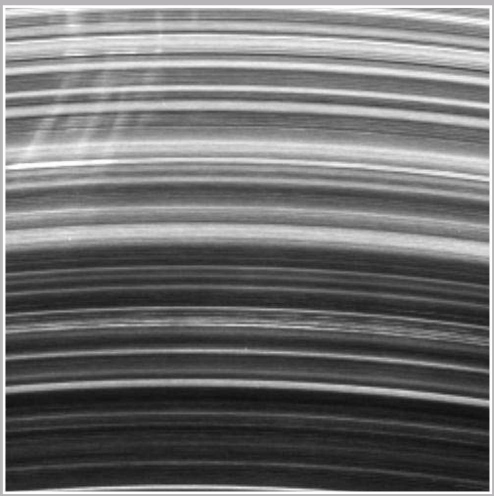
An inner-central part of the planet's B ring is shown in what is the highest-resolution true-color image of any part of Saturn's rings produced to date. The reason for the tan color is poorly understood, as is the origin of the sharply defined banded structure.



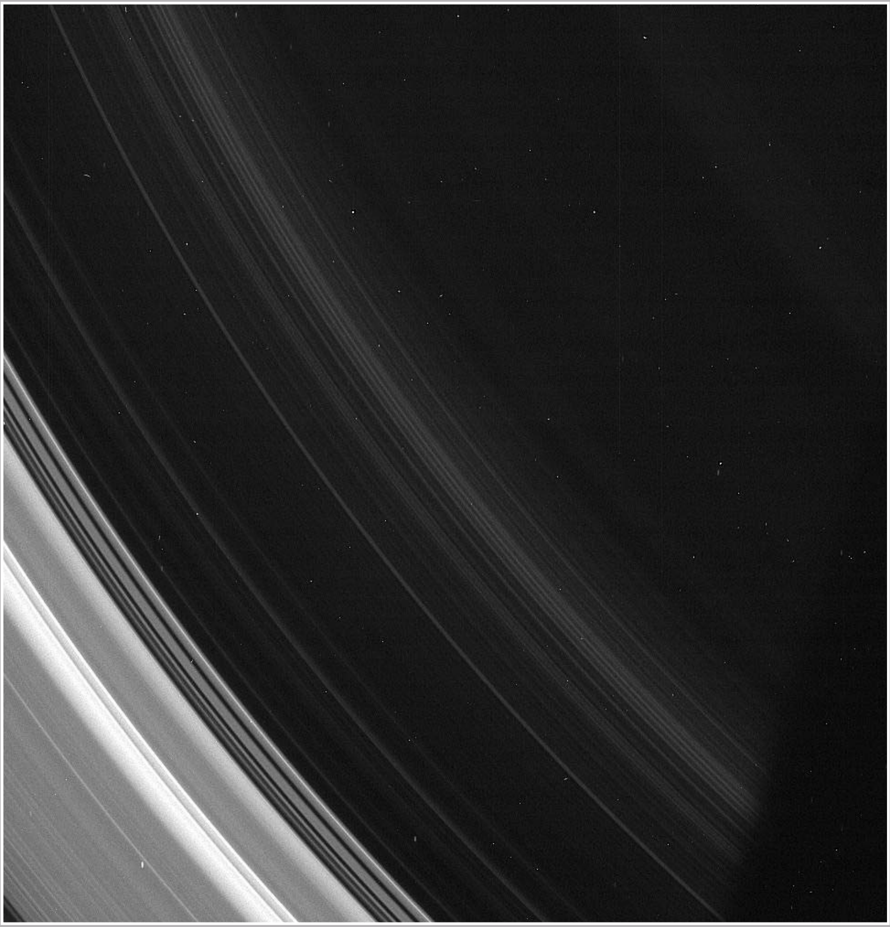
Close-up views show giant *propellers* caused by ring-embedded moons. The figures immediately below and to the right show the lit side of the rings, where brighter tones indicate more material reflecting light. The figures at far bottom and on the facing page bottom show the translucent, unlit side of the rings, where most dark tones indicate more material, such that the ring becomes opaque. Cassini has tracked the orbits of these individual propellers for more than a decade and has found subtle changes in the orbits that are likely due to moon-disk interactions. These individual propellers have been nicknamed "Santos-Dumont" (*below*), "Earhart" (*right*), and "Blériot" (*facing*).



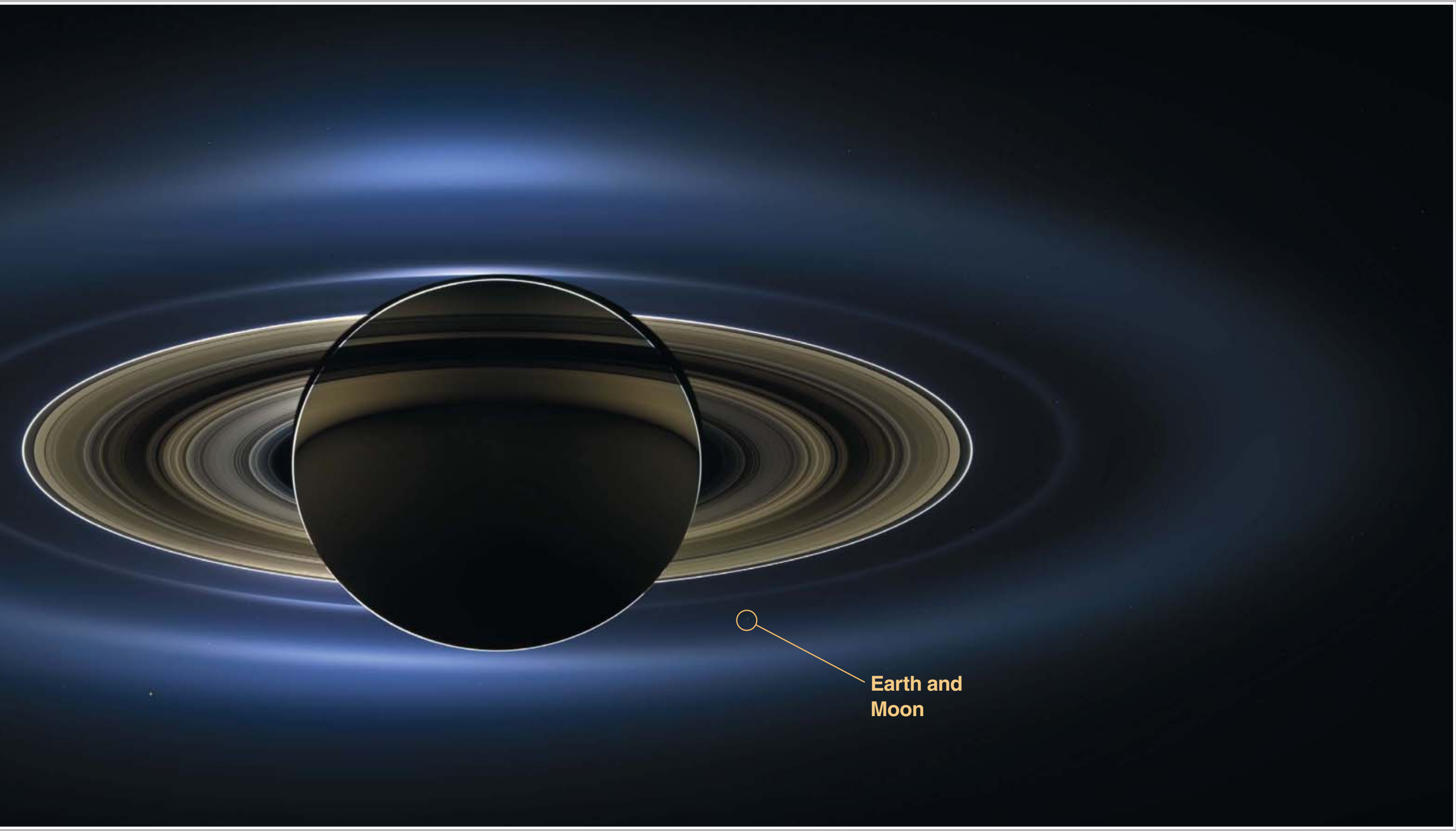




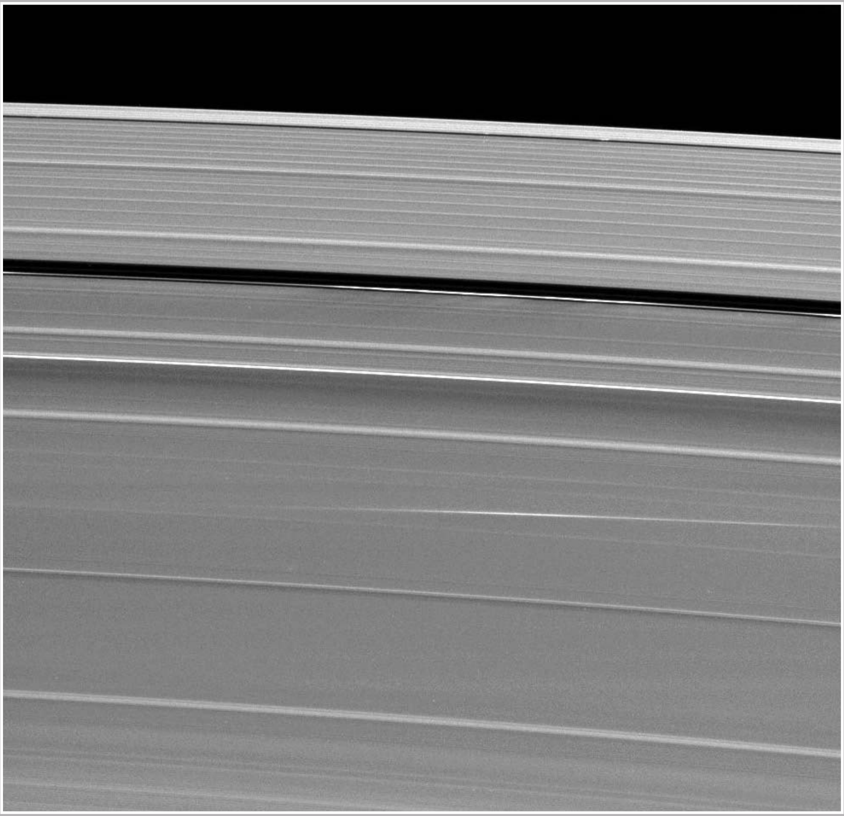
Ghostly white markings called *spokes* appear in the central region of the main rings, in a wide-angle image (*above*) and in a more detailed image (*left*). These spokes are composed of dust levitating above the ring plane. The dust was likely ejected by meteoroid impacts and then caught up into Saturn's magnetic field, although the details are not well understood.



Dusty lanes occur within the D ring, the innermost ring of Saturn. Some of the bands comprise a spiral that preserves traces of an inferred meteoroid impact that took place in the 1980s. The brighter bands at bottom-left are the innermost portion of the main rings.



A panoramic view of Saturn shows both its main rings and its outer, dusty rings. Because the Sun was behind Saturn when this image was taken (that is, Cassini was inside Saturn's shadow), the dusty rings appear nearly as bright as the main rings, and dusty regions that are more closely aligned with the Sun appear even brighter. The bluish E ring is composed of material ejected from the geysers on the south pole of the moon Enceladus. The Earth and its Moon are visible to the lower-right of Saturn.



An impact ejecta cloud tells the tale of a meter-sized object that plowed through the rings some hours before this image was taken. The ejecta cloud is near the center of the image and is nearly horizontal, canted at an angle from the concentric bands of the ring structure.

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

www.amsci.org/magazine/issues/2017/november-december

American Scientist Previous Page | Contents | Zoom in | Zoom out | Front Cover | Search Issue | Next Page

The New Language of Mathematics

Is it possible to take all words out of mathematical expressions?

Daniel S. Silver

Something strange is happening in mathematics seminar rooms around the world. Words and phrases such as *spider, birdtrack, amoeba, sandpile,* and *octopus decomposition* are being heard. Drawings that resemble prehistoric petroglyphs or ancient Chinese calligraphy are being seen, and are being manipulated like the traditional numerals and symbols of algebra. It is a language that would have been alien to mathematicians of past centuries.

The words and symbols of mathematics are intended to stimulate thought, promote curiosity, or simply amuse. At times they ignite public imagination. Occasionally they interfere with understanding. Always they are evolving. Today, as the boundaries of mathematical inquiry expand, their evolution seems to be accelerating. The words and symbols of mathematics have helped bring the subject to its present, bountiful state. But the question remains: Can the symbols of mathematics stand up on their own, without any words to support them?

Is Mathematics a Language?

Josiah Willard Gibbs navigated confidently in a sea of mathematical words and symbols. Gibbs was a founder of statistical mechanics and a professor of mathematical physics at Yale University during the latter half of the 19th century. Colleagues knew this outwardly plain and unassuming scholar as someone who rarely made public pronouncements. Imagine their surprise when, during a faculty meeting about replacing mathematics requirements for the bachelor's degree with foreign language courses, Gibbs rose and forcefully declared: "Gentlemen, mathematics *is* a language."

Gibbs wasn't the first notable scientist to call mathematics a language. Galileo Galilei beat him to it by more than 200 years. In *Il Saggiator* (The Assayer), published in Rome in 1623, the Italian astronomer wrote: "[The universe] cannot be read until we have learnt the language and become familiar with the characters in which it is written. It is written in mathematical language."

