How to improve childcare by RETHINKING REGULATION

Using beams of sound to HEAL THE BRAIN

Moving beyond Mendel in GENETICS EDUCATION

AMERICAN Scientist

January–February 2025

www.americanscientist.org

Audio Immersion

Reproducing the joy of live music

SIGMA XI

THE SCIENTIFIC RESEARCH HONOR SOCIETY

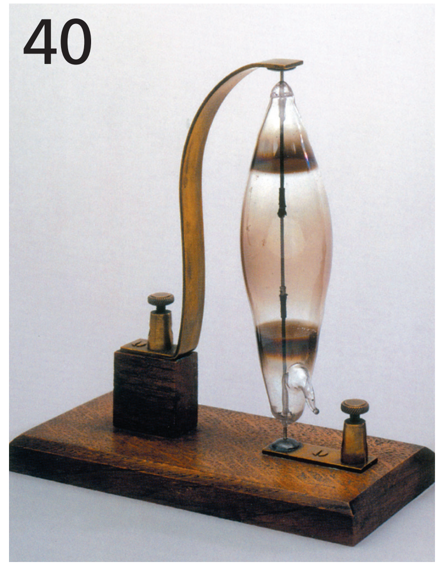
AMERICAN
Scientist
Volume 113 • Number 1 • January–February 2025

Departments

Scientists' Nightstand

From Sigma Xi

Feature Articles



The Cover

Few things in life are more satisfying yet elusive than a live musical performance. For nearly a century and a half, audio engineers have designed equipment to recapture the concert hall experience, without ever fully succeeding. In "The Science of Hi-Fi Audio" (pages 32–39), John G. Beerends and Richard Van Everdingen explain that part of the problem is human subjectivity. Beerends has helped develop precise standards for quantifying the perceived quality of spoken words, but no such standards exist for music because people vary so much in their individual responses. There is also the challenge of sonic immersion: Stereo systems produce too little, while home theater set-ups produce too much. Beerends and Van Everdingen have collaborated on a novel loudspeaker that addresses both issues, producing immersive sound that mimics the qualities of a concert hall, while allowing listeners to adjust that sound to their own preferences. (Cover illustration by Michael Morgenstern.)

From the Editors

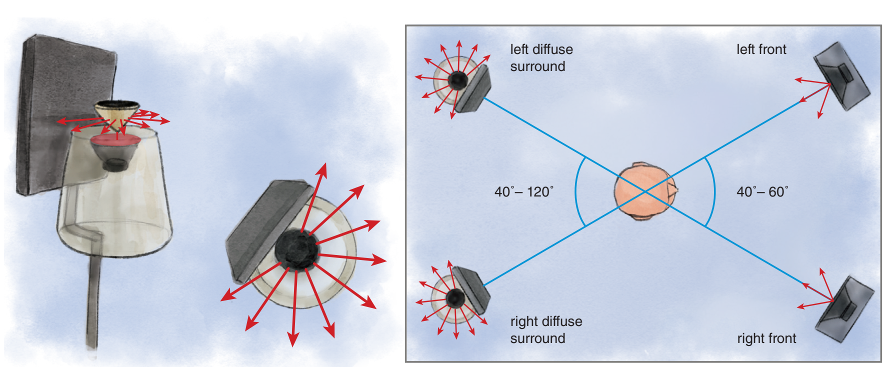
A Sense of Direction

Caught in the Moment Photography

When a full orchestra starts playing while you are sitting in a large concert hall, the music can feel like it's physically surrounding you. The sound waves are bouncing around the room, creating reflections that hit you from all sides with varying intensity. For decades, acousticians have attempted to re-create the same complex experience with recordings and stereo technology, so that listeners can get that feeling of being enveloped with music at home as often as they'd like. Multiple speaker setups can create some of the illusion, especially when replaying speech, but a virtual concert hall impression remains elusive. Part of the problem is that judgments of sound quality are subjective and difficult to quantify. But in "The Science of Hi-Fi Audio" (pages 32–39), John G. Beerends and Richard van Everdingen describe new research with directed speakers (*as shown below*) that take advantage of the acoustics of reflected sound, to produce a more realistic listening experience.

Music may be a matter of personal preference, but in science, the standards by which we judge new research findings are meant to be consistent and unbiased. That goal is part of the reason for the development of the peer-review system in research. But as Robert Pennock describes in "After Peer Review" (Science and Engineering Values, pages 22–27), the process of peer review has evolved over time, and the ways in which scientific research is deemed to be valid and distributed publicly are also expanding. Peer review is a useful tool, says Pennock, but referees should clearly understand what it can and cannot do.

A better grasp of the process of doing science is the objective of a new column that is being launched in this issue, called Scientific Method. In future installments of this column, authors will explore themes across research and delve into ongoing challenges and outstanding debates, everything from experimental setup and data gathering, to questions about how to interpret results, to philosophical discussions about the purpose of research itself. In "Finding the Rules that Work" (pages 16–21), Richard Fiene takes on the topic of regulatory science, discussing which rules actually improve quality. Do you have a suggestion for a topic we should cover in this new column? Write to us through our website to let us know. *—Fenella Saunders (@fsaundersamsci.bsky.social)*



Stephanie Freese

AMERICAN
Scientist
www.americanscientist.org

VOLUME 113, NUMBER 1

American Scientist gratefully acknowledges support for engineering content through the Leroy Record Fund.

Printed in the USA

Tetrahedron Angles

To the Editors:
In the Q&A with Miguel José Yacamán (First Person, September–October 2024), there is an arithmetic error. In the second question about five-fold symmetry, Yacamán says, "If I have a two-fold symmetry, if I rotate the crystal 180 degrees, all the atoms fall in the same place. When I rotate it by 70.2 degrees, the atoms will not match." The number should be 72 degrees because 360 divided by 5 equals 72, not 70.2.

William J. Saucier
Madison, WI

Dr. Yacamán responds:
The reader is correct in general, but this specific case is different. In a regular tetrahedron with the crystal structure of gold, the angle between the sides of the tetrahedron is 70.53 degrees. Therefore, if we pack five tetrahedra to form a decahedron (70.53 degrees 5 times), the total is 352.65 degrees rather than 360 degrees. Therefore, from the classic point of view the experimental observation of decahedra is not possible. I apologize for the confusion caused by my effort to avoid more technical data.

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.

Erratum
In "Life in the Deep Blue" by Kelly Sutherland (Nightstand, November–December 2024), the third full sentence in the right-hand column on page 378 should read: "Because short wavelengths, like blue, penetrate furthest, while longer wavelengths, like red, attenuate more quickly in surface waters, pelagic organisms have evolved in a predominantly blue world." This change has been made to the online version of the article.

American Scientist (ISSN 0003-0996) is published bimonthly by Sigma Xi, The Scientific Research Honor Society, P.O. Box 13975, Research Triangle Park, NC 27709 (919-549-0097). Newsstand single copy $7.95. Back issues $9.95 per copy for 1st class mailing. U.S. subscriptions: one year print or digital $33, print and digital $40. Canadian subscriptions: one year print $41, digital $33; other foreign subscriptions: one year print $49, digital $33. Print institutional rate: $75; Canadian $83; other foreign $91. Digital site license $200, print and digital institutional rate $275. Periodicals postage paid at Durham, NC, and additional mailing offices. Postmaster: Send change of address form 3579 to American Scientist, P.O. Box 193, Congers, NY 10920. Canadian publications mail agreement no. 40040263.

Online | @americanscientist.org

Unscrambling the Signal of Higher Vaccine Exemptions
Overcoming immunization opt-outs will require strengthening ties and trust between patients and the health sector. Brian G. Southwell, lead scientist for public understanding at RTI International; Mary Klotman, dean of the Duke University School of Medicine; and Reed V. Tuckson, managing director of Tuckson Health Connections, discuss how medical misinformation spreads and how health care providers can combat the problem.
www.amsci.org/node/5242

A Window into the Origins of Brain Surgery
Neurosurgeon Theodore H. Schwartz creates a vivid history of the field of neurosurgery in his book *Gray Matter: A Biography of Brain Surgery* (Dutton, 2024), reviewed by neurological surgery resident Margot Kelly-Hedrick.
www.amsci.org/node/5294

Some Assembly Required: A Bold New Vision of Life
NASA Sagan Postdoctoral Fellow Michael L. Wong reviews Sara Imari Walker's new book, *Life as No One Knows It: The Physics of Life's Emergence* (Riverhead Books, 2024), in which the astrobiologist and theoretical physicist posits a new theory about life.
www.amsci.org/node/5297

Annual Gift Guide
The *American Scientist* staff have made a list (and checked it twice) of their favorite STEM books from 2024 that will make perfect gifts for science enthusiasts of all ages.
www.amsci.org/node/5299

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

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

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

Find us on Instagram
instagram.com/american_scientist/

Follow us on BlueSky
bsky.app/profile/amsci.bsky.social

Brain Surgery Without a Knife

Focused ultrasound shows promise as a treatment for neurological conditions such as Alzheimer's disease and addiction.

There are few words that a patient wants to hear less than "brain surgery," but for a wide range of neurological diseases, surgery remains the only treatment option. But what about surgery without a knife?

A therapeutic technology called *focused ultrasound* (FUS) shows promise as a nonsurgical alternative to brain surgery for some ailments. In recent years, neurosurgeons have been using FUS to treat essential tremor disorder and Parkinson's disease. It is also being investigated as a way to treat a diverse range of neurological conditions including Alzheimer's disease, drug addiction, and even brain tumors, all without invasive surgery.

Patients who opt for FUS undergo magnetic resonance imaging (MRI) while wearing a special helmet that is attached to 1,024 ultrasound probes. "The sound waves are tuned so that they converge at precise locations in the brain," says Ali Rezai, the director of the Rockefeller Neuroscience Institute at West Virginia University. Where the waves converge, sonic energy is converted into heat that kills cells in a very specific location, as little as a millimeter wide. When treating patients with essential tremor or Parkinson's disease, "we can modulate the dosage or increase the energy to create a small thermal lesion to stop tremors," Rezai says.

For some patients, FUS can provide an incision-free alternative to *deep brain stimulation* (DBS), a common surgical procedure used to treat neurological disorders such as Parkinson's disease and essential tremor, as well as epilepsy and

A patient at Sunnybrook Hospital in Toronto holds out his arm to test for tremors after undergoing *focused ultrasound* (FUS) treatment. The procedure is performed within a magnetic resonance imaging machine, which allows doctors to guide ultrasound waves to specific brain regions. Prior to the procedure, the patient's severe tremors prevented him from using his right arm; after the procedure, his arm was steady, and he regained full use.

AP Photo/The Canadian Press, Frank Gunn

obsessive-compulsive disorder. DBS patients have electrodes surgically implanted directly into their brains. The electrodes, which are controlled by a pacemaker-like device implanted near the patient's collar bone, stimulate brain cells through a form of neuromodulation. The treatment works well, but it involves invasive surgery. "FUS allows patients to have a choice," says Vibhor Krishna, a neurosurgeon at the University of North Carolina at Chapel Hill. "There are patients who would choose DBS, and then there are patients who would never choose DBS. Today, we have an option for those patients."

Researchers are still trying to understand the neurological mechanisms that make treatments such as FUS and DBS effective; nonetheless, the results of these procedures can be astounding. Rezai shows me a video of one of his patients with essential tremor before an FUS procedure. "You can see that he has severe shaking of his arm and legs, and he has severe difficulty doing basic activities: writing, eating, drinking, brushing his teeth," Rezai says. "He gets the ultrasound with increased energy to the part of the brain causing the tremor, and we shut it down." In a postprocedure video, the tremors have vanished. The entire process takes only a couple of hours, and, according to Rezai, for many of the patients the tremors never return.

FUS could also potentially be used to deliver medications to the brain with greater efficacy. Medicating the brain has been a challenge for neuroscientists because of the blood–brain barrier, which separates neural blood vessels from surrounding brain tissue. Large, harmful molecules are stopped from passing from the blood to the brain, but therapeutic drugs are also swept up in the dragnet. For example, Alzheimer's disease medications such as aducanumab and lecanemab work by dissolving brain plaques that interfere with communication pathways. However, the blood–brain barrier limits these drugs from reaching their targets.

In January 2024, Rezai and his team published a paper in the *New England Journal of Medicine* that showed how FUS could be used to allow aducanumab and other medications to pass through the blood–brain barrier unencumbered. FUS agitates microbubbles that are injected intravenously. These bubbles expand and create gaps in the blood–brain barrier through which the drugs can pass. The team's study of administering aducanumab with FUS showed a 50-percent reduction in brain plaque coverage in the treated areas. "The temporary opening of the blood–brain barrier by focused ultrasound allows more of the antibody to get into the brain," Rezai explains.

Rezai is also looking to FUS therapies for treating different types of addiction through neuromodulation. Low levels of ultrasound energy are used to stimulate brain cells, but. unlike DBS, there is no need for an electrode implant. Preliminary results were published in the journal *Frontiers in Psychiatry* in 2023 as part of an ongoing study with the U.S. National Institutes of Health and the National Institute on Drug Abuse.

"A basic 20-minute neuromodulation treatment can result in sustained reduction in cravings and drug use more than three months after the procedure."

Rezai's hope is that FUS treatment "can reset parts of the brain that are focused on addiction or mental health problems without creating a lesion. Drugs, alcohol, gambling—it doesn't matter. It's the same part of the brain that is electrically supersensitive. We deliver ultrasound waves to calm the hypersensitivity of that part of the brain, calming the neurons." Although Rezai's team has published results from only four patients, "what we've seen is a basic 20-minute neuromodulation treatment can result in sustained reduction in cravings and drug use more than three months after the procedure," he says.

Substance use disorder and other types of addiction are more complex biopsychosocial phenomena than are the tremors brought on by Parkinson's disease. Rezai is emphatic that for people suffering from addictions and other mental health issues, technologies such as FUS will need to be accompanied by therapy and broader social supports.

"It's the only way," he says. "[FUS] is an adjunctive approach that empowers the therapist to do talk therapy and prescribe medications because the brain has been reset. The patients are more engaged in their treatment plan and more open to the therapist making an impact."

Neurosurgeons are continuing to investigate other potential uses of FUS. Krishna is currently involved in phase I clinical trials for the use of FUS for treatment-resistant epilepsy. He tells me about another trial at Sunnybrook Hospital in Toronto that used FUS on patients with brain tumors. "You open the blood–brain barrier and pair that with chemotherapy; it allows you to then treat it in a targeted fashion," Krishna says.

The opening of the blood–brain barrier also allows for a procedure called a *liquid biopsy*. "We could biopsy any material that leaks out of the tumors into the patient's blood," Krishna says. "We could diagnose tumors, but then do surveillance of the tumors to see how they respond to treatment." Rezai is pursuing a similar line of inquiry in neurodegenerative conditions. "You could sample antigens or molecules coming from the brain into the blood and get a blood test to phenotype or tissue-type a tumor without doing a surgical biopsy," he says.

This kind of procedure would certainly be welcome news to those facing the prospect of brain surgery. "It's surgery on the inside even though there's no incision on the outside," Krishna says.

Bibliography

Mahoney, J. J., et al. 2023. Low-intensity focused ultrasound targeting the nucleus accumbens as a potential treatment for substance use disorder: Safety and feasibility clinical trial. *Frontiers in Psychiatry* 14:1211566.

Rezai, A. R., et al. 2024. Ultrasound blood–brain barrier opening and aducanumab in Alzheimer's disease. *New England Journal of Medicine* 390:55–62.

Joseph Wilson is a doctoral candidate in linguistic and semiotic anthropology at the University of Toronto. His work examines how scientists use metaphor and other figurative language to communicate with one another in laboratory settings. Twitter: @josephwilsonca

Infographic | Wandering Walden

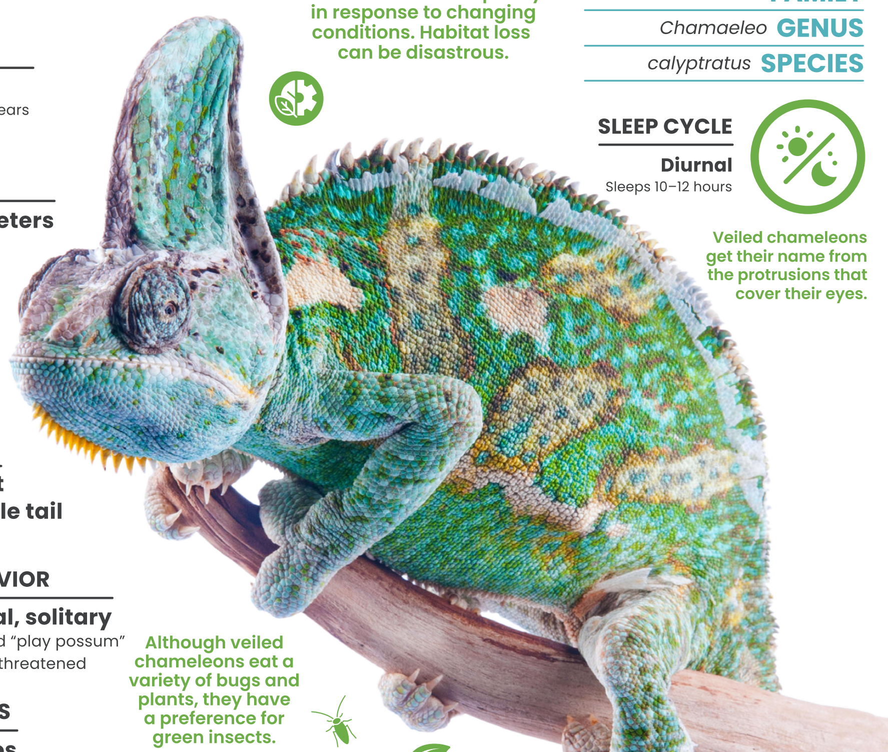
VEILED CHAMELEON

LC NT VU EN CR EW EX

IUCN Conservation Status: Least Concern — Stable

Animalia	KINGDOM
Chordata	PHYLUM
Reptilia	CLASS
Squamata	ORDER
Chamaeleonidae	FAMILY
Chamaeleo	GENUS
calyptratus	SPECIES

Because veiled chameleons are slow-moving, they are unable to relocate quickly in response to changing conditions. Habitat loss can be disastrous.



LIFESPAN
5–8 years
males live about 2 years longer than females

LENGTH
25–61 centimeters
10–24 inches

WEIGHT
85–170 grams
3–6 ounces

APPENDAGES
Grasping feet and prehensile tail

SOCIAL BEHAVIOR
Shy, territorial, solitary
Darken in color and "play possum" when surprised or threatened

ADAPTATIONS
Color changes
Can brighten, darken, or camouflage in response to environment and mood

SLEEP CYCLE
Diurnal
Sleeps 10–12 hours

Veiled chameleons get their name from the protrusions that cover their eyes.

Although veiled chameleons eat a variety of bugs and plants, they have a preference for green insects.

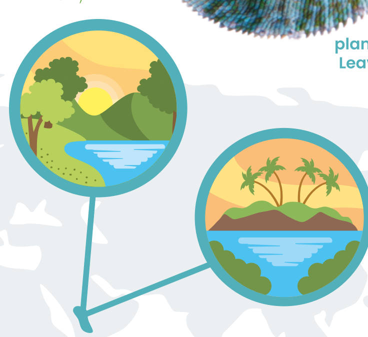
Veiled chameleons capture prey by projecting their sticky tongues.

Veiled chameleons are one of the few chameleon species that eat plants as well as insects. Leaves provide a source of water during dry seasons.

BREEDING AND PARENTAL CARE

MATING
- *Polygynandrous* (short-term pairs)
- Breed up to three times per year

OFFSPRING
- 30–85 eggs per clutch
- Incubation period: 165–200 days



Mature veiled chameleons are vibrant shades of green, tan, orange, white, and sometimes yellow.

These reptiles are arboreal lizards, preferring to live in treetops or near the ground in bushes and shrubs.

Males are fiercely territorial and should never be housed together.

This species is one of few chameleons that can tolerate a wide range of temperatures. Their preferred environment is 24 to 35 degrees Celsius.

Veiled chameleons are common pets, but they require special care and are best suited for experienced reptile owners.

RANGE
Veiled chameleons are native to the border region between Yemen and Saudi Arabia. They are found in a variety of habitats, including dry plateaus and river valleys, and can be found at elevations up to 914 meters.

Created by Abigail Lorincz
Source: Animal Diversity Web

Nanoscale Science

Paul S. Weiss is a pioneering nanoscientist at the University of California, Los Angeles, where he previously directed the California NanoSystems Institute. He studies the ultimate limits of miniaturization, exploring the atomic-scale properties of surfaces, interfaces, and biomolecular assemblies. He has developed and applied atomic-resolution scanning tunneling microscopes and spectroscopic imaging methods to measure the structure, function, and spectra of the smallest switches and motors in the world. To do so, he and his group also developed chemical patterning methods to place molecules and to control intermolecular interactions from the subangstrom to centimeter scales. He applies these advances in many areas, including quantum information, sensing, neuroscience, microbiome studies, tissue engineering, cellular therapies, and high-throughput gene editing. Weiss was the recipient of the William Procter Prize for Scientific Achievement at the 2024 International Forum on Research Excellence (IFoRE), and spoke with editor-in-chief Fenella Saunders after the conference about his work. (This interview has been edited for length and clarity.)

Courtesy of Paul S. Weiss

How did you end up specializing in nanoscience?

When I was an undergraduate, I got interested in how electronic structure and chemistry were coupled. I worked in crossed molecular beams of excited atoms, where we could point an orbital in space and see how that reacted. But I was looking for a more general way to approach the problem. I thought that on semiconductor surfaces, we could manipulate the occupation of the electronic orbitals. That would be a way to vary the chemistry.

The closest thing I could come up with was a group at Bell Laboratories, where people had figured out how semiconductors worked and invented the transistor. I worked with Mark Cardillo there, who had been slamming rare gas atoms into surfaces to excite the electrons on the semiconductor surface. I convinced him to start putting molecules on the surface, and we realized that we could detect a tiny fraction, a part in 100 million, of the reaction on the surface covered with the molecules we put down. What we didn't know was where the molecules were on the surface and what they were doing chemically.