Galileo wrote in Italian rather than scholarly Latin, hoping to reach readers who were literate but not necessarily scientific. But just as students at Plato's Academy were reputedly greeted with the warning "Let no one ignorant of geometry enter here," so Galileo's readers were being cautioned that the book before them had some language prerequisites. For Galileo, the letters of mathematics were triangles, circles, and other geometric figures. Gibbs, who is responsible for much of the vector calculus that we use, would have added a modern symbol or two of his own.

If mathematics is a language, then just as any ordinary language, such as French or Russian, does not rely on another one to be understood, so mathematics should be independent of ordinary languages. The idea does not seem so far-fetched when we consider musical notation, which is readable by trained musicians everywhere. If mathematics is a language, then we should be able to understand its ideas without the use of words. Let's see how that might be done.

Consider the task of adding the first few integers, say 1, 2, 3, 4, 5. Easy enough. Their sum is 15. But what about adding the first 100 integers?

The figure on page 366 displays 1+2+3+4+5 dots twice, once in black and again in red. By arranging all in rectangular fashion as we have, it is an easy matter to count the total, which is $5\times6 = 30$. In order to find our original sum, we need only divide by 2, thereby correcting for our double counting.

The novelty of the picture is that we can grasp the idea at a glance. Moreover, there is nothing special about 5 columns of dots. We could just as easily imagine 100. So $1+2+\ldots+100$ must be equal to 100×101 divided by 2, which is 5,050. Similarly we can find the sum of the first N integers, for any N whatsoever. The answer is $N(N+1)/2$.

Just as the 19th-century musician Felix Mendelssohn enjoyed writing *Lieder ohne Worte* (Songs Without Words), so mathematicians like to craft *Proofs Without Words.* Since 1975 the Mathematical Association of America has published a column devoted to them in its *Mathematics Magazine.* Its examples are intended to leave the reader speechless.

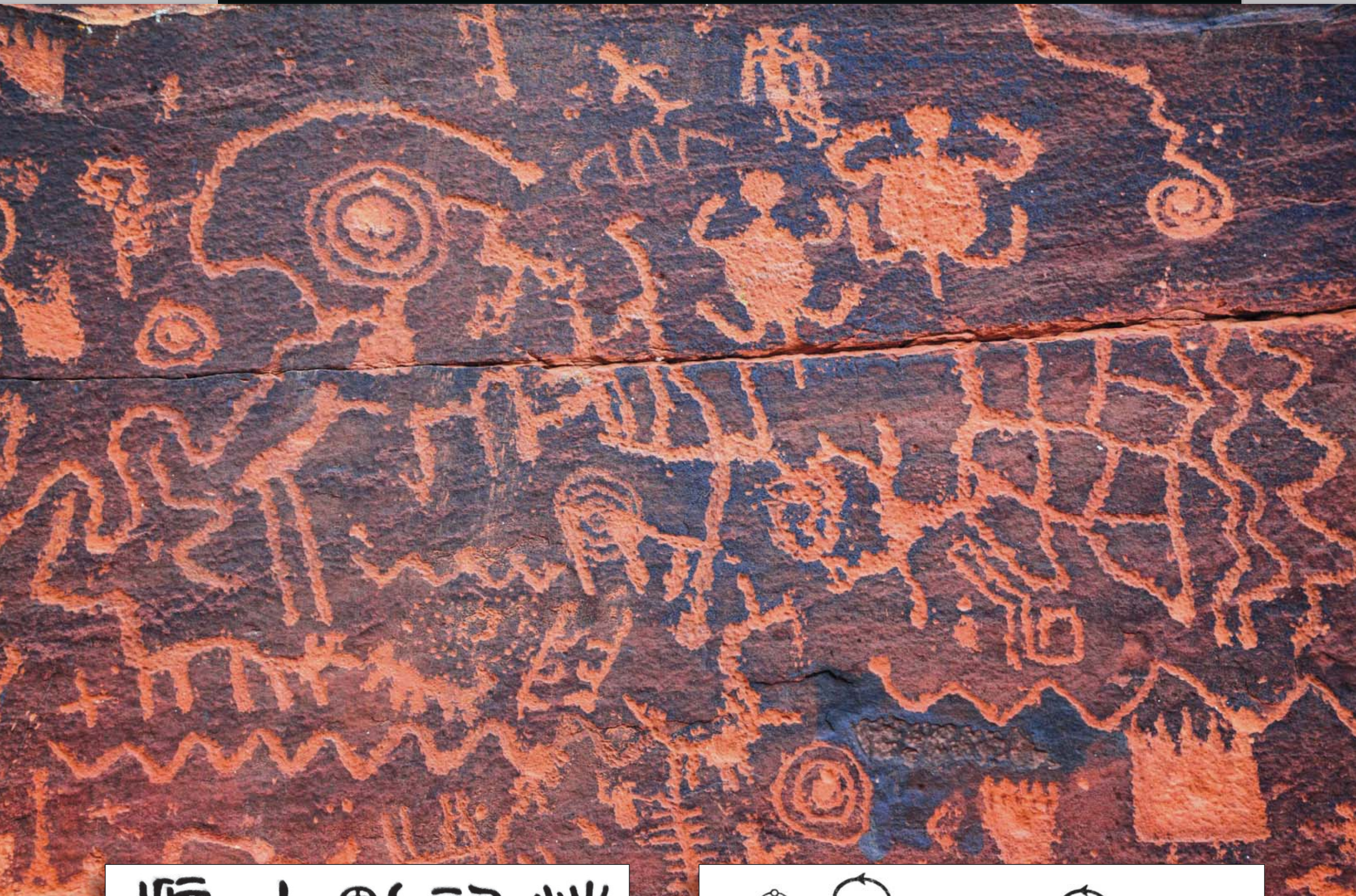
Despite the many proofs without words, mathematical thought *ohne Worte* might be impossible. Words come to us automatically when we view images. Images come to mind when we see words. It seems that we need words after all.

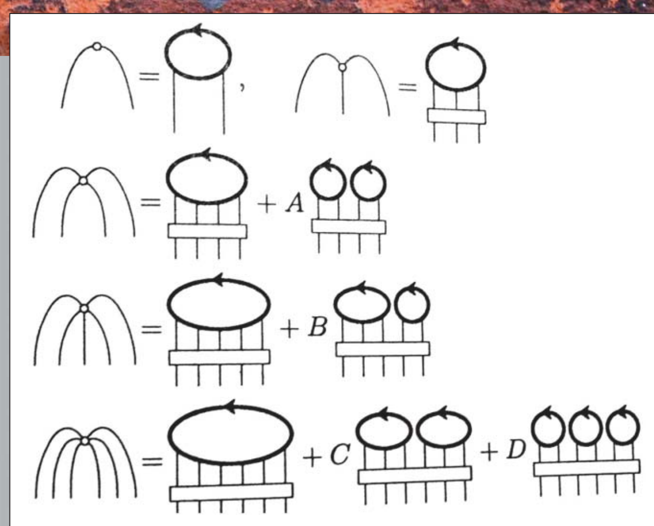
The role that words play in mathematics was on the mind of the French mathematician Jacques Hadamard when, during the 1940s, he asked colleagues around the world how they thought about their subject. Did they think in images or think in words? He summarized his findings in *The Psychology of Invention in the Mathematical Field.* One of the respondents was Albert Einstein, who wrote:

> The words or the language, as they are written or spoken, do not seem to play any role in my mechanism of thought. The psychical

Daniel S. Silver is an emeritus professor of mathematics at the University of South Alabama. His research explores the relation between knots and dynamical systems, as well as the history of science and the psychology of invention. Email: silver@southalabama.edu

American Scientist Previous Page | Contents | Zoom in | Zoom out | Front Cover | Search Issue | Next Page



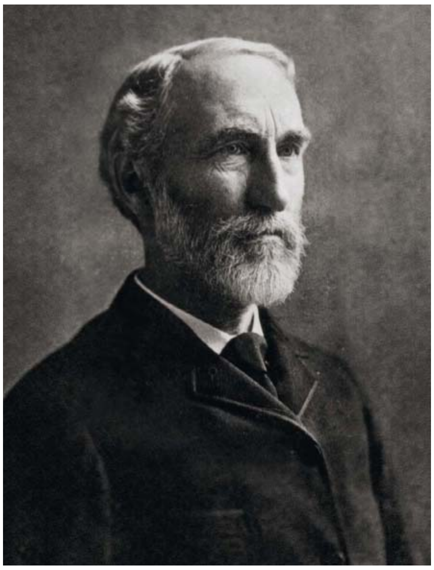


Symbolic representation of words and ideas has an ancient history, from North American petroglyphs (*top*) to early Chinese calligraphy (*above, left*). Modern mathematicians are exploring the usefulness of wordless symbols, including the "birdtracks" notation (*above, right*) created by mathematician Predrag Cvitanović.

> entities which seem to serve as elements in thought are certain signs and more or less clear images which can be "voluntarily" reproduced and combined.

Even those mathematicians who would agree with Einstein recognize the severe limitation on thought imposed by the use of images alone. In 1983 the New Zealander and American mathematician Vaughan Jones made the point by reproducing the image of a big black dot to illustrate the projection lattice of II_1 (read as "type two-one") factors, a sophisticated algebraic object that Jones later used to create new tools for knot theory, which won him the prestigious Fields medal seven years later. Although the dot was not entirely gratuitous—it came about by thinking of factors in terms of concentric black circles—it was intended as a humorous commentary on the constraints of imagery. "Some people have enjoyed the joke," Jones told me.

Whether or not mathematics is a language, its words share with ordinary language an important function: They transport essential images from one mind to another. For that reason



Wikimedia Commons

Josiah Willard Gibbs, a 19th-century mathematician at Yale University who was known for being reserved, nonetheless expressed very strongly his opinion that mathematics was itself a language.

the choice of words that we make is important.

Multiple Meanings

In 1948, Claude Shannon, working on communication theory at Bell Telephone Laboratories, created a beautiful and useful algebraic expression for a measure of average uncertainty in an information source. Its similarity in form to the statistical mechanics notion of entropy, introduced by Rudolf Clausius in 1864, was noted with wonder by many, including the mathematician John von Neumann. Shannon is said to have told an interviewer the following anecdote in 1961:

> My greatest concern was what to call it. I thought of calling it "information," but the word was overly used, so I decided to call it "uncertainty." When I discussed it with John von Neumann, he had a better idea. Von Neumann told me, "You should call it entropy, for two reasons. In the first place your uncertainty function has been used in statistical mechanics under that name, so it already has a name. In the second place, and more important, no one knows what entropy really is, so in a debate you will always have the advantage."

Although von Neumann's second reason presumably was intended as a joke, the exchange highlights the importance of careful language choice in fostering understanding. Like Shannon, many mathematicians agonize over the words that they invent, hoping that those they choose will survive for decades. Shannon's choice proved to be brilliant for three reasons, two in addition to those proposed by von Neumann:

First, mathematicians enjoy the borrowed authority of words from science, especially from physics. Von Neumann was correct when he said that the use of *entropy* conferred an advantage on the speaker. Audience members either know the meaning of the term or else they feel that they should.

Second, Shannon's appropriation of the term *entropy* provoked an insightful debate. Did his term in fact have a meaningful connection with statistical mechanics? The debate has been productive. Today Clausius's entropy is regarded by many as a special case of Shannon's idea.

Third, the wide popularity of Shannon's information theory has been helped by his use of a word that is recognized, if not truly understood, by everyone. Its associations with disorder in everyday life evoke a sympathetic response with everyone, much like another popular mathematical word, *chaos.*

What's in a Name?

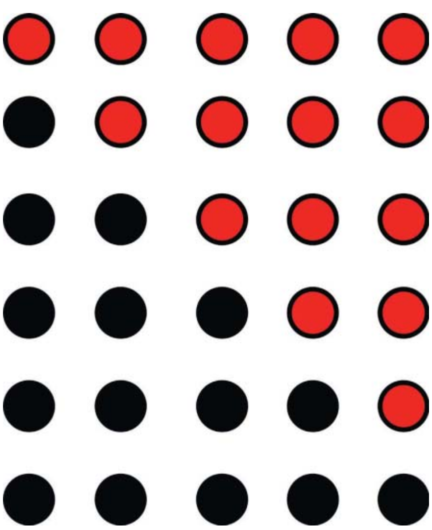
"Must a name mean something?" asked Alice doubtfully. "Of course it must," was Humpty Dumpty's emphatic reply to Alice's question in Lewis Carroll's *Through the Looking-Glass.*

Mathematician James Yorke of the University of Maryland would have given Alice the same answer. In a recent discussion I had with him, Yorke recalled his decision to adopt the word *chaos* for a mathematical phenomenon associated with a type of unpredictable behavior found everywhere in the world, from the drip patterns of water faucets to long-range weather behavior. "Your terms should mean something," he insisted. He dismissed the advice of colleagues urging him to choose a more dispassionate term. Yorke wished to capture the feelings that we all have about the randomness in our lives.

Mathematical monikers usually slip through the world quietly, recognized by only the researchers who use them. *Chaos* was an exception. In 1975 Yorke and his co-author, mathematician Tien-Yien Li of Michigan State University, proved a surprising theorem about continuous functions on an interval, the sort of functions that students learn about in calculus. They wrote up their proof in a short paper titled "Period Three Implies Chaos."