Right about at that time, during my PhD, the scanning tunneling microscope (STM) was invented, and really opened up the nanoscale world. We could start to image atoms. There was another postdoc at Bell Labs named Don Eigler who finished his time at Bell and moved to IBM Almaden in California. I followed him out there, and we built this low-temperature STM with the idea that we could not only measure the position of molecules, but we could do vibrational spectroscopy for chemical identification. We had this grand vision for which you needed an ultrastable microscope. This instrument was on its own foundation in a separate building inside the basement of the IBM building. Eventually, we made that experiment work. It took about 13 years. But it turned out to be not that useful in the end, because we don't know the selection rules. Sometimes the signal goes up and sometimes it goes down. What we got out of that instrument that we built with this exceptional stability, where we could stay over a single atom for 10 days at a time, was that we could map the surface around an atom or molecule.

It turned out that Don and I both have a favorite rare gas: xenon. He had a dog named Xenon at the time. We put xenon down on the surface, just to see if we could image it. It didn't have any electronic states near where we were probing. But sure enough, we could image those atoms. They were sitting out in the middle of an atomically flat terrace, which didn't make any sense chemically. We reprogrammed the microscope so we could move the atoms out of the way to find out what was underneath. That was the first instance of moving atoms around with an STM and really showing that this microscope could do more than measure the structure. Later, Don spelled out "IBM" with atoms.

We could look at an isolated molecule that wasn't moving on the surface, because it was at a very low temperature. We discovered that several atoms away, the electrons of the surface were perturbed. Normally, when we think about chemistry, a tiny change, a tenth of an angstrom, is enough to change from a single bond to a double bond. This was at 100 times greater distances. We could see with the microscope what turned out to be chemical effects. We looked at what roles those perturbations had, and it turns out they're relevant in catalysis and building surface structures and a number of other areas.

It turns out there are other things we could do with the microscope as well. Spectroscopies let us look at all kinds of aspects of assemblies of molecules. Later, we got into the switches and motors on the surface, looking at the function in addition to structure and spectra. Now, we combine all those modalities together to do things like atomically resolve structures of the amyloid plaques that are thought to be responsible for neurogenic disease.

How did you create nanoscale switches?

There had been a discussion of whether a molecule could function as a wire, and later whether it could function as a conducting switch. We developed the means to isolate a single molecule in a controlled chemical environment, and then we could position our STM tip over it, and we could see it switching stochastically, and later we learned to drive it from one state to another. We developed two capabilities that turned out to be important. One is that we added the chemical dimension to nanolithography, controlling the exposed functional group on a surface. And then we also developed microscopes where we could sit our probe tip over the functional part of the surface, and then do the same measurement over and over, tens or hundreds or thousands of times, so we could work out what the mechanism of function is. In the first switches we looked at, there had been six different mechanisms proposed for how they worked. We systematically showed that all six ideas were wrong. We had to come up with our own mechanism, which turned out to be, again, chemistry related. When the molecules tilted, they changed the bonding to the surface, and that was responsible for the change in conductance.

How did you move from nanoscale molecular switches to motors?

One thing that's fascinated me is how nature uses motors that convert chemical fuel to motion with more than 99 percent efficiency. My late colleague Paul Boyer figured out how those proton pumps in cell membranes work. They're amazing. You take one ATP molecule as a fuel, you rotate the motor 120 degrees, and in three rotations, you pop a proton across the membrane. They're so efficient that you can push the proton back through and get your fuel back. Nothing humans make at any scale is remotely that efficient. We use 50 to 75 kilograms of ATP every day, but we're not made up of half chemical fuels. Rather, we can go back and forth between the fuel and its product many times every day in all of our cells.

At the same time we were studying these motors, I had a neuroscience colleague, Anne Andrews, come into my laboratory a little frustrated, because when she tried to capture proteins involved in neurotransmission, all kinds of biomolecules stuck to the surfaces. She looked at our control of the exposed chemistry on the surface and said, "Some of what you're doing might be useful." It was both flattering and insulting at the same time. I asked her what she meant and she said, look, the brain's been doing nanoscience for hundreds of millions of years—you people are way behind. Let's make a surface that prevents molecules from sticking and put, on average, neurotransmitters every five nanometers. We started working together. Later, we got married. We moved to UCLA together. That work led us into understanding that controlling the exposed chemistry at the nanoscale let us measure and control interactive biological systems.

Is that how you started working with the BRAIN Initiative?

The BRAIN [Brain Research through Advancing Innovative Neurotechnologies] Initiative came about because we were asked to put a panel together for the Office of Science and Technology Policy under the Obama White House, to come up with a grand vision of what would capture the public imagination about nanoscience. The part that I promoted really came from Anne's work to listen in on the chemical communication in neural circuits and to understand how those circuits functioned, to be able to stimulate, predict, and understand the difference between healthy and diseased brains. We put together both the technologies for voltage measurements in parallel with chemical measurements. Anne leads that technology development for making chemically sensitive and specific arrays to go in the brains of animals, and soon also humans in some studies that are coming up.

> **"Nature uses motors that convert chemical fuel to motion with more than 99 percent efficiency. Nothing that humans make at any scale is even remotely that efficient."**

This nanoscience group was brought back again to help develop the U.S. Microbiome Initiative in 2016. Some of the same sensors apply, but we needed oceanography and soil science and gynecology and dermatology and the gut microbiome. A lot of those technologies also work for wearable sensors. Anne has sensors developed for measuring cortisol in sweat as a continuous stress monitor, and one for continuous glucose monitoring.

How does the U.S. Microbiome Initiative connect to nanoscience?

The same sensors can target just about any biomarker. For example, in our mouths we have 200 or 300 different microbial species. Some are mutually exclusive. There's one that gives you cavities and another one that doesn't. You can have only one at a time in a particular place. One of our colleagues figured out what molecules they use to communicate. We can develop the tools to measure those and learn about how to avoid cavities just by having the right bacteria in our mouths.

We live with a kilogram or two of bacteria, some of which help us live and some of which are trying to kill us. You want to support the good ones and not the bad ones. We can do that with fairly simple choices of what we do day to day. But it's quite amazing compared with all the treatments a person needs after they're already in trouble.

How can nanoscience improve other disease treatments?

Steve Jonas, an MD–PhD student and resident in pediatric hematology oncology, worked out a project with us to come up with a way to do efficient, economical, safe, high-throughput gene editing to treat diseases such as sickle cell and thalassemia, in which there are 300,000 patients per year for each disease. If you replace 10 or 20 percent of the bone marrow with corrected cells, that's a curative treatment. There are approved drugs for this replacement, but those are over $2 million a dose, so that doesn't scale. We wanted to be able to do this treatment in one hour for a 12 kilogram child, which meant we needed to transfect a billion hematopoietic stem cells. And we wanted to be able to do it at many doctors' offices around the world. With Steve and four other clinicians who do bone marrow transplants, we developed different ways to do this work. The same technology works for cancer immune therapy, for developing genetically modified T-cells.

Hsian-Rong Tseng in molecular pharmacology at UCLA had developed this sort of silica barbed wire to capture circulating tumor cells. He noticed that there's penetration of these needles into the cells. He used host-guest chemistry to build carriers to get DNA, RNA, and protein machinery into the cell nucleus. But what he found with the barbed wire was, he could get those payloads into the cells, but just like if you jumped onto barbed wire, it's easier to get on than it is to get off. He couldn't get the cells off, but he could show that he could get those payloads into the cells.

Before I came to UCLA, we'd studied what happens when you mechanically perturb membranes. We made what are called *giant unilamellar vesicles*. They're like soap bubbles that mimic cells. We made a cup shape to mimic red blood cells, and when compressed, we noticed that pores opened up, to our surprise. We figured out how big the pores were and how long they lasted, but we had no idea what to do with that. Later, Robert Langer, Dan Anderson, and Klavs Jensen at MIT showed that if you pushed cells through a constriction in a microfluidic channel, you opened up transient pores, as we had in the model cells. You could deliver payloads to the cell nucleus that way. So we realized, that's what you do with it. The struggle they had was their channels would clog. We instead used acoustics to define a virtual channel. It turns out one of our UCLA alumni, Tony Huang, had moved to Duke University. We called him up and said, let's define virtual channels in this way and see how effective that will be. I sent four students, and it took them a week to make the whole thing work. The nice part about that is, there's no constriction. The channels cannot clog. Every time you change cell types, you need to optimize the diameter and so forth, and with acoustics all you do is change the waveform. One device will serve all.

I assumed it would be best to make the constriction down the center of the channel, but we found out that it works best to use acoustofluidics to charge the outside of the payloads to hold them on one wall, and then bounce the cells off the wall using the acoustics. We can do 12 million cells per channel per hour at peak efficiency. We made that parallel, so with 100 parallel channels, that's more than a billion cells. Cell viability is very high, more than 90 percent. Right now, the efficiency is in the 20 to 40 percent range, depending on the particular cell, which is sufficient. It's very exciting that we can do this. We haven't tested it in animals or humans yet at all, but that's coming up shortly.

How else can you use nanoscience to manipulate cells for other purposes?

A PhD student, Amir Nasajpour, got an idea from cellular agriculture, trying to grow meat and fish in a laboratory, in which there are difficulties with the scaffold. It's often gelatin, which comes from boiled cow bones, which kind of defeats the purpose of not killing animals to grow meat in the lab. Or there are scaffolds that are plastic or silk, but they need to be decorated with proteins that again come from animal sources. In addition, it's hard to scale the growth, because you're trying to grow muscle, so you need to stimulate it either mechanically or electrically. In addition, we don't just eat muscle. We eat muscle and fat in particular arrangements, depending on preferences. It's difficult to co-culture in the current setups.

Amir reverse engineered all that and came up with a combination of molecules that come from plants that have been known since the 19th century and form a liquid crystal. All the relevant cells recognize them, so the adhesion is built into the system. They're also FDA-approved. They're in cosmetics, so they're produced at scale. He sets the formulation of these molecules so their transition temperature is the incubator temperature. It turns out that bioreactors are not perfect in temperature control. They're not made to be. The small fluctuations in temperature give rise to these very large motions of the underlying scaffold, which stimulates the cells. And all the cells recognize those scaffold molecules, so we don't need to encapsulate them.

Then we wanted to get a 3D movie of the growth. We went to our super-resolution microscopy facility and gave them the scaffold to grow any cell of their choice. A few hours later, we got a call saying, sorry, we used this eye cancer, uveal melanoma, cell line that we like to use as a model, and the organoid grew too quickly for us to measure. We just looked at the phone and said, what? We went through a whole bunch of other cancer cell lines. We've now grown lung cancer, colon cancer, and pediatric brain cancer in a cell line nobody else could grow. We can also co-culture, because to make a biological twin of a tumor, you need to have the healthy cells around to recapitulate the tumor microenvironment. We've integrated with the robotic high-throughput screening that we have, the same as a drug company would use for personalized medicine, for individual patients, so you can test potential therapeutics against many copies of a tumor. And then you can also use that to look at several different patient lines and test potential therapeutics. If we can do it on the time scale of treatment, then we're in good shape to go ahead. ■

In this roundup, associate editor Nicholas Gerbis summarizes notable recent developments in scientific research, selected from reports compiled in the free electronic newsletter *Sigma Xi SmartBrief*: www.smartbrief.com/sigmaxi

When Ants Took Up Farming

The cataclysm from the meteor that wiped out nonavian dinosaurs 66 million years ago bolstered the success of the first fungus-farming ants. So say evolutionary timelines derived from the largest genetic dataset of relevant species to date. A team led by researchers at the Smithsonian Institution's National Museum of Natural History says that small, subsurface organisms nourished by plant detritus would have enjoyed a survival advantage amid the global environmental upheaval.

Don Parsons

Moreover, this shared food source would have brought ants into contact with fungi, providing more opportunities for domestication and symbiosis. The analysis of reliable genetic markers from hundreds of ants and fungi helped reconcile conflicting ideas about ant–fungi coevolution and about how current farming practices developed. Ants began farming fungus around the time of the Chicxulub meteor impact, but whether before or after is unknown. Ant and fungal evolutionary timelines closely align, which suggests that ant traits for farming and fungal traits that enabled cultivation developed around the same time.

Schultz, T. R., et. al. The coevolution of fungus–ant agriculture. Science *386:105–109 (October 3, 2024).*

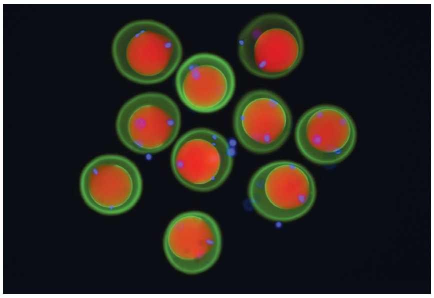
Undersea Nanocrystal Energy

Deep-sea hydrothermal vents can generate energy via channels of flowing ions, or charged atoms and molecules, according to a team led by scientists from the RIKEN Center for Sustainable Resource Science in Saitama, Japan. The findings show how a key capacity of living cells—a way to get energy from ions flowing across a membrane—can arise in nonliving systems, and hint at how life on Earth might have begun. Ions move along chemical and electrical gradients, creating flows that can be converted into energy via special nanoscale gatekeepers; in cells, protein complexes do the job. Researchers found that hydrothermal vents, too, create energy flows via ionic gradients and nanostructures. Specifically, they studied vents influenced by serpentinite from Earth's mantle, a metamorphic rock that produces a highly alkaline fluid when exposed to heated seawater. The high-pH conditions strengthen the electrochemical gradient created by ions swirling in the fluid as it flows through hydrothermal vents. The vents' walls contain layers of outward-pointing hydroxide crystals honeycombed with nanopores, which acquire different surface charges depending on which ions adhere to them. Thus, they play the same role as a cell's protein complexes.

Lee, H.-E., et al. Osmotic energy conversion in serpentinite-hosted deep-sea hydrothermal vents. Nature Communications *15:8193 (September 24, 2024).*

Sperm-Egg Bridge Found

Scientists have at last discovered how proteins in eggs and sperm recognize and bind to each other; their findings in both fish and mammals likely hold true for all vertebrates, including humans. Until now, researchers knew that certain proteins are needed for fertilization, but not always why. Curiously, proteins needed by sperm for fertilization have remained mostly unchanged since mammals diverged from their fishlike ancestors, whereas their egg-protein counterparts are evolutionarily



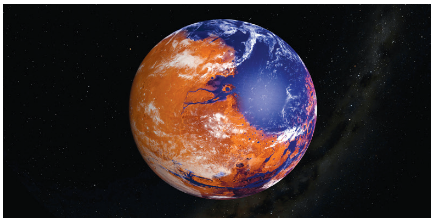
Yonggang Lu/Osaka University/IMP via AP

and structurally distinct from one another (possibly because of dissimilar challenges posed by external and internal fertilization). How could the same sperm proteins work for both? Using lab tests and a deep neural network to predict protein interactions, a team led by scientists at the Vienna BioCenter in Austria discovered a new sperm protein, Tmem81, that combines with two known sperm proteins to form a complex that can bind with both fish and mammal egg proteins, bridging the evolutionary gap. The same complex formed when researchers cultured the corresponding proteins in humans. The authors believe Tmem81 helps stabilize the sperm protein complex until fertilization.

Deneke, V. E., et al. A conserved fertilization complex bridges sperm and egg in vertebrates. Cell *187:1–13 (October 10, 2024).*

Mars Likely Lifeless

The most direct sampling and analysis ever of carbonates on Mars raises doubts that the Red Planet's surface ever supported life, according to NASA researchers. Ratios of carbon and oxygen isotopes



NASA/Lunar and Planetary Institute

(versions of elements with different masses) released and measured by instruments aboard the Curiosity rover, which gathered the samples from four Gale crater sites, suggest no biosphere existed when the carbon–oxygen compounds formed three to four billion years ago. The isotope ratios also point to extreme rates of evaporation and of carbon flowing into the atmosphere—released from solution like bubbles from soda when a pressurized can is opened. Two ideas together best explain the high proportions of heavy carbon and oxygen isotopes: A geologically rapid vacillation between habitable wet and less-habitable dry climates, and frigid midlatitude regions where water, locked up in ice, was unavailable for biochemical uses. Such ideas are not new, but this study marks the first time isotopic evidence from rock samples has been available to support them. The findings don't rule out subsurface life or exclude the possibility that surface life existed before these carbonates formed.

Burtt, D. G., et al. Highly enriched carbon and oxygen isotopes in carbonate-derived CO_2 at Gale crater, Mars. PNAS *121:e2321342121 (October 7, 2024).*

A Biting Buzz

Bees transfer vibrations better and increase pollen rewards when they grasp flowers with their mandibles.

A tasty pollen reward comes from flowers that bees visit. If the bees shake the anthers, the part of the flower that holds the pollen, do they get more out? "In a similar way to shaking a ketchup bottle, vibrating the anthers speeds up the release of pollen," says Charlie Woodrow, an ecologist at Uppsala University in Sweden. But maybe flowers have evolved shapes to limit how much pollen bees can take in one sitting. Woodrow and his colleagues decided to see if they could delve into this sort of morphological arms race between bees and flowers.

About half of the world's 20,000 or so bee species use a process called *buzz pollination*, in which the bee curls its body around the anthers and emits short, rapid bursts of vibration. And tens of thousands of plant species have coevolved to require this specific behavior in order to induce the release of their pollen—lower amplitude vibrations such as from wind gusts won't do it, Woodrow explains. The vibration was thought to transfer to the flower primarily through the bee's body contact.

Researchers also knew that buzz-pollinating bees often hold onto anthers with their mandibles, and thought that this biting was likely to prevent the bees from being shaken off the flower during the vibrations. But the bees don't constantly hold onto the anthers with their bites—they do a quick pattern of biting and releasing while they buzz. Woodrow and his colleagues wondered why the bees would let go if the purpose was to stabilize themselves.

Buzzing is an acoustic behavior, but Woodrow and his colleagues figured out that the most reliable method to measure the vibration of the bee and the flower simultaneously was visual, with high-speed video, which avoided the need to contact the bee during the behavior. To make measurements, Woodrow hand-marked a miniature scale on each of the approximately 100 flowers that the team used. "I tried having a ruler in the image, but the problem is, when the bee is on the flower, it swings the flower back and forth, so the error on the measurement becomes much larger. So I thought, if we can have the scale actually on the flower, then we know that the change in that scale with distance is going to have a much smaller effect on how we interpret the data." The team was then able to convert the displacement of the bee and the flower in the video into a measurement of vibration.

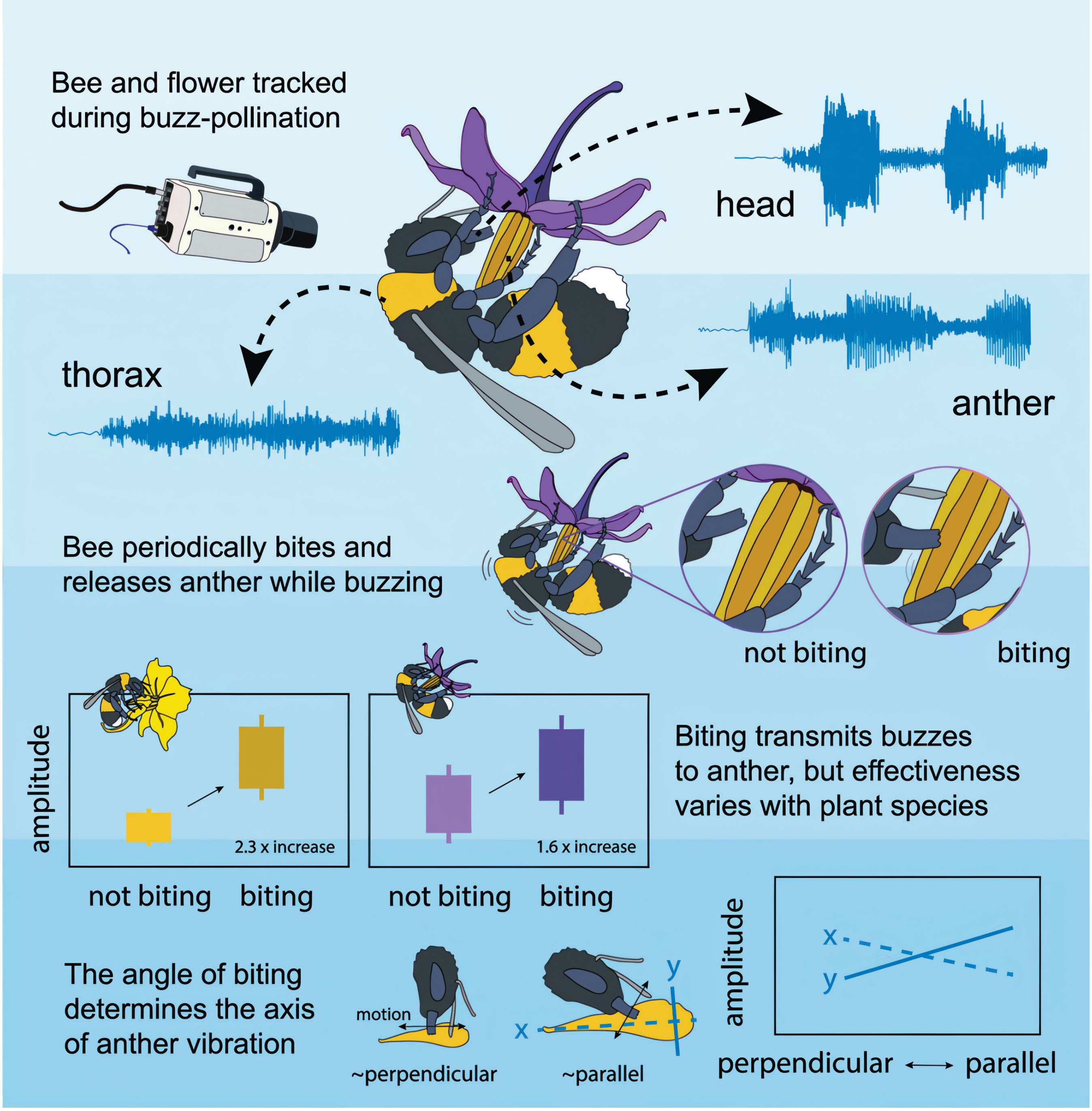
The video allowed Woodrow and his colleagues to isolate the vibration level at the bee's head and body, as well as at the anther, all at the same moment. Their data showed that biting the anther caused significantly more vibration than buzzing alone, and that the mandibles vibrate much less during biting (*see graphic at top right*). "That finding suggested to us that this biting and non-biting pattern is the way that the bees are transmitting the vibrations," Woodrow said. "It's not just a way to hold on to the flower."