What did Yorke and Li prove? Think of such a function as a machine: Insert a number a from the interval and get a value b. If we put b into our machine, then we get a third number c. Now insert c to get d. If it happens that d is

Mathematical monikers usually slip through the world quietly. The term *chaos* was an exception; it seized the public imagination.



An array of dots can be used as a nonverbal demonstration that the sum 1+2+3+4+5 is equal to 5 (the number of columns needed to represent all the digits) multiplied by 6 (the number of rows needed to represent the largest digit) and then divided by 2.

equal to *a*, then we say that *a* has period three. In a similar way, a number might have period four, five, or any number. What Yorke and Li proved, assuming continuity, is that if some number has period three, then one can find numbers of any period one chooses.

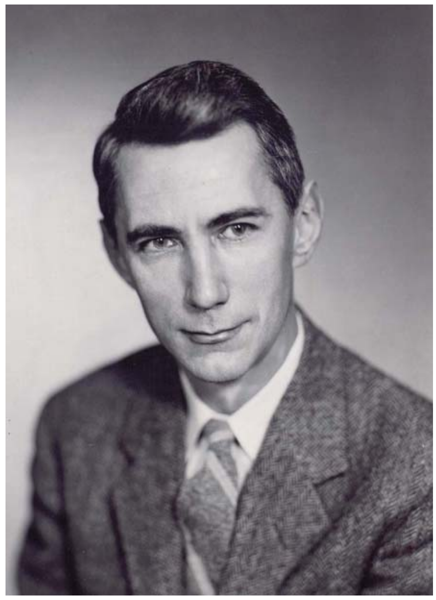
Yorke and Li also showed that many numbers have successive outputs that never return to their starting value. More surprisingly, pairs of such nonperiodic numbers can be found as close together as desired, yet having successive outputs that move apart. It is an example of sensitive dependence on initial conditions. The situation is, well, chaotic.

Had Yorke and Li published their paper in a specialists' journal in their field of dynamical systems, it is likely that *chaos* would never have caught the public's imagination. Few people know that the word had already appeared with another meaning in the wonderfully oxymoronic title "The Homogeneous Chaos." The author was Norbert Weiner, the creator of cybernetics.

Weiner published his article in the 1938 volume of the research-oriented *American Journal of Mathematics.* Instead Yorke and Li published in the *American Mathematical Monthly,* an expository journal published by the Mathematical Association of America, intended for a broad readership from college students to researchers. When biologist Robert May read the article and wrote about its implications for population models a year later in the journal *Nature,* the word *chaos* appeared on top of the first page. According to Yorke, that is when the term took off.

How long does it take a mathematical word to seize the public imagination? Within 10 years, James Gleick's *Chaos: Making a New Science* would become a bestseller. The book, which marks its 30th anniversary this year, celebrated its subject, beginning with the ideas of mathematician and meteorologist Edward Lorenz. In 1961 Lorenz had noticed that over extended periods of time, his model of weather patterns behaved differently when its initial conditions were varied only slightly. Although such a phenomenon was not new to mathematicians—the French mathematician Henri Poincaré had written about it at the turn of the 20th century—it came as a shock to many scientists. Storm conditions today, Lorenz suggested, might have been caused several weeks ago by a butterfly flapping its wings in Brazil. *Butterfly effect* was the term he coined to reinforce the metaphor.

If a mathematical term is to catch fire with general readers, it must spark their imagination. A powerful image helps. In the case of chaos, Lorenz himself had supplied it. It was a bundle of curves, today called the *Lorenz attractor,* suggesting the infinite number of solution curves to a system of three differential equations. And it resembled the two wings of a butterfly. Start at any point of one of its solution curves and follow around, and you visit the two wings successively in some infinite pattern. Start at a nearby point on a different curve and the pattern might become wildly different after some time. The Lorenz attractor became the emblem of a new and lasting field, chaos theory.



Claude Shannon of Bell Telephone Laboratories in 1948 borrowed the term *entropy* from statistical mechanics to name his algebraic expression for the amount of uncertainty in an information source. His term has attracted widespread attention. Today some consider the statistical mechanical notion to be a special case of Shannon's entropy. (Photograph courtesy of Tekniska museet/Flickr cc.)

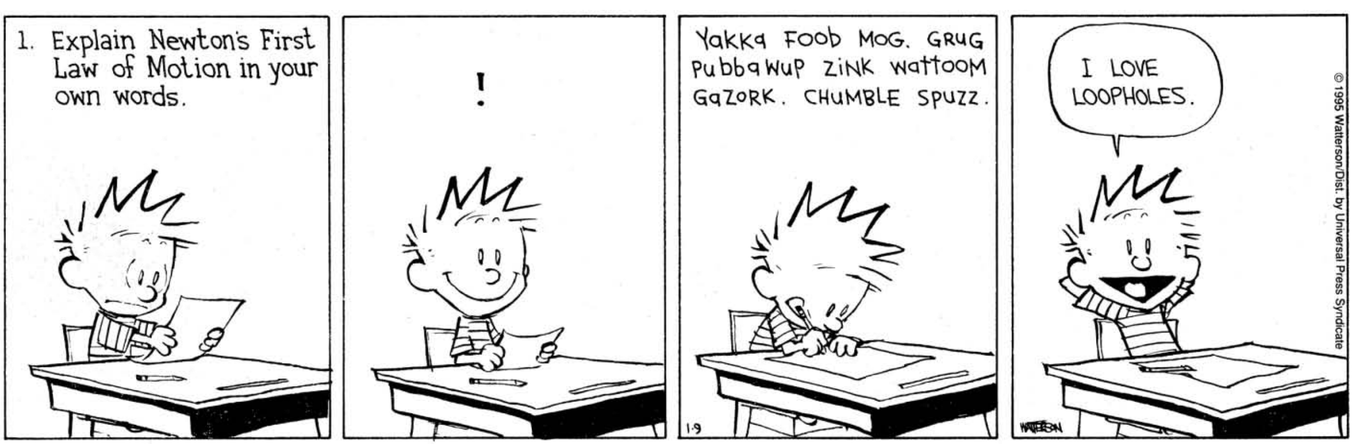
Choosing Words, Sometimes with Care

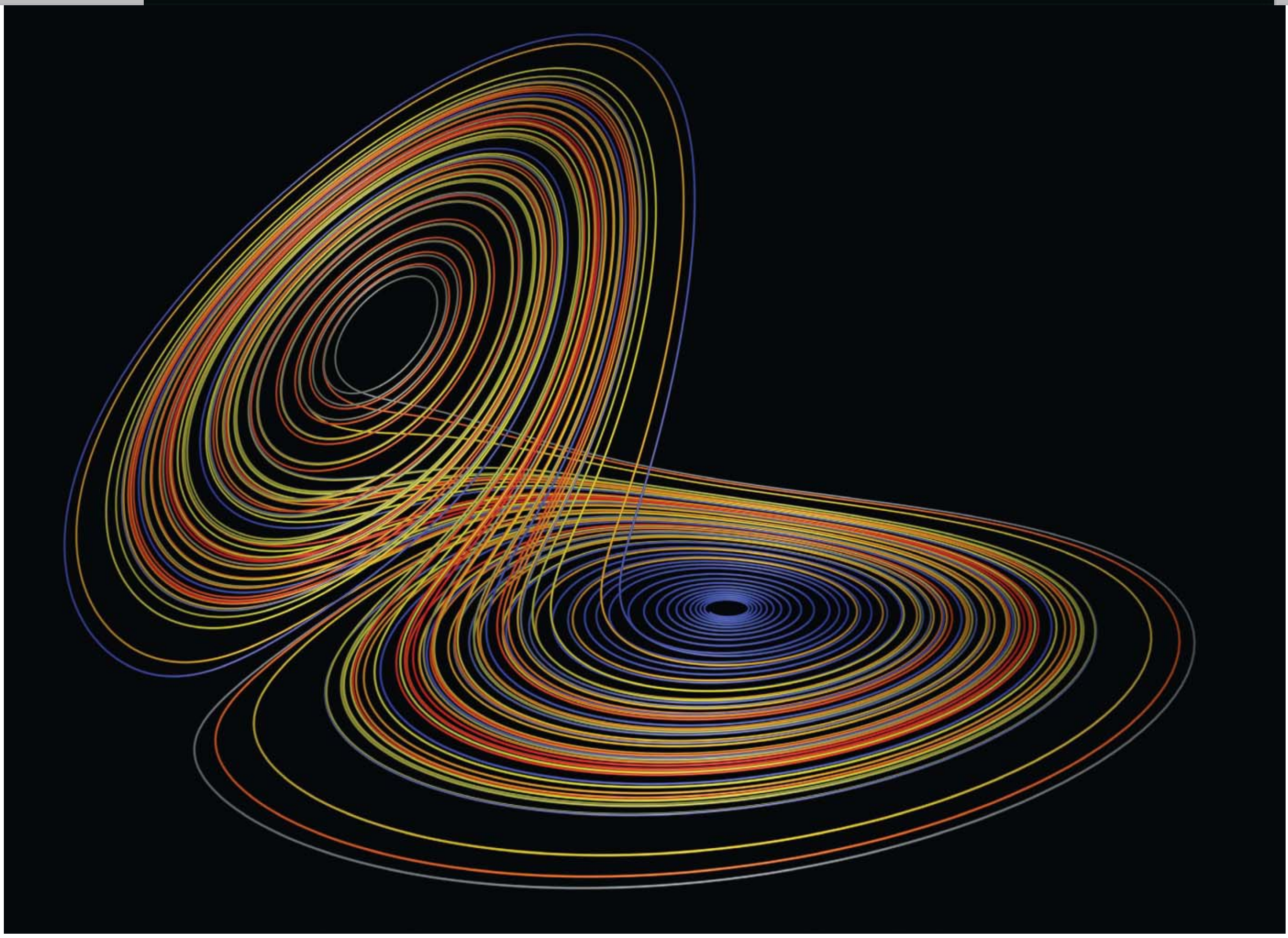
Poorly chosen words and phrases can interfere with the progress of mathematics. René Descartes showed us how.

While pondering solutions to algebraic equations, Descartes was compelled to consider the possibility that some numbers when multiplied by themselves could give a negative result. Certainly no "real numbers," that is, no numbers on the familiar number line, behave that way. Descartes called the numbers *imaginaire* (imaginary). In *La Géométrie,* his effort in 1637 to unify geometry and algebra, he explained that "true or false roots can be real or imaginary."

The expression *imaginary number* caught on. Carl Friedrich Gauss, one

Although mathematicians sometimes joke about the terms they try to introduce, some agonize over the words that they invent, and others wonder whether neologisms cause confusion.





The Lorenz attractor is a bundle of curves plotting the solutions to a system of three differential equations. It resembles the wings of a butterfly, and became an emblem for the subject of chaos. Some of its infinitely many curves are colored to distinguish them. (Image by Paul Bourke.)

of the greatest mathematicians of all time, despised it. In 1831 Gauss wrote: "If this subject... [has been] enveloped in mystery and surrounded by darkness, it is largely an unsuitable terminology which should be blamed."

Gauss preferred the less censorious term *complex numbers*, which includes both ordinary (real) numbers and imaginary numbers appearing together in a single expression, such as $2+\sqrt{-3}$. Regrettably, Descartes's coinage remains in circulation, adding to the handicap that mathematics teachers endure as they try to convince some students that complex numbers are more than imaginary, having significant applications throughout the sciences.

In 1670 Isaac Newton contributed the intensely bothersome words *fluents, fluxions,* and even *ffluxions.* They were intended to describe the path and velocity of fluid in motion, but defining them rigorously proved too difficult even for the Cambridge genius. For more than a century mathematicians tried and failed to clarify the meanings of these words with their implied metaphysical notions of time. During the 1800s rigorous and effective definitions, those of *limit* and *derivative*, were finally developed without the mention of time, and mathematicians stopped using Newton's f-words.

Leibniz was so smitten with symbols that he dreamed of a purely symbolic language, one with which nations might someday settle arguments, using computation rather than swords and cannon.

British mathematicians of the 19th century loved inventing new words and phrases for mathematical concepts. Unfortunately, they got off to a rather bad start.

In 1801 a book by Cambridge professor John Colson, who held the Lucasian chair of Mathematics (earlier held by Newton and later by Stephen Hawking) was published as an English translation of *Instituzioni analitiche ad uso della gioventú italiana* [*Foundations of Analysis for the Use of Italian Youth*], a textbook about calculus written by Maria Agnesi in 1748. It is a pity that Colson didn't know much Italian. He learned just enough to follow the path of translation, dropping a hazardous rock onto the road along the way.

In her book Agnesi had described a particular curve, referring to it as "la curva... dicesi la Versiera." The word *versiera* was an adaptation by another Italian author, Guido Grandi, from the Latin word *versoria,* meaning "a rope that turns a sail." It was a helpful image. Colson could not find Agnesi's

American Scientist Previous Page | Contents | Zoom in | Zoom out | Front Cover | Search Issue | Next Page

"*la versiera,*" because it was not in any dictionary. He used the closest word that he could find, *l'avversiera.*

In Italian *l'avversiera* means "the witch" or "the she-devil." Colson wrote: "...the equation of the curve to be described, which is vulgarly called the Witch." Today's calculus textbooks continue to use the term Witch of Agnesi, and students continue to stare at the curve and wonder what demonic forces shaped it. It is ironic that Agnesi was a devout woman who spent most of her adult life aiding the poor.