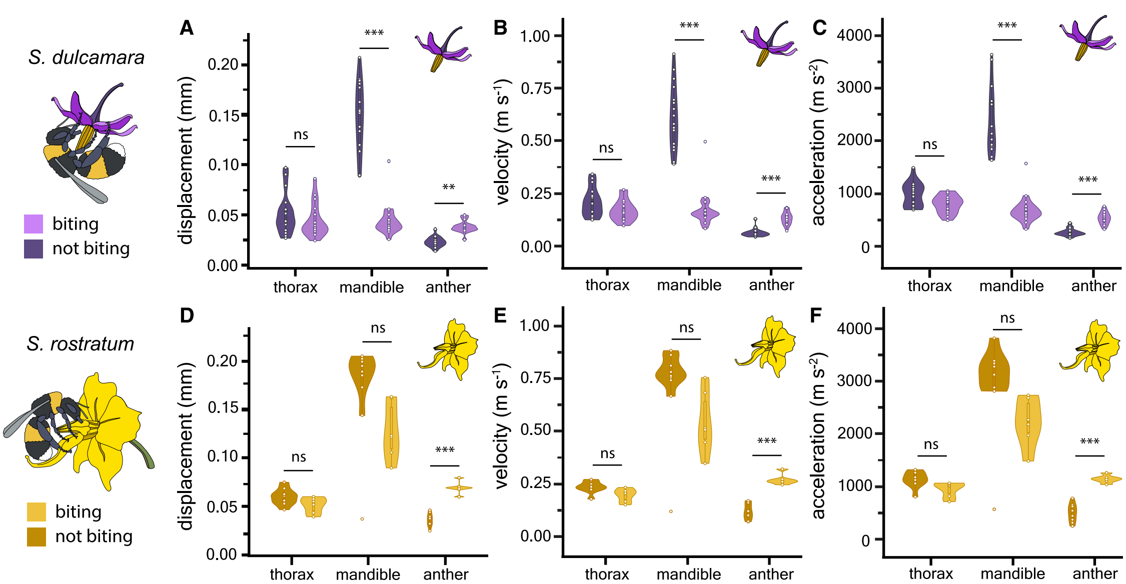
The team used bumblebees (*Bombus terrestris*) as an example bee species, and two different flowers, *Solanum dulcamara* (commonly called Bittersweet) and *Solanum rostratum* (known as Buffalo bur). Both flowers are known to need buzz pollination, but their shapes are quite different: The former is more cone-shaped and hangs downward, whereas the latter is larger, flatter, and more open. The team wanted to investigate whether the differing shapes would modify the ways that the bees interacted with the flowers. Their data showed that the biting behavior was more effective with *S. rostratum*, increasing the amplitude of vibration 2.3 times, versus 1.6 times with *S. dulcamara*.

The reason for this difference seems to come down to the morphology of the plant, because the angle that the bee can attach itself seems to be the limiting factor in an efficient transfer of buzz. "If the bee is biting perpendicular to the flower parts, then when it's vibrating the flower, it's essentially pulling on the entire flower," Woodrow explains. "But if it bites closer to parallel, then it shakes the flower across an axis that has more freedom of movement, which should in theory release more pollen." The hanging *S. dulcamara* seems to be difficult for the bees to attach to in a parallel direction, and that appears to cause a slower pollen release. The larger, more laterally angled *S. rostratum* makes it easier for the bees to get a strong parallel grip, releasing the pollen faster (*see graphic at bottom right*).

And it may be an advantage to the flowers to decrease the reward the

High-speed video captured the movement of bees as they grasped and transferred quick bursts of vibrations onto the pollen-bearing anthers of flowers. As shown in the graphic at top right, bees bite and release the anthers, and the level of vibration in the flower increases during the biting. The effectiveness of this buzz transfer varies by plant species, and by the angle at which the bees can attach themselves to the flower. The shape of the flower seems to limit this angle of attachment, which affects the rate of vibration transmitted to the anthers (*bottom right*). In the graph, *A* and *D* show displacement, *B* and *E* show velocity, and *C* and *F* show acceleration; asterisks indicate significance and "ns" indicates results that were not statistically significant. (All images from C. Woodrow et al., *Current Biology* 34:4104–4113.e3, CC-BY.)


Bee and flower tracked during buzz-pollination
head
thorax
anther
Bee periodically bites and releases anther while buzzing
not biting
biting
amplitude
2.3 x increase
1.6 x increase
not biting
biting
not biting
biting
Biting transmits buzzes to anther, but effectiveness varies with plant species
The angle of biting determines the axis of anther vibration
motion
~perpendicular
~parallel
x
y
amplitude
perpendicular ⟷ parallel


S. dulcamara
biting
not biting
S. rostratum
biting
not biting
A
B
C
D
E
F
displacement (mm)
velocity (m s⁻¹)
acceleration (m s⁻²)
thorax
mandible
anther
ns
***
**

bees receive. "Not only does it limit the amount of pollen the bee can take, so it doesn't overexploit the flower, but also it increases the time that the bee is in contact with the reproductive parts of the flower to transfer pollen from the last flower it was on," Woodrow says.

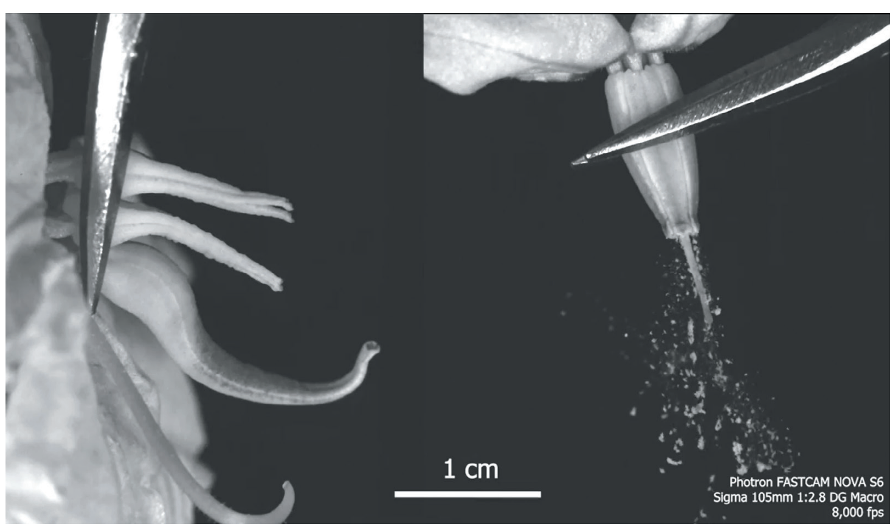
The question remains as to why the bees have this bite-and-release pattern on the flowers. Woodrow thinks it could be for several reasons, which may also vary by the type of flower: "One reason is the energetic cost; it's difficult to maintain holding onto the flower because they're using their muscles to move the mandibles as well. Another may be so that the bee could move between different parts of the flower: Maybe it's using a behavior referred to as *anther milking*, in which the bee starts at the top of the flower and moves down during the course of a buzz, which might help to release the pollen further. Another suggestion is what we call the *bellows hypothesis*, which might work in hanging flowers. If you just squeeze the flower without the bee, the pollen comes out. Maybe by biting and not biting they can force the pollen out through two different mechanisms, which should increase the speed that they can acquire this resource." Woodrow and his team plan to use artificial vibrations, either continuous or with a start-stop pattern, to count how much pollen is released, which may help determine the answer.

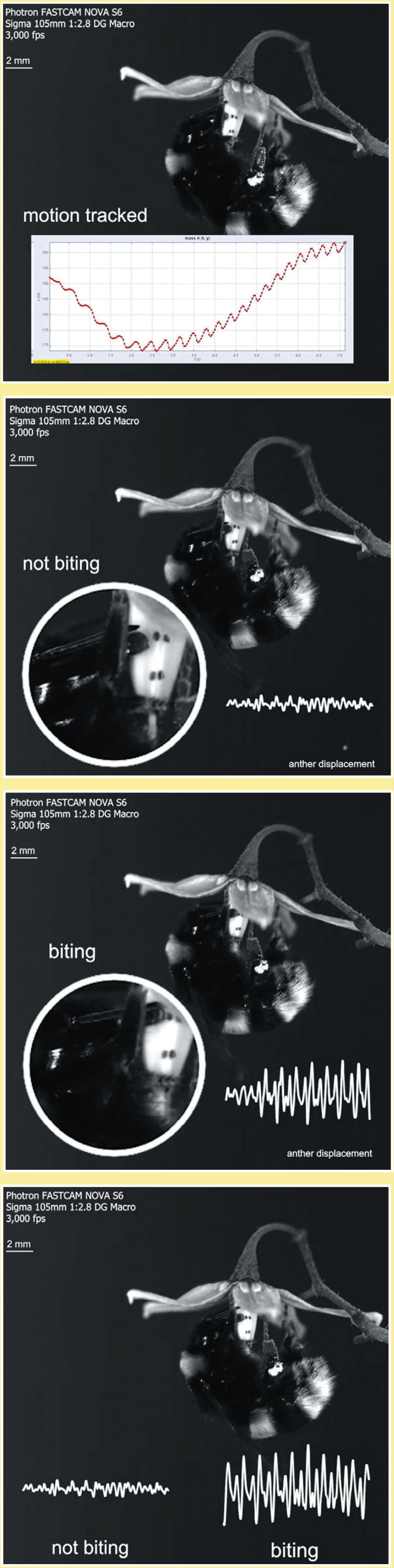
Flowers that require buzz pollination have anthers that look more like tubes, with the pollen inside rather than on the surface. Mechanical manipulation of two different flowers shows how pollen release can be controlled by compression, which may be a reason that bees bite them. The effect differs between flower species.

Better understanding this widespread pollination behavior could help a large number of bee species, many of which have seen their populations in steep decline recently. Because half of bee species are buzz pollinators, habitat loss and decrease in biodiversity may limit the options for affected bees and flowers. "Maybe there are some really exclusive couplings in which one bee species requires one buzz-pollinated flower, or vice versa," Woodrow says. "If we know this, then these habitats are the places we can focus conservation efforts."

Buzz pollination also occurs in food crops, including tomatoes, potatoes, blueberries, and kiwifruits, which can only be pollinated by bees. Woodrow and his colleagues are exploring whether it's possible to artificially emulate the buzz pollination process, to supplement declining bee populations. "If we understand the way that bees are holding the flowers, maybe we can develop microrobots that can do similar things," Woodrow says.