No British mathematician was a greater wordsmith than James Joseph Sylvester. Born in London in 1814, he attended Cambridge University but could not be awarded a degree because, as a Jew, he did not subscribe to the Thirty-Nine Articles of Religion of the Anglican Church. He managed to support himself by teaching at the secular University College London and working as an actuary. Finally, in 1876, Sylvester's career reached its full promise when he was appointed to a professorship at the new Johns Hopkins University in Baltimore, Maryland. There he began the United States' first mathematics research department and its first mathematics research journal, the aforementioned *American Journal of Mathematics.*

Sylvester loved language almost as much as mathematics. He composed unappreciated poetry, and in 1870 he proudly self-published a slim volume, *The Laws of Verse,* in which he proposed rules for effective versification. Not surprisingly, Sylvester coined many mathematical words. At the end of his article, "On a Theory of the Syzygetic Relations of Two Rational Integral Functions," published in 1853, he attached a glossary of "New or unusual Terms, used in a new or unusual sense in the preceding Memoir." It began with *allotrious, apocopated, bezoutic,* sprinkled *monotheme, perimetrical, rhizoristic* along the way, and finished up with *umbral, weight,* and *zeta.*

Most of Sylvester's words have been forgotten, but some have survived. Perhaps his most notable contribution to mathematic's lexicon is *matrix,* a square or rectangular arrangement of terms in rows and columns. It is a Latin word that means womb. Sylvester was most interested in the case of a square matrix, one for which the numbers of rows and columns are equal. Because a rectangular matrix can give birth, so to speak, to a square matrix by striking out unwanted rows or columns, *matrix,* with its suggestion of fertility, seemed appropriate.

The word *matrix* lives on in mathematics at all levels. And like *chaos,* it has captured popular imagination. Sylvester would have been delighted to know that at the end of the 20th century both a Hollywood science-fiction film and a compact car would carry the name.

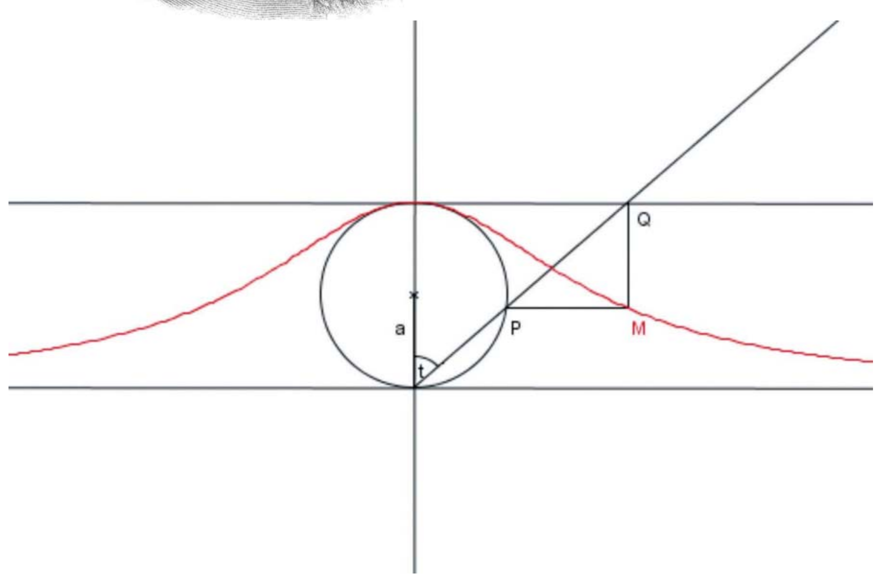
A Picture Worth 1,000 Symbols

The symbols of mathematics appear to be separate from the words alongside them. In fact, they have evolved from words. The 19th-century German philologist and historian Heinrich Nesselmann was the first to recognize this concept. He identified three stages of their evolution: rhetorical, syncopated, and symbolic.

An illustration of Nesselmann's theory is furnished by the humble minus sign that we use for subtraction. In medieval Europe the operation of subtraction was written out; we find it recorded as *minus* (Latin), *moins* (French), or *meno* (Italian). However, by the 15th century, the word had been syncopated, shortened to $\overline{m}$. The symbol – began to replace $\overline{m}$ as early as 1489, the year that Johannes Widmann published a book of commercial arithmetic with the phrase (in translation): "What – is, that is less, and the + is more."

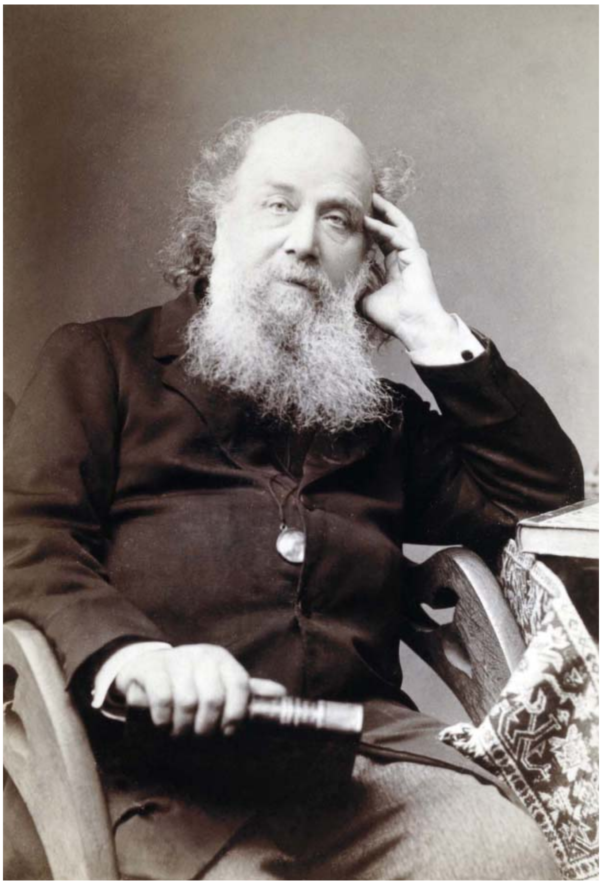
No matter what their origin, symbols have been a source of mystery for mathematicians. They do more than merely stand in for words. Gottfried Wilhelm Leibniz, who shares with Newton credit for discovering calculus, believed it. Leibniz was so smitten by symbols that he dreamed of a purely symbolic language, one with which nations might someday settle arguments, using computation rather than swords and cannon. He imagined an alphabet of ideas.

Maria Gaetana Agnesi (*left*) authored a mathematics textbook in Italian in 1748. In the volume she describes a particular curve (*below*), calling it *la versiera,* and adaption from the Latin word *versoria,* meaning a rope that turns a sail. But when her textbook was translated into English by Cambridge mathematician John Colson and published in 1801, he mistook the word for another that means "the witch." The mistranslated moniker still sticks with the curve in modern textbooks, which call it *Witch of Agnesi,* to the confusion of students who are not told of the reasons for any supernatural connections. The story demonstrates the importance of choosing potential mathematical terms with care.



Wikimedia Commons

American Scientist Previous Page | Contents | Zoom in | Zoom out | Front Cover | Search Issue | Next Page



Chris Hellier / Science Source

James Joseph Sylvester, who taught mathematics at Johns Hopkins University in the late 1800s, invented a host of mathematical terms, but the most enduring one is *matrix*. Sylvester borrowed the term from the Late Latin word for womb, likely to imply how a rectangular arrangement of terms in rows and columns can "give birth" to a square one.

Leibniz never found his universal language, but others have had success searching for ideographic languages with more specialized functions. The German mathematician and philosopher Gottlob Frege thought that he had found a way to communicate with just symbols. In 1879 he published *Begriffsschrift*, an 88-page booklet in which the quantifiers of formal logic first appeared. Frege described his work as "a formula language, modeled on that of arithmetic, of pure thought." Diagrams stretched not only across its pages but up and down them as well.

Today Frege's work is regarded by many as the most important single work in logic. However, in its day, the book was derided. Logician John Venn (remembered today for his eponymous diagrams) described *Begriffsschrift* as "cumbersome and inconvenient." Another influential logician, Ernst Schröder, called it a "monstrous waste of space" and complained that it "indulges in the Japanese custom of writing vertically."

Frege faced resistance not just from readers. His two-dimensional formulas and other exotic symbols precipitated complaints from his typesetter. "After all, the convenience of the typesetter is certainly not the *summum bonum* [highest good]," Frege argued some years later. Costs associated with typesetting, once a bane for mathematicians, have disappeared today thanks to effective software that enables authors to create camera-ready manuscripts for journals and periodicals. Nevertheless, the mathematical community's reluctance to adopt pictorial arguments has slowed their acceptance.

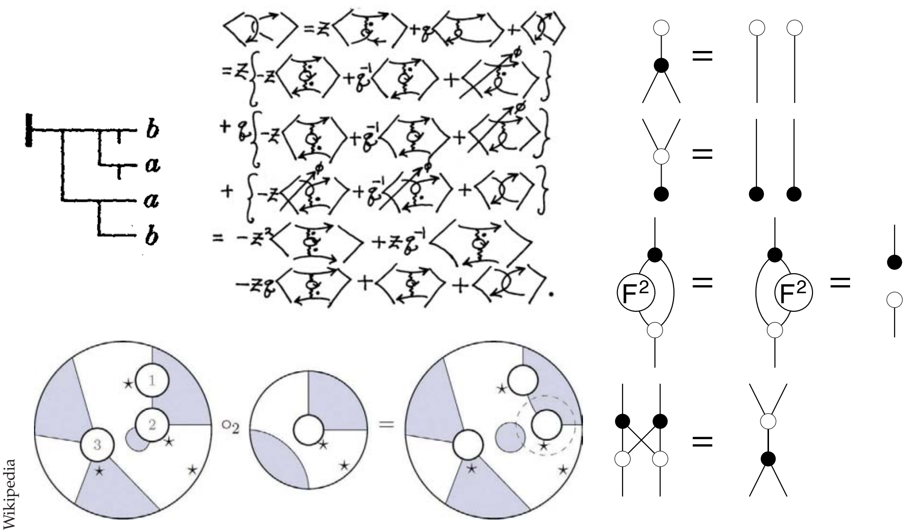
Pictures have accompanied mathematical proofs since the days of Euclid. But a picture can mislead us by suggesting that the special case it illustrates is sufficiently general for our argument. Beginning in the late 19th century, intuition-defying examples from the expanding subject of topology reinforced doubts about our spatial intuition and proofs that rely on them. Giuseppe Peano's *space-filling curve* was one such example: Try to imagine an unbroken curve inside a square that does not miss a single point. Few can, but Peano could. He constructed such a curve using an infinite recursive process. No picture can fully describe it. (More about Peano can be found in "Crinkly Curves," *Computing Science,* May–June 2013.)

Mathematicians' reluctance to accept images in place of words has softened but not vanished. It is a common and uncomfortable experience for someone presenting a pictorial argument to hear a skeptic ask, "Is that really a proof?" Predrag Cvitanović of Georgia Institute of Technology knows the experience. He recalls the taunt of a colleague staring at the pictograms that he had left on a blackboard: "What are these birdtracks?" Cvitanović liked the ornithological term so much that he decided to adopt it for the name of his new notation.

So what are these birdtracks? Briefly, they are combinations of dots, lines, boxes, arrows, and other symbols, letters of a diagrammatic language for a type of algebra. They were inspired by the famous paper about QED, written in 1948 by physicist Richard Feynmann, and also later articles by mathematical physicist Roger Penrose. As Cvitanović explained in his 1984 book *Group Theory: Birdtracks, Lie's, and Exceptional Groups,* the diagrams represent an evolution of language. They are not merely mnemonic devices or an aid for computation. Rather, they are "everything—unlike Feynman diagrams, here all calculations are carried out in terms of birdtracks, from start to finish." Could one perform the calculations without them? Yes, but as Cvitanović warns the reader, it would be like speaking Italian without using your hands.

Birdtracks are not the only picture-language game in town. Others have been, and are being, invented for different purposes. Louis H. Kauffman, a mathematician at the University of Illinois at Chicago and one of the most inventive and influential topologists of our time, uses a variety of new diagrammatic languages for the study of knots. An algebraist familiar with bird-

It is a common and uncomfortable experience for someone presenting a pictorial argument to hear a skeptic ask, "Is that really a proof?"



Wikipedia

Attempts at purely symbolic mathematical languages have taken on a range of formats. One diagram from the work of Gottlob Frege in 1879 asserts that: "If *b* implies *a*, then the negation of *a* implies the negation of *b*" (*top left*). Louis Kauffman used picture-language to argue that a certain quantity for a knot does not depend on how the knot is drawn (*top, middle*). The quon language is used at top right to express the properties of a pair of expressions in Frobenius algebra. Disks called *planar tangles* are used to express properties in planar algebra (*bottom*).

tracks would recognize much but not all of what she might read in Kauffman's book *Formal Knot Theory*, much as a traveler might understand spoken cognates of a foreign tongue.