To find out more, Woodrow hopes to move beyond bumblebees and take high-speed video of a wide range of wild bee species on flowers. "Bumblebees are so fluffy, it's really hard to track anything on them," he explains. "Some orchid bees, for example, are super reflective, which would make a really nice video." He plans to mark up a lot more flowers next summer. "We will test our work by bringing some of our greenhouse plants into the botanical garden here, where we have lots of different bee species flying around, and then see if anything visits and try to get some good videos," he says. "We just have to sit and wait and watch a flower for many hours."—*Fenella Saunders*





Stills from high-speed video show how the motion of the bees and flower were tracked (*first image*) and the amplitude change in vibration of the flower during not biting (*second image*) and biting (*third image*). A comparison between the vibration amplitudes (*fourth image*) shows the increase in vibration of the anther during biting.

Finding the Rules that Work

An emerging paradigm promises to close the gap between regulatory compliance scores and the quality of childcare services.

Richard Fiene

An old fable recounts how a father and son, taking a donkey to market to sell it, encounter a string of critical villagers who each inform the pair they're "doing it wrong." Their efforts to please each subsequent critic end, absurdly and tragically, with them carrying the beast of burden themselves, ultimately causing its death.

Like the advice of those villagers, regulations are proffered in the name of safety and good practice. And, like that father and son, programs that try to follow every single rule to the letter may soon find themselves too weighed down to achieve (or perhaps even recall) what they set out to do. As the saying goes, "When you're up to your behind in alligators, it's hard to remember that you set out to drain the swamp."

In my four decades as a regulatory scientist studying childcare, I've seen this pattern play out time and again: In the lead-up to evaluations, staff at perfectly compliant programs spend so much time dotting i's and crossing t's that they have little left over for working with classrooms or teachers, whereas staff at slightly less compliant facilities, though equally careful about observing rules, fuss less with paperwork and work more with teachers on improving skills and curriculum.

Needless to say, developmentally appropriate curricula change kids' lives; boasting a perfect record does not. This observation neither dismisses the 200 to 400 rules and regulations set by respective U.S. states nor undermines the importance of complying with them, either as individual rules or in the aggregate. And full compliance does improve safety. But, as data gathered by my research team repeatedly demonstrates, a vague, uncomfortable gap separates full, costly regulatory compliance from program quality.

It is never about more or fewer rules; it is about which rules are really productive and which are not.

Moreover, early care and education providers often voice concerns that licensing inspectors inconsistently administer and apply particular rules. At issue, then, are not regulations' overall value per se, but rather the value of individual rules relative to fanatical box-checking. Given their limited resources, how can the early care and education fields get the most bang for their buck?

Such a discussion is long overdue. The unequal worth of many general licensing and quality standards, including those driven by a regulatory political bent rather than empirical evidence, produce markedly uneven developmental outcomes for kids. Today, an outcomes-based scientific reference frame is already influencing the human services industry (childcare, child welfare, and child and adult residential services), particularly in the early care and education fields (childcare centers and family childcare homes for children between infancy and 12 years old). The point of my team's approach, which I call the *theory of regulatory compliance*, is not to ask whether we need more or fewer rules, or more thorough or less thorough compliance, but rather to evaluate which rules truly prove effective.

Modernizing Measurement

Regulatory scientists use tools, standards, and methodologies to assess the safety, efficacy, and quality of programs under government regulation. Ideally, they help regulatory agencies achieve the best possible public health and safety outcomes.

The regulatory science field has a lot of ground to make up. At about 30 years old, it lags its subject matter by a good century (Pennsylvania passed the first orphanage licensing law in the United States almost 140 years ago). Human services licensing grew slowly prior to the late 1960s to early 1970s, when American President Lyndon B. Johnson began the Great Society initiatives such as Head Start, which kicked off the rapid multipli-

QUICK TAKE

Contrary to historical assumptions, the quality of childcare programs does not increase linearly as their compliance with rules and regulations approaches 100 percent.

All-or-nothing, one-size-fits-all approaches to compliance and licensing generate skewed data, raise risks of false negatives and false positives, and burden staff with bureaucratic tasks.

Substantial regulatory compliance is an alternative approach that emphasizes compliance with the most productive rules, preserves safety, and allows staff to concentrate more on children.

DGLimages/Shutterstock

Staff of fully compliant childcare programs say they spend too much time box-checking and not enough working with teachers, whereas staff at slightly less compliant facilities, though equally scrupulous, bother less with form-filling and spend more time in the classroom. An outcomes-based substantial regulatory compliance approach lets licensors strike that balance.

cation of childcare programs. Those decades also saw human services, especially childcare, begin transforming from cottage industries, with program monitoring and measurement conducted qualitatively via case notes and anecdotal records, to more rigid systems that entailed oversight, case reviews, and state agency inspections. In the 1970s, these systems, which often varied from state to state, gave way to improvements brought by the Federal Interagency Day Care Requirements.

The watershed moment for regulatory science as it pertains to children's programs came in the 1980s. The previous decade's major childcare expansion in the United States had created a backlog of licensing assessments, caused unmanageable monitoring delays, and laid bare the logistical limits of case studies. These factors, combined with advances in computing, led states to introduce an empirical, quantitative, and instrument-based approach, complete with sophisticated software systems designed by state agencies and private vendors to track regulatory compliance and quality assessment data. Empirical evidence not only moved regulatory science from qualitative to quantitative analysis, it also revealed surprising patterns.

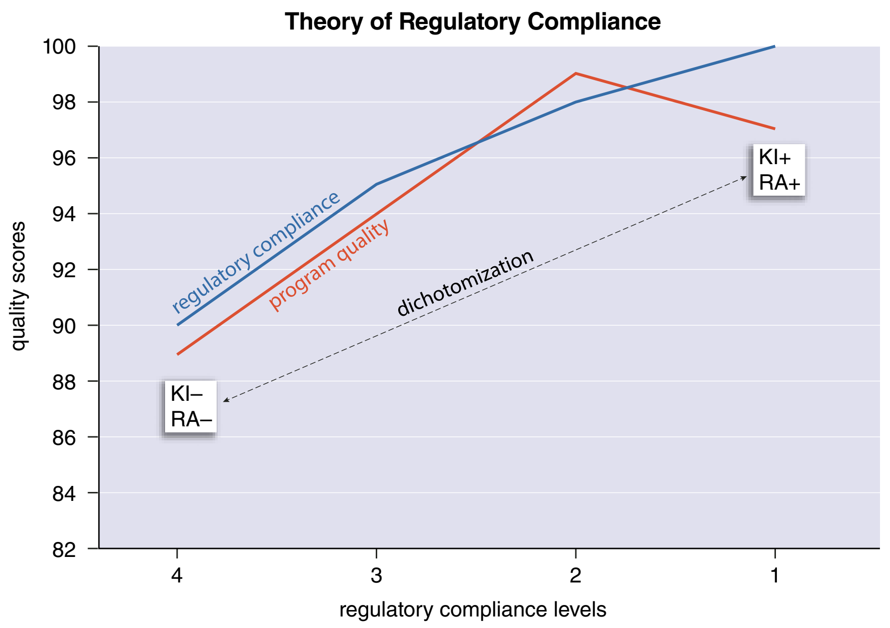
But first, some background: As the U.S. Department of Health, Education, and Welfare took over running the show for all U.S. early care and education programs in the 1970s, *uniform program monitoring* had become the rule. Uniform monitoring derived from the philosophical assumption that fuller regulatory compliance would produce, linearly, better quality across U.S. early care and education programs. As the former went up, so would the latter. From a public policy standpoint, this notion sounds aspirational, but sensible: Any licensing agency looks for service quality to increase as its rules, regulations, and standards are followed.

But as expert opinion and anecdotal evidence gave way to better-designed studies and empirical data, and as larger studies became possible thanks to data computerization by state licensing agencies, cracks appeared. When researchers compared violations found during licensing reviews and inspections to the quality of the violating programs, they found that a linear relationship did indeed exist between quality and compliance—but only as one moved from low compliance levels to *substantial regulatory compliance* (that is, 98–99 percent). Between that and 100 percent compliance, quality consistently plateaued and, as some 2010s replication studies suggested, even showed diminishing returns.

A New Paradigm

These results called into question the notion that state agencies should issue licenses solely to fully compliant programs. If, as data suggested, substantially compliant programs provided the same or better care as fully compliant ones, then clearly, we needed to rethink our program evaluation strategies.

In the United States, state licensing and regulatory agencies establish childcare regulations, but federal agencies such as the Office of Child Care and the Administration for Children



Adapted from Richard Fiene

This graph shows the quality scores (*y*-axis) associated with four categories of regulatory compliance (*x*-axis, defined by the number of rules violations, ranging from 0 [Level 1] to 10 or more [Level 4]). Note that compliance scores (*blue line*) and quality scores (*red line*) rise together, but only until substantial compliance (99-97 percent compliance with all rules [Level 2]) is reached. This finding argues for the adoption of substantial compliance as a standard, and for utilizing differential/relative monitoring to better capture nuances of quality and more efficiently allocate resources. The alternative—a punitive, gatekeeping licensing approach requiring full compliance (a yes/no proposition)—has led to highly skewed data. Here, the author has split (dichotomized) these skewed data into two extremes: Programs with regulatory compliance scores in the top 5-10 percent (*upper right*, labeled KI+/RA+ to indicate positive key indicator and risk assessment findings) and the bottom 5–10 percent (*lower left*, labeled KI-/RA-). The graph shows how scores in key indicators and risk assessment effectively predict program quality.

and Families also influence rules, as does Congress through its funding purse strings. Sometimes cities and counties, too, set regulations or standards, especially concerning physical environment, health, safety, and zoning. (Here, the term "regulations" means those defined by the National Association for Regulatory Administration's Licensing Curriculum.)

For an individual program or facility to operate, a state licensing agency must judge that it follows these standards. Examples include certifications for teacher qualifications, first aid, CPR, and the facility environment, along with requirements for ongoing training and professional development. State licensing staff evaluate compliance via inspections, document reviews, audits, and interviews, usually on a yearly basis. Inspections check for health, safety, cleanliness, educational standards, and staff-to-child ratios, as well as less obvious standards such as playground and transportation safety. Noncompliant programs may face fines, mandated corrective actions, training, or technical assistance, or may undergo license suspension or even permanent closure.

Licensing requirements vary depending on the childcare offered (such as family childcare homes, center-based care, or school-based programs), with larger centers typically facing more stringent requirements. Along with compliance ratings and violations issued by licensing inspectors, these facilities voluntarily seek ratings from quality initiative offices within human services agencies.

Here, and in my research, I primarily deal with center-based care programs, but the findings apply to other service types as well, such as family childcare homes and school-age programs, as well as human services categories such as child residential, child foster care, adult residential, and adult personal care homes. My data and research concern the relationship between quality and compliance, and how to improve it. They stem from studies of hundreds of programs I conducted at the state level from the 1970s through the 2010s, when I directed various research and training institutes at Pennsylvania State University. In these controlled and replicated studies, trained observers collected both regulatory data and program quality data from eight states, three Canadian provinces, and the U.S. Head Start program. The work ran the gamut, from site selection via stratified random samples, to dispatching data collectors to specific programs, to providing individual states with an overall blueprint describing how to conduct their studies.