Planar algebras are another significant picture language. Jones (the one with the big black dot mentioned earlier) introduced them in 1999, and they provide a general setting for important quantities in the study of knots. The basic pictures of the language are *planar tangles*—disks with smaller disks inside them—connected by line segments and decorated with stars and shading. Planar tangles can be combined to form new characters. The properties they exhibit mirror those of different algebraic and topological structures that are already familiar to researchers. Consequently, planar algebras have a variety of applications.

Birdtracks and planar algebras are picture-languages that draw their initial inspiration from physics. Yet another is *quon language*, first introduced in December 2016 on arXiv.org, a repository of online research-paper preprints that is used by scientists throughout the world. Quon language, created by Harvard University mathematicians Zhengwei Liu, Alex Wozniakowski, and Arthur M. Jaffe, is derived from three-dimensional pictorial representations of particlelike excitations and transformations that act on them. An earlier, simpler version has much in common with the languages of Cvitanović and Kauffman.

According to its inventors, the quon language can do more than aid in the study of quantum information. It is also a language for algebra and topology, with the ability to prove theorems in both subjects. In an interview with the *Harvard Gazette*, Jaffe remarked, "So this pictorial language for mathematics can give you insights and a way of thinking that you don't see in the usual, algebraic way of approaching mathematics." He added that "It turns out one picture is worth 1,000 symbols."

Not Just Another Language

Something strange is indeed happening in mathematics seminar rooms today, but it amounts to more than amusing sights and sounds. Mathematicians are attempting to break through the barriers of traditional language in order to think more deeply about fundamental questions. Their strange words and images are attracting attention, motivating all of us to learn more about them.

In a series of lectures at Cornell University in 1965, Richard Feynman contemplated the effectiveness of mathematics in science. He sympathized with the lay reader who asked why it was not possible to explain mathematical ideas with ordinary language. Correcting Gibbs, who started our discussion, Feynman replied that it is "because mathematics is *not* just another language." Simply put, mathematics is more general than any language that tries to express it.

Traditional methods of learning mathematics are discursive, demanding a sequential, step-by-step understanding. Assuming that a lesson succeeds, there is a moment of "aha!" when lights turn on and the room is illuminated for us. One day it might become possible to flip on the light switch as soon as we enter the room. Philosopher and logician Susanne K. Langer once called for a revolution in our modes of communication, moving beyond our "tiny, grammar-bound island." Our journey from words to symbols to picture language is bringing Langer's revolution just a bit closer.

References

Cvitanović, P. 2008. *Group Theory: Birdtracks, Lie's, and Exceptional Groups.* Princeton, NJ: Princeton University Press.

Feynman, R. 1965. *The Character of Physical Law.* London: Cox and Wyman Ltd.

Gleick, J. 1987. *Chaos: Making a New Science.* New York: Penguin Books.

Hadamard, J. 1945. *The Psychology of Invention in the Mathematical Field.* Princeton, NJ: Princeton University Press.

Jones, V. F. R. 1999. Planar algebras, I. arXiv:math/9909027v1.

Kauffman, L. H. 1991. *Knots and Physics.* Singapore: World Scientific Publishing Co.

Langer, S. K. 1942. *Philosophy in a New Key.* Cambridge, MA: Harvard University Press.

Sylvester, J. J. 1853. On a theory of the syzygetic relations of two rational integral functions. *Philosophical Transactions of the Royal Society of London* 143:407–548.

Li, T-Y. L., and J. A. Yorke. 1975. Period three implies chaos. The *American Mathematical Monthly* 82:985–992.

Liu, Z., A. Wozniakowski, and A. M. Jaffe. 2017. Quon 3D language for quantum information. *Proceedings of the National Academy of Sciences of the U.S.A.* 114:2447–2502.

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

www.amsci.org/magazine/issues/2017/november-december

American Scientist Previous Page | Contents | Zoom in | Zoom out | Front Cover | Search Issue | Next Page

The Persistence and Peril of Misinformation

Defining what truth means and deciphering how human brains verify information are some of the challenges to battling widespread falsehoods.

Brian G. Southwell, Emily A. Thorson, and Laura Sheble

Misinformation—both deliberately promoted and accidentally shared—is perhaps an inevitable part of the world in which we live, but it is not a new problem. People likely have lied to one another for roughly as long as verbal communication has existed. Deceiving others can offer an apparent opportunity to gain strategic advantage, to motivate others to action, or even to protect interpersonal bonds. Moreover, people inadvertently have been sharing inaccurate information with one another for thousands of years. However, we currently live in an era in which technology enables information to reach large audiences distributed across the globe, and thus the potential for immediate and widespread effects from misinformation now looms larger than in the past. Yet the means to correct misinformation over time might be found in those same patterns of mass communication and of the facilitated spread of information.

Misinformation vs. Disinformation

Misinformation is concerning because of its potential to unduly influence attitudes and behavior, leading people to think and act differently than they would if they were correctly informed, as suggested by the research teams of Stephan Lewandowsky of the University of Bristol in the United Kingdom and Elizabeth Marsh of Duke University, among others. In other words, we worry that misinformation (or false information) might lead people to hold misperceptions (or false beliefs) and that these misperceptions, especially when they occur among large groups of people, may have downstream consequences for health, social harmony, and political life.

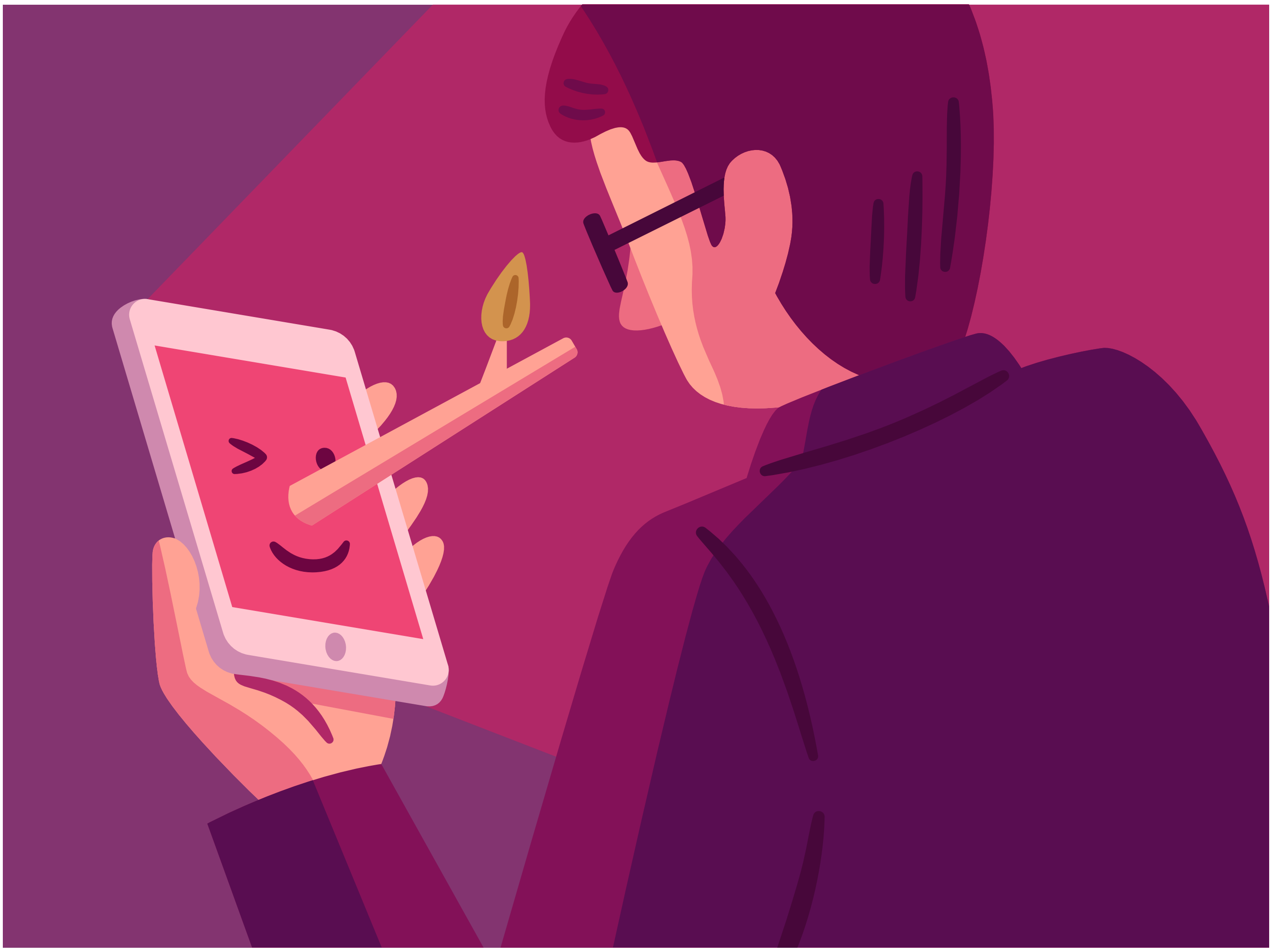
Concern about proving that any information is objectively true can complicate the distinction of misinformation from information.

Does misinformation require intentional deceit on the part of the presenter? Philosopher Jürgen Habermas, who taught at institutions such as Goethe University Frankfurt and the Max Planck Institute in Germany until formally retiring in the 1990s, focuses on a speaker's intent to deceive to distinguish between *misinformation* and *disinformation*. Habermas views truth as only possible collectively among people as a product of consensus; one's collegial participation in such collective understanding also matters. Misinformation from such a perspective, then, is contentious information reflecting disagreement among people, whereas disinformation is more problematic, as it involves deliberate alienation or disempowerment of other people. Lewandowsky and his colleagues have carried forward this definition of *disinformation* as intentionally incorrect information.

From a critical theory perspective, we might worry about the impossibility of proving that any information is objectively true, a concern that complicates any distinction of misinformation from information on the basis of the former being false and the latter being true. Nonetheless, we respectfully reject a worldview in which no degree of consensus among people—agreement as to what is true—is possible. If we allow that a claim acknowledged by consensus to be true is valid, we then can position misinformation as a category of claims for which there is at least substantial disagreement (or even a consensus for rejection) when their truthfulness is judged by the widest feasible range of observers. That approach would include disinformation as a special type of misinformation distinguished by the promoter's intent.

From an ethical perspective, many people worry most about active promotion of disinformation. We think it is best, however, to continue to use the word *misinformation*, because we should acknowledge that false information can mislead people even if unintentionally promoted or mistakenly endorsed as the truth. This approach opens the door for certain claims to evolve from accepted information to become misinformation, and vice versa, as a function of a society's changing consensus. (And, of course, such evolution of evidence can be useful.) We can define *misinformation* as claims that do not enjoy universal or near-

Brian G. Southwell is director of the Science in the Public Sphere Program at RTI International in North Carolina and a faculty member at Duke University and the University of North Carolina at Chapel Hill. Emily A. Thorson is an assistant professor of political science at Syracuse University. Laura Sheble is an assistant professor of information science at Wayne State University in Michigan. Adapted, with permission, from Misinformation and Mass Audiences, *edited by Brian G. Southwell, Emily A. Thorson, and Laura Sheble; © 2018 by the University of Texas Press. Email for Southwell: bsouthwell@rti.org.*



Michele Rosenthal

Although the practice of purposeful deception, or unintentional spreading of falsehoods, is not new, the recent, rapid growth of mobile devices and social media have greatly expanded the conduit for both information and misinformation to rapidly reach a large, global audience.

universal consensus as being true at a particular moment in time on the basis of evidence.

Why Misinformation is Problematic

At least three observations related to misinformation in the contemporary mass-media environment warrant the attention of researchers, policy makers, and really everyone who watches television, listens to the radio, or reads information online. First of all, people who encounter misinformation tend to believe it, at least initially. Secondly, media systems often do not block or censor many types of misinformation before it appears in content available to large audiences. Thirdly, countering misinformation once it has enjoyed wide exposure can be a resource-intensive effort.

Knowing what happens when people initially encounter misinformation holds tremendous importance for estimating the potential for subsequent problems. Although individuals generally have considerable routine experience encountering information now considered to be false, the question of exactly how—and when—we mentally label information as true or false has garnered philosophical debate.

The dilemma is neatly summarized by a contrast between how the philosophers René Descartes and Baruch Spinoza described human information engagement centuries ago, with conflicting predictions that only recently have been empirically tested in robust ways. Descartes argued that a person only accepts or rejects information after considering its truth or falsehood; Spinoza argued that people accept all encountered information (or misinformation) by default and then subsequently verify or reject it through a separate process. In recent decades, empirical evidence from the research teams of Erik Asp of the University of Chicago and Daniel Gilbert at Harvard University, among others, has supported Spinoza's account: People appear to encode all new information as if it were true, even if only momentarily, and later tag the information as being either true or false, a pattern that seems consistent with the observation that mental resources for skepticism physically reside in a different part of the brain than the resources used in perceiving and encoding.

We also know from work by one of us (Southwell) and others that people judge source credibility as a cue in determining message acceptability and will turn to others for confirmation of the truth of a claim. If the people surrounding someone tend to initially believe misinformation, then the specter of network reinforcement is raised, meaning that the false claim becomes more difficult to debunk as it is believed by more people.