Initially, the ceiling effect between regulatory compliance and program quality came as a surprise; we did not predict that full compliance would fail to outperform substantial compliance. It also drew pushback from the licensing field. Thus, I replicated the study many times over to assess my assumptions. But the finding persisted: Program quality scores rise with regulatory compliance until programs reach substantial compliance, after which quality declines. Although until 1980 states required childcare programs to show full compliance and zero violations, since 2015 most states have allowed licensing for facilities that are substantially compliant.

Differential Monitoring

If substantial compliance with some rules rather than full compliance with all rules best ensures the childcare program quality, then the question naturally arises: "Which rules?" Conceivably, some rules should weigh more heavily than others—say, the ones that data show most closely relate to safety and quality. Such is precisely the idea behind *differential monitoring*.

Differential monitoring emerged in 1979 during my discussions with federal agencies such as the Administration for Children, Youth and Families and the Children's Bureau, who felt dissatisfied with the traditional uniform monitoring approach. They knew about my team's work in Pennsylvania and invited me to give a series of talks to their staff. The result was a move away from the older, one-size-fits-all approach to differential methods focused on *key indicators* and *risk assessments*.

Key indicators are statistical predictors of overall compliance—rules that, if a facility follows them, strongly suggest they will follow other rules as well. They very efficiently determine a facility's overall regulatory compliance without requiring a comprehensive inspection. Far from negligent, this approach works because not all rules are created and monitored equally.

Risk assessment focuses on those rules and regulations which, when breached, place children at greatest risk, such as rules that deal with supervision or hazardous materials handling, among others. Generally, jurisdictions, states, and provinces engage major early care and education stakeholders (service providers, parents, advocates, and licensing staff) in weighting rules or regulations based on their risks to children's health and safety. Commonly, participants assign weights via a *Likert scale*—a common survey and questionnaire tool that lets respondents indicate the strength of their agreement or disagreement (or, in this case, their assessment of risk) with a statement about attitudes, opinions, or perceptions. The weights range from 1 to 10, where 1 indicates little risk if a program fails to follow the specific rule or regulation and 10 corresponds to high risk. Rules heavily weighted as associated with sickness, injury, or death join the risk assessment rules measured by inspectors in every differential monitoring review.

As an aside, I should point out that full compliance remains the standard for maintaining health and safety. So why incorporate risk assessments into differential monitoring and, by extension, the substantial compliance paradigm, as its own separate metric? In truth, I had no such intention when I wrote my 1985 research papers about differential monitoring and the theory of regulatory compliance. Rather, risk assessment morphed from a way to provide the needed data variance for key indicator scoring into its own submethodology. As it found its way into the implementation of national standards and guidelines, risk assessment subsequently emerged as a separate methodology.

Our findings repeatedly show that using the combined methodologies of key indicator predictor rules and risk assessment rules to identify the "right rules" and to ensure compliance with them, rather than to seek full compliance, makes the differential monitoring approach the most effective and efficient program monitoring system. Also, studies show that abbreviated, targeted, and focused reviews take approximately 50 percent less time than comprehensive reviews.

Compliance Measurement Systems

scoring level	individual rule		aggregate rules	individual rule
scale	**instrument based**	***scale***	**differential**	**integrated**
7	full compliance	7	full compliance	exceeds compliance
–	–	5	substantial	full compliance
–	–	3	mediocre	substantial
1	out of compliance	1	low	mediocre/low

adapted from Richard Fiene

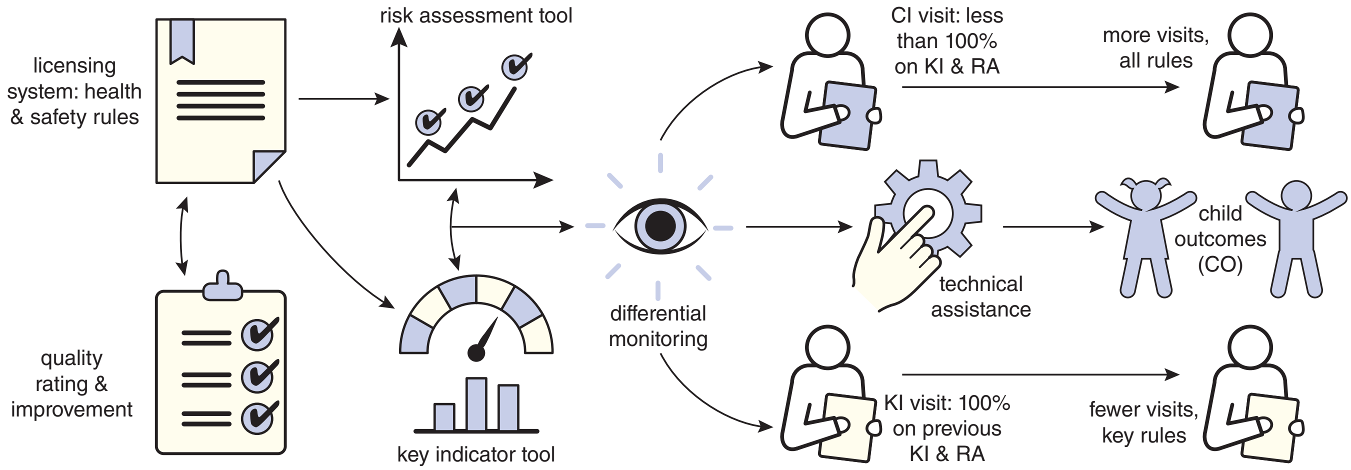
This table compares different approaches to measuring compliance: A licensing-focused approach in which programs are classified as either compliant or noncompliant based on rules violation counts, with no middle ground (*columns 1 and 2*), and a more nuanced ordinal approach using a Likert scale. This experimental metric, called the Regulatory Compliance Scale (*column 3*), is currently being tested at the aggregate rule level (*column 4*) and may be expanded to the level of individual rules (*column 5*) in the future. Note that aggregate rule scores are not equal to the sum of all individual rule scores because not all rules are created or administered equally.

Unfortunately, although many licensing bodies use risk assessment or key indicator methodologies, few use both. *Monitoring Practices Used in Child Care and Early Education Licensing*, a federal accounting of how states conduct program monitoring, reported that 10 states used key indicators, 17 states used risk assessments, and only one state used both. Hopefully, this pattern will change as the regulatory science field matures over the coming decades.

If, as data suggested, substantially compliant programs provided the same or better care as fully compliant ones, then clearly we needed to rethink our program evaluation strategies.

Since I first proposed it in the mid-1980s, the theory of regulatory compliance has faced numerous critics in the human services licensing field, especially among advocates of uniform monitoring and full compliance. Only after years of licensing validation studies conducted by my team and others repeatedly demonstrated that full compliance did not produce the highest quality did states begin licensing programs in substantial rather than full regulatory compliance. Today, although various U.S. states apply the differential monitoring review approach unevenly, nearly all have adopted the policy of granting licenses for substantial rather than full compliance. The latest revision of the legislation for the Child Care and Development Block Grant (a U.S. federal funding program that helps states, territories, and tribes assist low-income families in accessing affordable childcare) cites differential monitoring as an alternative to uniform program monitoring.

Of all the approaches and methodologies that flow from the theory of regulatory compliance, differential monitoring most significantly alters the program monitoring, inspection, review, and licensing landscape. Its reviews occur just as often as do uniform monitoring assessments but focus specifically on rule breaches shown to place children at risk. That said, differential monitoring did not replace but rather supplemented its predecessor: Comprehensive reviews must still occur every three to four years to validate the performance of key indicators and risk assessment rules. But what does that report card look like in terms of analyzable data?



Barbara Aulicino

This illustration shows the various components that contribute to a differential monitoring approach and how agencies can use them to evaluate the effectiveness and validity of different approaches. Differential monitoring allocates resources based on risk assessment (client morbidity and/or mortality) and key indicators (rules whose compliance is strongly predictive of program quality). These data, provided by mandatory licensing processes and voluntary quality rating services, reveals which programs are highly compliant with key rules (though not all rules) and therefore require fewer visits versus programs that are less compliant and require additional visits and technical assistance to achieve similar child outcomes.

Rethinking Nominal Data

Traditionally, licensing data are categorical (sorted into groups such as "approved" or "denied"), unordered (there's no built-in way for such groups to be sequenced), and mutually exclusive (state agencies cannot simultaneously deem a facility both "approved" and "denied"). In statistical terms, such data are nominal, like a table listing cars by make or model; you cannot "do math" on such a table like you can on, say, on a table listing automobile curb weights and fuel economies. It is also binary: A program either follows a rule, or it doesn't.

Presently most jurisdictions deal in these absolutes and exclude gray areas. This approach, much like uniform program monitoring and full compliance, makes intuitive sense: We create rules and regulations because we believe in the value of following them, and because licenses mean nothing if licensees are not held to a standard. But here again, we must look deeper and ask, "What consequences follow from this either/or approach to measuring compliance, and who decides whether or not a particular box gets checked?"

Let's begin with the latter question. In an ideal world, judgments made by assessors would perfectly reflect a program's actual regulatory compliance state. But research that tests reliability and replicability in the licensing field empirically shows a concerning degree of disagreement when a second observer validates the decision regarding regulatory compliance. These disagreements suggest a worrying number of false positives and false negatives.

A false positive occurs when a program follows a rule or regulation, but the assessor rules that the facility is noncompliant (which might sound backwards, but the metric is *non*compliance, not compliance, so finding a false violation means finding a false positive). But even more concerning are false negatives, in which an evaluator says a program complies with a rule that it breaches, thereby placing clients at risk. Detecting false negatives is one of the chief reasons we periodically validate the predictive value of key indicator rules through comprehensive reviews.

As for the first question, the answer is simple: Nominal, binary licensing data is severely skewed. Upon reflection, the reason becomes obvious. When a regulated industry such as childcare mandates compliance before a program can operate and excludes gray areas, most facilities will achieve full compliance or lose their licenses. Because unlicensed providers don't last long, the childcare sector produces data that skew toward licensed programs. To grasp such skewed continuous or multicategory data, we must first dichotomize it into two distinct groups.

Such sorting into piles raises statisticians' hackles; unless carefully done, it accentuates differences and forces tradeoffs between precision and sensitivity, which can mean swapping false positives for false negatives. But the nature of licensing data—a skewed collection of mostly or fully compliant programs dumped in a single bucket—makes the split both necessary and warranted. By setting a threshold of certainty or agreement among evaluators, we can more effectively reduce false negatives, that is, cases in which evaluators say a program follows a rule when it doesn't.

This need becomes even clearer when one considers the demands posed by differential monitoring and its methodologies, key indicators, and risk assessments. For a program to receive licensure, it is not enough to ask if it "complies enough overall"; we must also know if it follows the specific rules that most ensure safety. By comparing highly compliant programs only with low-compliant programs, we accentuate the differences between the two and bolster our data analyses as well as overall safety. This comports well with licensing decision-making, which can consider a program compliant or noncompliant not only in aggregate, but with respect to *individual* rules.

Infusing Quality

The all-or-nothing approach to regulatory compliance and licensing fails as a standard because it generates skewed data, raises the risks of false negatives and false positives, and springs from a false assumption that program quality increases in step with 100 percent compliance. But I am far from the first to notice that approach's weaknesses in evaluating how good a program or facility actually is. Indeed, its shortcomings helped drive the creation of a separate industry of voluntary accreditation programs such as the National Association for the Education of Young Children, state-run quality rating and improvement systems, and third-party tools and assessments. It's time we folded quality assessments into regulatory compliance.