What about our claim that misinformation often can appear in electronic or print media without being preemp-

photo Gage Skidmore/Flickr cc

The phenomenon of fake news has become commonplace on the Internet. Some sites disguise themselves as legitimate news outlets and release false stories, such as a 2016 hoax about actor Sylvester Stallone's death. Although other sites, such as the one in the screen shot here, quickly counter the stories, hoaxes are often shared widely before being disproven—and, ironically, repeating the falsehood can further engrain belief in it.

tively blocked? One might consider the nature of regulatory structures in countries such as the United States: Regulatory agencies tend to focus on post hoc detection of broadcast information. Organizations such as the U.S. Federal Trade Commission, the Federal Election Commission, and the Food and Drug Administration (FDA) offer considerable monitoring and notification functions, but these roles typically do not involve preemptive censoring. The FDA oversees direct-to-consumer prescription drug advertising, for example, and has developed mechanisms such as the "Bad Ad" program, through which people can report advertising in apparent violation of FDA guidelines on drug risks and benefits presentation.

People appear to encode all new information as if it were true, even if only momentarily, and later tag the information as being either true or false.

Such programs, although laudable and useful, do not guarantee that false advertising never appears on the airwaves and, moreover, do not prevent false news stories from appearing. In addition, as shown by one of us (Thorson), even misinformation that is successfully corrected can continue to affect attitudes.

Following on this last point, countering misinformation with new information requires effort not only to develop new content that is understandable but also to ensure adequate message exposure. As Robert Hornik of the University of Pennsylvania has argued, a communication campaign can succeed or fail, at least in part, as a function of exposure or lack thereof.

A campaign to correct misinformation, even if rhetorically compelling, requires resources and planning to accomplish necessary reach and frequency. For corrective information to be persuasive, audiences need to be able to comprehend it, which requires either effort to frame messages in ways that are understandable or effort to educate and sensitize audiences to the possibility of misinformation. That audiences might not be aware of the potential for misinformation also suggests the potential utility of media literacy efforts as early as elementary school. Even with journalists, pundits, and scholars pointing to the phe-

Wikimedia Commons

How the brain encodes information and when it judges it to be true or false has been a subject of philosophical debate for centuries. René Descartes (*left*) argued that people would consider the truthfulness or falsity of information before accepting or rejecting it. Baruch Spinoza (*right*) argued that people accept all information by default, then verify or reject it through a separate process.

In Kano, Nigeria, the polio virus thrives in sewage-contaminated water. Misinformation, however, caused many residents to refuse the polio vaccine, leading to the disease's reemergence a few years ago, after it was nearly eradicated.

ZUMA Press, Inc./Alamy Stock Photo

nomenon of "fake news," people often do not distinguish between demonstrably false stories and those based in fact when scanning and processing information.

We live at a time when widespread misinformation is common. Yet at this time many people also are passionately developing potential solutions and remedies. The journey forward undoubtedly will be a long one. The way in which media systems have developed in many democratic societies inherently calls for both vulnerability to occasional misinformation and robust systems to detect and address it.

Future remedies will require not only continued theoretical consideration but also the development and maintenance of consistent monitoring tools and a willingness among neighbors and fellow members of society to agree that some claims that find prominence on shared airwaves and in widely available media content are insufficiently based in scientific consensus and social reality and should be countered. Misinformation arises as a function of systems structure, human fallibility, and human information needs. To overcome the worst effects of the phenomenon, we will need coordinated efforts over time, rather than any singular, one-time panacea we could hope to offer.

Everett Collection

In 1938, actor Orson Welles rehearses for a radio broadcast of H. G. Wells's *The War of the Worlds*. Some Americans panicked when they believed the radio drama—and then newspapers sensationalized the panic to demonize their radio competitors. Such incidents demonstrate that misinformation can be mistakenly spread as fact whether or not the originator had that intention.

Bibliography

Aikin, K. J., et al. 2015. Correction of overstatement and omission in direct-to-consumer prescription drug advertising. *Journal of Communication* 65:596–618.

Asp, E. W., and D. Tranel. 2012. False tagging theory: Toward a unitary account of prefrontal cortex function. In D. T. Stuss and R. T. Knight, eds. *Principles of frontal lobe function*. New York, NY: Oxford University Press, pp. 383–416.

Gilbert, D., R. Tafarodi, and P. Malone. 1993. You can't not believe everything you read. *Journal of Personality and Social Psychology* 65:221–233.

Lewandowsky, S., U. K. H. Ecker, C. M. Seifert, N. Schwarz, and J. Cook. 2012. Misinformation and its correction: Continued influence and successful debiasing. *Psychological Science in the Public Interest* 13:106–131.

Thorson, E. 2016. Belief echoes: The persistent effects of corrected misinformation. *Political Communication* 33:460–480.

American Scientist Previous Page | Contents | Zoom in | Zoom out | Front Cover | Search Issue | Next Page

Scientists' Nightstand

The Scientists' Nightstand, American Scientist's *books section, offers reviews, review essays, brief excerpts, and more. For additional books coverage, please see our* Science Culture *blog channel, which explores how science intersects with other areas of knowledge, entertainment, and society: americanscientist.org/blogs/science-culture*.

ALSO IN THIS ISSUE

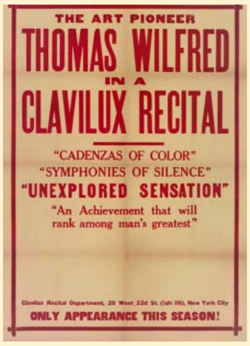
THE RISE OF BIG DATA POLICING: Surveillance, Race, and the Future of Law Enforcement. By Andrew Guthrie Ferguson. page 378

ONLINE

Upcoming on our *Science Culture* blog: *americanscientist.org/blogs/science-culture*

CLIMATE CHANGE AND THE HEALTH OF NATIONS: Famines, Fevers, and the Fate of Populations. By Anthony J. McMichael.

We Choose to Go: A Conversation with *Apollo 8* Author Jeffrey Kluger



From *Lumia*.

Sumptuous Science

Not so long ago, when digital books were ascendent, avid readers would speculate from time to time about the future of print books. As the number of readers browsing print copies of newspapers and magazines dwindled, it seemed plausible that book readers might migrate en masse as well.

More recently digital book sales have been declining and, for a multitude of reasons, print books don't appear to be destined for the endangered species list anytime soon. For this, as a reader, reviewer, editor, and certified book addict, I am grateful. Although there's nothing handier than being able to carry the equivalent of a bookcase full of titles on a device the size of a paperback, it remains true that some books are best experienced on paper. And in a few cases, the full wonder of a title may be otherwise lost.

Such is the case for an impressive clutch of recent releases. Each will richly reward readers who leaf through their pages, linger over their images, and launch into the text. Moreover, each demonstrates how vibrant and gorgeous science-related titles can be.

The experience of reading Terry Burrows's ***The Art of Sound: A Visual History for Audiophiles*** (Thames and Hudson, 2017) is akin to strolling through a thoughtfully staged museum exhibit on the history of sound recording and playback—one that presents technology matter-of-factly while also dipping into pop culture. The book is divided into chapters representing four eras of audio evolution: acoustic, electric, magnetic, and digital. Sections of patent drawings punctuate the transitions between chapters, and it's easy to lose yourself in their details as you take in the change over the decades from the elegant mechanics of the acoustic era to the intricate circuitry threading through digital devices. (One patent illustration, submitted by the Victor Talking Machine Company in 1927, depicts an instructional tag that offers a suggestion that remains refreshingly useful: "Read your Instruction Book carefully—it will save you expense and trouble.") Each chapter opens with a timeline that anchors its topic and is followed by a lucid essay by Burrows, an author who specializes in music history, that discusses the developments of that era and effectively contextualizes the images to follow. These images—delightful, varied, and attentively captioned—make up the bulk of each chapter, and rightfully so. The real stars are Simon Pask's copious photographs, which manage to be both luscious and crisp, of items from the EMI archives. From a 1902 gramophone that played discs made of chocolate, to a 1940 U.S. Army–issued wind-up record player for use in war zones, to a shiny array of Sony Minidisc Walkman devices from the 1990s through the early 2000s, these artifacts transfix. I'd encourage anyone who suspects that *technology* and *beauty* are antithetical terms to spend time immersed in these pages.

Instead of sound, Keely Orgeman lavishes attention on light in ***Lumia: Thomas Wilfred and the Art of Light*** (Yale University Art Gallery and Yale University Press, 2017). The book's release coincided with the opening of an exhibition of Wilfred's work earlier this year at the Yale University Art Gallery, where Orgeman curates American artwork. *Lumia* is the first book devoted to Wilfred's work in 40 years, and it aims to introduce the artist to new generations. Enthusiasts familiar with contemporary artists such as James Turrell (who contributes the book's foreword), Jenny Holzer, and Olafur Eliasson, whose light-based installations are well-known and widely exhibited, may never have heard of Thomas Wilfred, who pioneered the art form. (His

Courtesy of the Peabody Museum of Natural History, Yale University

Rudolf Zallinger's 1943 tempura study for his masterwork, *Age of Reptiles*, prepared him for the daunting task of painting the 110-foot-long mural at Yale's Peabody Museum of Natural History. This detail shows a few of the creatures depicted in the mural, specimens spanning from the Devonian to the Cretaceous periods, based on paleontologists' understanding at the time. From *Paleoart*.

name for the medium, *lumia*, never caught on.) His experiments with manipulating light into art began in 1919, which meant that it was necessary for him to invent many of the technologies that made his work possible. Early in his career, he focused on producing vibrant, abstract performance artworks comprised of vividly hued, sinuous light forms that would flicker, swell, and diminish on a projection screen for audiences who would watch the complex interplay of colors. The effect somewhat resembled the aurora borealis, a look that remained his signature style throughout his career. Wilfred invented a device, which he dubbed a *clavilux*, for these performances; it resembles modern soundboards, and he'd play it somewhat like an organist. Although the performances were silent, he thought of them as orchestral, and he named his work in the style of symphonic music, a practice he continued after moving on to create installations that played automatically, many of which were designed to project a flow of spontaneously generated light forms (that is, they didn't repeat) and a few that were designed to be enjoyed in the home. These later works, which began appearing around 1930, prefigured the television, yet the devices somewhat resemble cabinet TVs if they'd been made by modernists. *Lumia* makes it clear why Wilfred deserves a spot in the 20th-century canon, and Keely does a terrific job discussing not only the artistic merits of the work, but also its significance in the context of technology and physics at the time. "Broadly speaking, Wilfred's precise light-bending technique mimicked that of physicists' laboratory experiments," Keely notes, "which often used angled mirrors to test the velocity of light when its path was redirected by reflection. The sophistication that the artist demonstrated in his technical work shows his deep understanding of the behaviors of light—how it traveled and how it interacted with other elements in its environment." Essays contributed by a historian, a curator-conservator pair, and a media studies scholar, focusing, respectively, on Wilfred's art in a postwar context, on the process of conserving the works' mechanical structures, and on the artist's legacy, make for fascinating reading as well. The still images of Wilfred's works do not disappoint, and photos of the equipment he developed and some of his conceptual drawings are helpful and revealing. Incidentally, the Wilfred exhibition shown at Yale has traveled to the Smithsonian American Art Museum, where it will be on display until January 8, 2018. Seeing these works in all their performative glory could well be worth a trip to the capital—there are some things even the most beautiful book can't capture.

That doesn't mean that authors and publishers won't try, though, and sometimes the effect is downright dazzling. The juggernaut of this fetching group of volumes is ***Paleoart: Visions of the Prehistoric Past*** (Taschen, 2017), by Zoe Lescaze, a writer whose background includes art history and archaeological illustration. Paging through it is like taking a time- and space-bending, multicentury global tour of natural history museums and amusement venues and getting to compare what scientists and artists from era to era believed prehistoric creatures must have looked like. The experience begins before you've even opened the book. The cover's pebbly, reptilian surface may take a few minutes to discover, as its jacket has been designed to stop readers in their tracks. There, front and center, is a grotesquely beautiful painting of an *Inostrancevia* shown in profile, its hide flecked with coppery foil that lends its tawny folds dimension. Crimson viscera dangles from its fangs like a holiday garland. Its amber eye gives the impression of regarding the viewer with cool contemplation. Painter Alexei Petrovich Bystrow has emphasized the mammal-like appearance of the saber-toothed reptile and his creation seems to exist, powerfully and unnervingly, in an interstitial world. Its body plan caught between the reptilian and the mammalian, the Bystrow's *Inostrancevia* appears to inhabit a time that is at once ours and its own. In fact, that visual collapsing of time—whereby ancient creatures appear to have traveled forward in time—is a hallmark of *Paleoart*. Across its pages, readers may certainly track the advance of paleontological knowledge, but they can also take in the effect of different modes and fashions of painting. Some of these modes make for

Top: Interfoto/Alamy Stock Photo, Bottom: EMI Archive Trust

The 1967 Telefunken Magnetophone CC Alpha (*top*) was among a crop of battery-operated devices that implemented new compact cassette technology to make recorded music affordably portable. More than 40 years earlier, design innovations had enabled manufacturers to offer smaller gramophones. A tin children's model made by Nirona around 1925 (*below*) features a small bell-shaped horn that could emit sound at a high volume despite its compact size. From *The Art of Sound*.

stunning paleoart. Works by mid-20th-century English artist Maurice Wilson, who had an affinity for Japanese and Chinese painting techniques, exhibit rare warmth and delicacy. Lescaze points out that Wilson was known for "eschewing three-dimensional illusionism altogether," rendering stylized scenes that draw and hold the viewer's eye. His watercolor of an archaeopteryx, depicted in russet and golden hues, perched on a branch with its wings spread wide, is simply breathtaking. Lescaze proves an intrepid guide through the scientific and artistic eras, and her lovely writing and good humor make the text a pleasure to read. This is a huge book—in size (oh, the stunning pullouts and gatefolds!) and in stature. Many books about prehistoric life cross my desk in a given year, but I've never seen anything like this.

The measure of a book is usually tallied by page count, binding style, or physical dimensions. For these three titles, I offer instead their combined weight, which is nearly 20 pounds. A strong bookshelf is a must for these volumes, but the yield will be hours of enlightening enjoyment.

Dianne Timblin is book review editor for American Scientist.

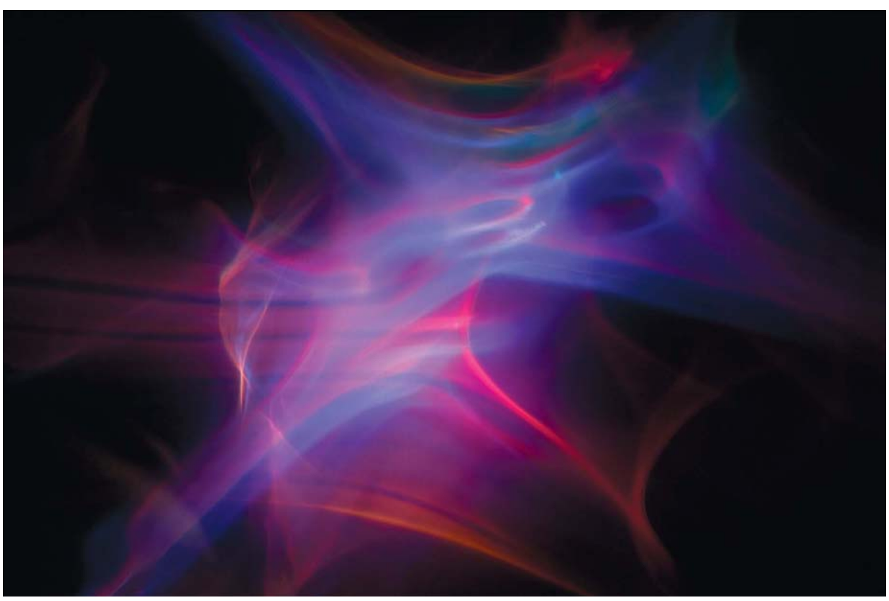


Photo courtesy of Yale University Art Gallery

Thomas Wilfred's 1959 work *Study in Depth, Op. 152* was commissioned by Clairol, which displayed the luminous, continually moving artwork on a cinema-size screen in the lobby of its headquarters building in New York. From *Lumia*.

Beyond Data-Driven Policing

Andrew Guthrie Ferguson

Big data is undeniably a big deal. The process of computationally analyzing huge volumes of data to uncover patterns and associations has become a mainstay in fields as diverse as finance and health care. In his book The Rise of Big Data Policing: Surveillance, Race, and the Future of Law Enforcement, *Andrew Guthrie Ferguson, a law professor specializing in predictive policing, big data surveillance, and the Fourth Amendment, examines the good, the bad, and the ugly of law enforcement's increasing reliance on big data analysis and its driving algorithms. Ferguson provides in-depth discussion of the often troubling ethics involved, particularly injustices stemming from implicit racial bias that all too easily becomes encoded in the algorithms. (For more on ethics and algorithms, see "A Peek at Proprietary Algorithms" on page 326 of this issue.) In this passage, Ferguson takes a step back to consider how different municipalities and their police forces intervene to try to prevent violent crime and protect potential victims after these individuals have been flagged as being at risk.*

A knock on an apartment door. A man gives the prognosis to a worried mother. Your son might die. He is at grave risk. Others he knows have already succumbed. An algorithm has identified the most likely to be stricken. He is one of a few thousand young men who may die. In Chicago, Illinois, this scene has played out hundreds of times at hundreds of doors. The danger, however, is not some blood-borne pathogen. This is not a doctor giving a cancer diagnosis but a police detective giving a life diagnosis. Violence is contagious, and you are exposed. As a young man in Chicago, due to your friends, associates, and prior connection to violence, you have been predicted to be the

American Scientist Previous Page | Contents | Zoom in | Zoom out | Front Cover | Search Issue | Next Page
Qmags THE WORLD'S NEWSSTAND®

victim or perpetrator of a shooting. Your name is on the "Strategic Suspects List," also known as the "heat list," and a detective is at your door with a social worker and a community representative to tell you the future is not only dark but deadly. A vaccine exists, but it means turning your life around now.

Data and Its Limitations

Only a tiny percentage of people commit violent crime. Police tend to know the people who commit those crimes, but the difficulty has always been getting "the criminals" to know that the police know they are committing the crimes. Enter "focused deterrence"—a theory that seeks to understand and dismantle the networks of criminal actors that drive violent crime. Focused deterrence involves a targeted and explicit message to a narrow slice of the population that police, prosecutors, and the community know who is engaged in violence and that the killings must end.

Person-based predictive policing involves the use of data to identify and investigate potential suspects or victims. Part public health approach to violence and part social network approach to risk assessment, big data can visualize how violence spreads like a virus among communities. The same data also can predict the most likely victims of violence. Police data is shaping who gets targeted and forecasting who gets shot.

While these predictive technologies are excitingly new, the concerns underlying them remain frustratingly old-fashioned. Fears of racial bias, a lack of transparency, data error, and the distortions of constitutional protections offer serious challenges to the development of workable person-based predictive strategies. Yet person-based policing systems are being used now, and people are being targeted

The Chicago Way

In Chicago, 1,400 young men have been identified through big data techniques as targets for a roster called the *heat list*. Software generates a rank-order list of potential victims and subjects with the greatest risk of violence. Based on an algorithm designed by Miles Wernick of the Illinois Institute of Technology, the heat list uses 11 variables to create risk scores from 1 to 500. The higher the score means the greater the risk of being a victim or perpetrator of gun violence. Who gets shot? The algorithm knows. And the heat-list algorithm has been tragically accurate. On a violent Mother's Day weekend in 2016, 80 percent of the 51 people shot over two days had been correctly identified on Chicago's heat list. On Memorial Day 2016, 78 percent of the 64 people shot were on the list. Using the heat list, police have prioritized youth violence to intervene in the lives of the most at-risk men.

The algorithm remains a police secret, but reportedly the factors include past criminal history, arrests, parole status, and whether the target has been identified as part of a gang. As described by the Chicago Police Department, "The software is generated based on empirical data that lists attributes of a person's criminal record, including the record of violence among criminal associates, the degree to which his criminal activities are on the rise, and the types of intensity of criminal history." The algorithm ranks these variables to come up with a predictive score of how "hot" individuals might be in terms of their risk of violence.

Selection to the heat list can be accompanied by a "custom notification visit." As described at the beginning of this chapter, it involves a home visit, usually by a senior police officer, a social worker, and a member of the community (perhaps a football coach or pastor). During the visit, police hand deliver a "custom notification letter" detailing what the police know about the individual's criminal past, as well as a warning about the future. As described in another police department document, "The Custom Notification Letter will be used to inform individuals of the arrest, prosecution, and sentencing consequences they may face if they choose to or continue to engage in public violence. The letter will be specific to the identified individual and incorporate those factors known about the individual inclusive of prior arrests, impact of known associates, and potential sentencing outcomes for future criminal acts." These custom notification letters symbolize formal messages of deterrence written in black and white. Mess up and you will be prosecuted to the fullest extent of the law. The message is also quite personal. You—the person named in the letter—are known by police and are being watched.

But the hard reality is that violence in Chicago has only increased. In fact, 2016 has seen a heartbreaking uptick in violent shootings, leading to public criticism of the model. Questions remain about the program's effectiveness and in particular whether enough has been done to remedy the social and economic risks identified. For example, there exists the open question of whether the algorithm adequately distinguishes between targets who are "high risk" (those who might get shot) and those who are "high threat" (those who might shoot). Intensive surveillance and police intervention for those who might be victims may not be as important as targeting those who might engage in violence. But if the heat-list formula counts the risk and threat equally, police resources may be misdirected.

> **Without targeted social service interventions, the algorithm just became a targeting mechanism for police.**

Stepping back, two important insights arise from the heat-list experiment. First, the public health approach of mapping social networks of violence may successfully identify those who might be involved in violence. Second, studying the data to forecast who might be engaged in violence does not automatically end the violence. Custom notifications, while well meaning, may not have the intended effect if not implemented with a focus on addressing underlying social needs. The second step of providing interventions, resources, and redirection must also accompany the risk identification. Without targeted (and funded) social service interventions, the algorithm just became a targeting mechanism for police. Plainly stated, mapping the social network of vio-

Fotan/Alamy Stock Photo

It doesn't take an algorithm for New Orleans officers to know that the French Quarter regularly requires a police presence. To predict and prevent violence more generally, the city works with a technology partner to analyze large volumes of data. Crucially, law enforcement, community organizations, and social service agencies work together on interventions.

lence may be easier than ending the violence. Data identifies the disease but offers no cure.

Big Data Meets the Big Easy

An example of a more successful, holistic approach to reducing violence can be found in New Orleans, Louisiana, once the murder capital of the United States. In 2013, the Big Easy averaged 1.46 shootings *a day*, and Mayor Mitch Landrieu turned to data to get a handle on the societal problems driving violence in the city. Private technology company Palantir has partnered with the mayor's office to identify the 1 percent of violent crime drivers in the city. The city's ambitious public health approach to violence has relied on these insights, as they illuminated largely hidden relationships in already-existing city databases.

Mayor Landrieu's project—NOLA for Life—began with data. Because the data sources included large-scale city systems with continuously generating records, Palantir engineers had to carefully integrate existing police and public-safety data into the system. This data included police calls for service, electronic police reports, probation and parole records, sheriff's office arrest and booking records, gang databases, field information cards, ballistics, and the current case-management system. The analysts also added community and infrastructure details, including the location of schools, hospitals, libraries, parks, police districts, liquor stores, and even streetlights.

Using crime-mapping software, particular violent hot spots were identified. Using social network analysis, particular individuals were identified as being most likely to be victims of violent crime. As in Chicago, New Orleans analysts could identify particular individuals at risk for violence. Analysts predicted they could identify between 35 percent and 50 percent of the likely shooting victims from a subpopulation of 3,900 high-risk individuals. In addition, linkages between these individuals in terms of rivalries, retaliations, and relationships could explain why these men were in danger. Analysts using Palantir systems identified 2,916 individuals from the general New Orleans population of 378,750 (the 2013 city estimate) most likely to be the victim of homicide.

The data insights, however, continued beyond people. On the basis of the Palantir analysis, the Fire Department increased its presence around particular schools, and the Department of Public Works repaired missing streetlights. The Health Department targeted high-risk schools for violence prevention, and police mapped gang territory to identify areas of tension. Alcoholic-beverage enforcement targeted liquor-store violations, and neighborhoods were targeted for street cleanup. All of these localized interventions came from the same data set that mapped crime, governmental services, and public infrastructure.

Building from the data, the city launched a holistic strategy to address violence reduction focused on those who were identified as being most at risk. Some focused-deterrence policies were implemented with call-ins, "stop the shooting meetings," and police targeting of suspected offenders. Since 2013, a multiagency gang unit has indicted 110 targeted gang members. But a host of other, non-law-enforcement social-services programs also were enacted (and funded). These programs included violence-reduction measures involving mediators, violence interrupters, community first responders, and other individuals who worked to defuse conflict and thus prevent retaliatory shootings. City officials also improved social-services programs addressing family violence, mentoring, fatherhood classes, behavioral interventions, and other mental and physical health concerns for those who were at risk. Programs focused on redirecting tension in public school by addressing trauma and on building restorative justice principles into discipline systems. All told, New Orleans adopted 29 different programs focusing on family, school, job training, reentry, and community and economic development. The goal of all of the changes was to target those individuals most at risk of committing violence and then to provide alternatives, support, and an opportunity to change.

From 2011 to 2014, New Orleans saw a 21.9 percent reduction in homicide, a better statistic than could be found in similar big cities such as Baltimore, St. Louis, Newark, or Detroit. More impressively, the city saw a 55 percent reduction in group or gang-involved murders.

New Orleans's holistic approach to predictive policing did much more than identify the high-risk people in a community. By using public data to address factors that created an environment for crime, the New Orleans project looked to widen the lens of big data technologies. Funding public resources to respond to the underlying economic and social problems appeared to offer more long-term solutions. Big data alone cannot stop the shooting without resources to address the underlying causes of violence.

Nov.–Dec. 2017
Volume 26
Number 6

Sigma Xi Today

A NEWSLETTER OF SIGMA XI, THE SCIENTIFIC RESEARCH HONOR SOCIETY

Q & A with Fenella Saunders

Fenella Saunders is the new editor-in-chief of *American Scientist* and director of science communications and publications for Sigma Xi.

Are there changes to *American Scientist* that readers can expect under your leadership?
American Scientist remains a premier publication for longform narrative stories about all areas of science, as told by the researchers themselves. We will continue to work with scientists to bring their process of discovery to a wider audience. In recent years the magazine staff have worked hard to offer our readers articles that convey researchers' unique voices, vividly tell their stories, and feature dynamic layouts with great custom graphics. We'll continue to customize our content to bring readers the best experience possible in print, while also expanding online blogs, features, and multimedia. Watch for more podcasts, as well as live interviews and other opportunities for readers to interact with scientists.

Where do you hope to focus your efforts as director of science communications and publications?
In addition to *American Scientist*, Sigma Xi publishes *Chronicle of The New Researcher* (CTNR), a pioneering journal that gives budding researchers in high school the opportunity to publish professionally reviewed research. This cause is close to my heart, because I started my own research career in high school. So I hope to expand CTNR's reach and impact. Sigma Xi has a great history of involvement in science ethics and nonpartisan policy, and the Society has vast potential to contribute more to these areas.

Sigma Xi Today is edited by Heather Thorstensen and designed by Justin Storms.

From the President

Who is Responsible?

President Stuart L. Cooper

I would like to reflect on some of the core elements of Sigma Xi's reason for being and what members bring to the table. The Society was formed more than a century ago as a means to celebrate accomplishments in research as well as to promote collaboration and companionship throughout the various disciplines. Sigma Xi is an honor society, meaning that peers and mentors nominate members, who must be approved for induction by a well-defined process at the local chapter level or by the Committee on Qualifications and Membership. It is expected that members will hold to the highest standards of research ethics and promote the public understanding of science and technology.

Many of these values are emphasized at Sigma Xi headquarters. An example of this involves our leadership giving voice to concerns of the scientific community in terms of the health and funding of science and technology and the need for research to underpin governmental policy issues. Our involvement in the March for Science this past April exemplified these goals. Other important goals, such as promoting public understanding of the importance of research and encouraging youth toward careers in science and engineering, are driven forward by the joint actions of our headquarters and our local chapters. Finally, after all is said and done by our organizational and local chapter leadership, it is incumbent on individual members to have a role in these endeavors.

What does this mean? Sigma Xi members have been elected to a scientific research honor society and have assumed nontrivial responsibilities of membership in our Society. It is more than just a line on one's resume. Foremost is a requirement that as long as one's career has a research component that the member pursue new discoveries with the same drive and enthusiasm that brought him or her to membership in Sigma Xi. Members should also be active in nomination of the next generation of inductees and, if possible, participate in local chapter activities.

The future workforce needs to be well populated by individuals with science and engineering skills that are applied with diligence and integrity. Get involved wherever possible in K–12 activities, whether by visiting classrooms, judging at science fairs, mentoring, or supporting excellence and innovation in precollege science teaching. Engaging the public about the importance of research is something in which we can all participate and should be willing to do.

Sigma Xi will continue to have a meaningful role as long as individual members are active participants in that for which we stand.

Stuart L. Cooper

Stuart L. Cooper

Symposium Will Address Climate and Health

High school, undergraduate, and graduate students will compete for poster presentation awards at the Student Research Conference after Sigma Xi's symposium. (Image courtesy of Robb Cohen Photography and Video.)

Symposium on Atmospheric Chemistry, Climate, and Health Featuring a Public Town Hall about Climate Change
November 10, 2017
Student Research Conference
November 11, 2017
Raleigh Convention Center
Raleigh, North Carolina: www.sigmaxi.org/fallsymposium-src

You are invited to Sigma Xi's 2017 fall symposium, a gathering of climate researchers, Sigma Xi members, and the public on November 11 in Raleigh, North Carolina, to discuss critical issues facing our planet. Join us for in-depth presentations and thought-provoking discussions about the effects of climate change on human health and the environment.

The registration fee is $75 and includes lunch. Veterans and active military members receive a 50 percent discount in honor of Veterans Day.

The symposium will feature the following sessions:

The Past Is Prologue: Lessons Learned for Science-Based Policies to Address Air Quality and Climate
How did science-based policies of the past lead to solutions for air quality and climate? What are some potential paths forward for current problems?

The Ozone Hole: From Discovery to Recovery
What science, policies, and actions were successful in putting the ozone layer and ozone hole on the path to recovery?

Air Pollution and Health: Scientific and Public Policy Controversies
Does a safe threshold of pollution exist?

Humanity and Global Warming: Views from the Carbon Cycle
How will today's energy decisions impact the future?

Simulation and Forecast of Infectious Disease: Environmental Determinants and Transmission Dynamics
How do scientists predict the spread, growth, and timing of disease outbreaks?

Student Research Conference
Science and engineering students are invited to the 17th Annual Student Research Conference on November 11 for a poster presentation competition, professional development workshops, networking, and a chance to connect with the world's largest research honor society.

Get the group discount! Groups of 25 students or more from the same institution receive a 10 percent registration discount.

Members Share Solar Eclipse Stories

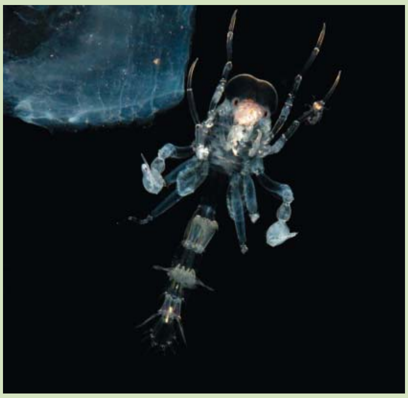
Brian C. Lee watched the total solar eclipse on August 21 from his driveway in Corvallis, Oregon.

Game Day…for Science!
The College of Charleston in Charleston, South Carolina, threw an eclipse party for 1,500 students, faculty, and staff members, and the party served as a staging area for NASA TV's national total solar eclipse coverage. The excitement led Alicia Caudill, College of Charleston executive vice president of student affairs, to say, "It looks like game day…for science!" Astronomy faculty and students took the event off campus to 16 locations, including parks and retirement homes. *—Jon Hakkila, associate dean of the Graduate School and professor of physics and astronomy, College of Charleston*

We Drove 700 Miles
My wife and I traveled some 700 miles from Fort Lee, New Jersey, to Madisonville, Tennessee, to see the total solar eclipse. We now want to do the same thing again in 2024! *—Dickson Despommier, emeritus professor of public and environmental health, Columbia University*

Read more eclipse stories at www.sigmaxi.org/2017-eclipse.

Hiding Where There Is No Place to Hide



A hyperiid amphipod (Image courtesy of Laura Bagge.)

Most people never see one of Laura Bagge's favorite animals, a crustacean called *Phronima* that is nearly transparent to hide itself from predators in the open sea. Bagge finds them by looking for the one part of their body they can't camouflage—the four tiny black dots of the animals' eyes.

"My colleagues often joke that I can't be sure if I've actually caught one of the clear crustaceans in a jar because you can't see anything in there," she said.

Sigma Xi awarded Bagge, a Society member who recently received her PhD in biology from Duke University, funding from its Grants-in-Aid of Research program to support her study on seven species of transparent crustaceans called hyperiid amphipods. She published her research in *Current Biology* in 2016, and her work has been featured in the popular press, such as *National Geographic*.

She found these crustaceans carry an antireflective optical coating that may be made of living bacteria. The coating reduces surface reflections, similar to the thin-film coatings that can be applied to eyeglasses to reduce glare. It helps make the clear crustaceans even harder for predators to find.

Learning how organisms make their bodies transparent is interdisciplinary work, requiring expertise in biology, chemistry, physics, and nanoengineering. Bagge hopes what she and others learn can be applied to nanomedicine, the physics of imaging systems, and even military defense.

As a postdoctoral associate in the Department of Chemical Engineering and Materials Science at University of California, Irvine, she will continue her research on how ocean animals use camouflage.

Duke University PhD graduate Laura Bagge in the submersible Alvin.

"I am most inspired and motivated by placing myself in the environment where these animals live," she said, "because it helps me understand the pressures these animals face, such as trying to camouflage themselves in the open water where there is nowhere to hide."

New Partnership Will Support STEM Professionals and Students

STEM IN THE PARK
Research Triangle Park

A new partnership between Sigma Xi and STEM in the Park will work toward the goal of ensuring that every science, technology, engineering, and math (STEM) professional has career support and that every child is given a fair opportunity to pursue a career in STEM.

STEM in the Park is a program of the Research Triangle Foundation in Research Triangle Park (RTP), North Carolina, and a US2020 city coalition site. Its mission is to unlock and share STEM knowledge and experiences from local STEM companies and higher education institutions to K–12 students and educators in the Triangle region of central North Carolina.

"We are looking forward to working with STEM in the Park to create new outreach opportunities for our members and to help increase diversity in STEM," said Jamie L. Vernon, executive director and CEO of Sigma Xi.

STEM in the Park will provide meeting space and consult on programming for *American Scientist's* Pizza Lunches, monthly public presentations about cutting-edge scientific research. Made possible by funds from the N.C. Biotechnology Center, Sigma Xi, and the Research Triangle Park Chapter of Sigma Xi, the Pizza Lunches are held in RTP.

Additionally, STEM in the Park will become a consultant on Sigma Xi's K–12 programming, which currently includes a scientific journal for high school research, an opportunity to network with professional researchers as a Sigma Xi Explorer, an in-person research presentation competition, and an online research presentation competition.

STEM in the Park will also provide Sigma Xi's members with new opportunities to mentor and reach out in other ways to 5th- through 12th-grade students, especially minority groups, girls, and students from low-income backgrounds.

American Scientist Previous Page | Contents | Zoom in | Zoom out | Front Cover | Search Issue | Next Page
Qmags THE WORLD'S NEWSSTAND®

New Quarterly Conversations

Last quarter's conversation was about the reproducible science issue; this quarter's topic is open access

Sigma Xi launched an initiative in July to promote and cultivate quarterly conversations about critical issues in research. Executive Director and CEO Jamie L. Vernon prompted members to discuss how Sigma Xi can contribute to solving the reproducible science crisis, in light of recent findings about the irreproducibility of many published studies that have hurt the public perception of science.

"Not only should Sigma Xi members be calling on all of their qualified students and colleagues to take the Sigma Xi pledge, but also we can take meaningful steps toward addressing the problem of reproducibility," he wrote in his monthly Sigma Xi Speaks message.

Members sent in their thoughts and contributed in the online community, The Lab: Members to Members. The ideas so far include taking the following steps toward a solution:

- **Create a certified reviewers group**. As opposed to reviewing submissions for publication, this group would write critiques of published papers, acting as a watchdog for the research enterprise.

- **Start a Best Papers Club** wherein membership requires excellence in scientific writing.

- Find ways to address the problem that **too many researchers are chasing too little funding.**

- **Create a website** that documents the number of times a given result has been replicated. Provide attribution to the researchers involved and the details of the replication process.

- **Educate the public** about the challenges and processes of scientific reproducibility, including scientific methods and analysis, so they understand why studies sometimes contradict one another or may fail to be reproduced.

Though some of these ideas draw from existing initiatives, they represent Sigma Xi's first foray into this area. We will explore ways to put these recommendations into action, possibly as a report that could be shared with the research community or as a conversation during the 2018 Sigma Xi Annual Meeting. We encourage all members to gather together to discuss this important topic.

Q2 FY2018 Topic: Open Access

The process of making scientific results freely available to anyone could improve the dissemination and use of scientific results, especially for students, who often can't afford subscriptions to research publications, and for the public, who through taxes pay for much of the research conducted at universities and government laboratories. Publishers, however, add infrastructure and validity to the publishing process, which requires a viable business model. Should Sigma Xi promote open access, and to what extent? What can Sigma Xi do to help improve access to scientific results for the public and students? How should scientific publishers be compensated for their work in an open access system?

Contribute to quarterly conversations in Sigma Xi's community, The Lab: Members to Members, at community.sigmaxi.org or email executiveoffice@sigmaxi.org.

Vote in the Sigma Xi 2017 Elections

All active Sigma Xi members are encouraged to vote in the Society's 2017 elections. Voting takes place online and opens at 8 a.m. EST on November 8. Active members will receive an email from elections@vote-now.com that day with instructions on how to vote. The polls will close at 11:59 p.m. EST on December 7.

Active members may vote for president-elect, treasurer, and other open positions in the region and constituency assigned to their chapter. Members-at-large will vote for open positions in the Membership-at-Large Constituency.

The Society will announce the results on its website, www.sigmaxi.org. Members with questions or concerns should contact the executive office by emailing elections@sigmaxi.org or calling (800) 243-6534.

Open positions in this election are the following:

- President-elect
- Treasurer
- Directors: Baccalaureate Colleges Constituency, Canadian/International Constituency, Northwest Region, Southeast Region
- Associate Directors: Baccalaureate Colleges Constituency, Membership-at-Large Constituency, Research and Doctoral Universities Constituency, North Central Region, Northeast Region, Southwest Region
- Representatives for the Committee on Nominations: Area Groups, Industries, State, and Federal Laboratories Constituency; Baccalaureate Colleges Constituency; Comprehensive Colleges and Universities Constituency; Membership-at-Large Constituency; Mid-Atlantic Region; Northeast Region; Northwest Region

Information about the candidates is available on the Sigma Xi website at www.sigmaxi.org/about/2017-society-elections.


USA SCIENCE & ENGINEERING FESTIVAL
WASHINGTON, DC