I have already explained how the theory of regulatory compliance improves program quality and safety by focusing on substantial, not full, compliance and by using differential monitoring to ensure programs follow the most protective and impactful rules. But to further cast off the limitations and lopsidedness of a uniform monitoring and full compliance mindset, and to make room for data capable of tracking quality, we must also replace rigid either/or logic with a more nuanced ordinal measurement: a scaling technique.

Recall that assessors can evaluate compliance in two ways: They can consider aggregate rules—collections of rules that fall into categories such as staffing or safety practices—or individual rules. Each has its own studies and research literature. Research on aggregate rules from the 1970s, 1980s, and the 2010s established substantial compliance as a "sweet spot" of best outcomes and showed that the time had come to replace nominal metrics (such as "compliant" and "noncompliant") with ordinal ones (such as "98 percent compliant").

Inspired by this research, I have proposed replacing older nominal techniques with an ordinal scale like the Likert scale already used in quality measurements (usually but not always ranging from 1–7, with 1 being inadequate and 7 being excellent). This technique, currently under review by the National Association for Regulatory Administration, will help reviewers consider the importance of substantial compliance. Moreover, it will add the currently absent quality elements to each rule and regulation. However, this approach involves aggregate rules only; further research is needed to determine if the same shift from nominal to ordinal metrics should also occur at the individual rule level.

Should those findings bear out the value of evaluating individual rules via the 1–7 regulatory compliance scale, I propose that it should contain the following categories: exceeding full compliance, full compliance, substantial compliance, and mediocre compliance (*see figure on page 19*). These categories differ from the aggregate rule compliance scale currently under evaluation (full, substantial, mediocre, and low compliance) because aggregate compliance only considers health and safety elements, whereas an individual scale would also take quality into account.

Research supports the value of transitioning from uniform monitoring and full compliance to differential monitoring and substantial compliance. Practice has shown the value of retaining the older to help ensure the validity of the newer. Looking to the future, I believe we can further improve compliance evaluations by developing and evaluating *integrative monitoring*, which incorporates program quality into rule formulation and moves the key indicators from predicting compliance to forecasting quality.

The all-or-nothing approach fails as a standard because it generates skewed data, raises the risks of false negatives and false positives, and springs from the false assumption that program quality increases in step with 100 percent compliance.

Looking Forward

The regulatory compliance scale is a new and evolving metric. It transforms licensing data from a mere violation tally into a more useful and intuitive scale, one more consistent with the program quality measurements supported by research. Hereafter, I hope that the approach will incorporate quality measurements and more nuanced weighting into the evaluation of individual rule compliance. But discussions are just beginning, and this shift will pose a substantial challenge for agencies, which must also cope with the aftermath of the COVID-19 pandemic and a rising tendency toward deregulation.

The theory of regulatory compliance concerns the relationship between regulatory compliance and program quality, not health and safety, where full compliance remains the goal. It is, however, the preferred methodology for eliminating false negatives and decreasing false positives. Add to that the fact that the theory of regulatory compliance predicts a nonlinear relationship between compliance and quality but a linear relationship linking regulatory compliance and safety, and regulatory scientists clearly have our work cut out for us. Untying this knot will require greater collaboration between the historically siloed public policy worlds of licensing, accreditation, quality rating and improvement systems, and professional development systems.

I hope that the regulatory science field takes these paradigm shifts into consideration as it builds licensing decision-making systems and considers how states issue licenses. And although this work deals primarily with my own experience in the early care and education field, I wonder if other human service sectors, such as the foster care or child and adult residential areas, demonstrate similar patterns. Other disciplines that deal with regulations and compliance may similarly find it fruitful to discuss the nuances of their own evaluation metrics in order to achieve the best overall outcome with the most efficient use of limited resources.

Bibliography

Fiene, R. 2023. Licensing measurement and monitoring. National Association for Regulatory Administration licensing curriculum. Published online www.naralicensing.org/key-indicators

Fiene, R. 2022. Regulatory compliance monitoring paradigms and the relationship of regulatory compliance/licensing with program quality: A policy commentary. *Journal of Regulatory Science* 10:1–7.

Fiene, R. 2019. A treatise on regulatory compliance. *Journal of Regulatory Science* 7:1–3.

Richard Fiene is an emeritus professor of psychology at the Pennsylvania State University and has had a lengthy career in regulatory science and developmental psychology, in which his research has led to the development of national early care and education standards. He directs the Research Institute for Key Indicators Data Lab. Email: RFiene@RIKInstitute.com

After Peer Review

What role do referees play in science?

Robert T. Pennock

At the Lansing, Michigan, site of the 2017 March for Science, an international event to advocate for the importance of science, I was especially taken by one marcher's placard that was a humorous twist on a classic demonstration call and response. It read:

What do we want?
Evidence-based policy.
When do we want it?
After peer review.

The message was properly nonpartisan—the expressed sentiment was simply that governmental policies ought to be grounded in evidence rather than ideology, with the understanding that patience would be required for the necessary review. The motivating idea here is that peer review helps make policy trustworthy. In democratic deliberation, such trust is essential if decision-makers are to have a common, justified basis for action. What is important to remember, however, is that trust in science for policy and other external applications is derivative. The primary value of peer review is to increase trust internally—that is, with regard to the reported finding.

This value of peer review seems straightforward, but its basis is often misunderstood. If we can clearly identify its justified rationale, then we will be better able to work out how to support and perhaps even improve its role in scientific practice.

The Peer Review Process

Today, academic papers are standardly sorted into those published in peer-reviewed journals and those published in non-peer-reviewed venues, such as chapters in an edited book. The latter actually do get reviewed by the book's editor, who in almost all cases is a peer expert, but they are not generally sent to external reviewers. That notion of external evaluation is the narrower notion that is taken to be the relevant standard. There are interesting institutional reasons for this current standard, which puts greater weight on the reviewers than was the case in the past, when the editor had the greater weight of responsibility in the decision to publish (or to fund, in the case of grants). We may reasonably differ about the degree to which editors or referees should bear the greater burden in the assessment, but here the point is that this is a matter of degree rather than kind, and that editors still retain considerable responsibilities even in the current American model.

Papers submitted to scientific journals are given a preliminary review by the journal editors. Some get a "desk rejection" when there are obviously major problems or when they don't fit within the journal's scope. Those that pass initial review typically are sent to two independent referees for review. Most journals ask reviewers to answer a set of specific evaluative questions and then to provide an overall recommendation of whether the paper should be rejected, accepted, or resubmitted after minor or major revisions. In cases where they disagree, an editor may sometimes be able to make a deciding judgment or may send the paper out for a third referee report. Grant-making agencies typically have even stricter standards, often requiring three to five referee reports. Submitters of course find the process stressful, and everyone has complained at some point about the dreaded "Reviewer #2" who is unsatisfied with anything. Occasionally, the process is too lax and a paper slips though that should have been rejected. However, for the most part the process works well.

Science set the model for peer review, and that is the focus here, but it is worth noting that scholarly publications in the humanities adopted similar review procedures. I was fortunate to get an early view of the peer-review process during my last couple of years of graduate school, when I served as the assistant to the editor-in-chief of *Philosophy of Science*, the flagship journal in its field. Its editorial process was mostly indistinguishable from those of scientific outlets. As a faculty member, I regularly serve as a referee for conferences, grant-making foundations, and journals both in science and the humanities. I submit papers in both as well and have always been grateful for comments and suggestions from referees, which may be a hassle to address, but invariably improve the work.

QUICK TAKE

The primary value of peer review is to increase trust internally with regard to a reported finding, but peer review can extend to increasing trust in policy decisions.

The process of peer review has evolved since its beginning as public demonstrations of experiments, but many of the core tenets and purposes remain the same.

Peer review does not guarantee the truth of a paper's results, nor is it the only way to report valid science. But its values hold true to reproducibility, objectivity, and humility to evidence.

John Englart/CC BY SA 2.0

During the 2017 March for Science, protesters turned up worldwide to advocate for scientific literacy, open communication, informed public policy, and stable investment in science and research. These protestors in Melbourne, Australia, emphasized the importance of evidence-based research and peer review before adoption. Such placards that played on traditional call-and-response demonstration slogans were common at various march sites around the globe.

Although these filtering mechanisms are no longer unique to science, that is where they have their roots, and it is worthwhile to dig down a bit to see how peer review developed to its present state.

The Roots of Peer Review

The way that science conducts peer review today is considerably different than in the past. The Royal Society of London, the first scientific society, would hold demonstrations as part of its meetings, in which members would re-create their experiments so that others could directly observe them themselves. *Philosophical Transactions of the Royal Society*, established in 1665, is the oldest scientific journal and represented a revolutionary advance for promulgating results to a wider community beyond any audience that could attend in person. *Transactions* was originally edited by Henry Oldenberg, the first secretary of the society, who conferred informally with other members about whether to accept an application to demonstrate or to publish. He sometimes solicited advice from experts who lived elsewhere, and this correspondence resides in the society's archives as the first historical records of written peer reviews.

In the mid-19th century, *Transactions* briefly experimented with a collaborative open review system that polymath William Whewell advocated for, but the journal soon returned to confidential assessments solicited to serve as expert advice for the editor. In recent years, in the service of historical interest, the Royal Society has begun to make some of these early records publicly available. After a reasonable embargo of 70 years, they recently released peer reviews dating from 1949 and 1954, adding to those already in the online archive. Two highlights include the report that Nobel Prize–winning chemist Dorothy Hodgkin wrote on the paper by Francis Crick and James Watson about the complementary structure of DNA (she recommended acceptance and made what she said is just a trivial suggestion to improve the photo of the model) and one by physicist Charles Galton Darwin on a paper by Alan Turing on the chemical basis of morphogenesis (he recommended various revisions). One may dig further back in time and find a report by his famous grandfather Charles Darwin on a submission about the origin of a geological feature called the parallel roads of Lochaber, in the Scottish Highlands (he noted a couple of ways that paper might have been improved, but strongly recommended publication), or by other eminent figures including Michael Faraday, William Whewell, and others.

These handwritten reports are time capsules that reveal changes in peer review. Just as the form of scientific report became standardized over time so that the formula includes clear statements of the evidence and conclusions drawn, so too did requests for referee reports become more formalized. *Transactions* began to send standardized questions in a printed form for referees to fill in. Eventually, it became common to put this information in a form that could be sent to the submitting author and to provide a mechanism for confidential advice that only the editor would see. Blind reviewing was another innovation that required submitting authors to provide an anonymized version of the paper to be sent to referees. Reviewing practices continue to evolve to this day.

Linchpin or Traffic Cop?

Nearly all discussions of peer review quote the influential view articulated by physicist John Ziman, who wrote in his 1968 book *Public Knowledge* that "the referee is the lynchpin about which the whole business of science is pivoted." Ziman saw the standardization of journal publications and the referee process as the central example of the thesis of his book that science should be understood socially as public knowledge that results from scientific consensus. "The fact is that the publication of scientific papers is by no means unconstrained," he wrote. "An article in a reputable journal does not merely represent the opinions of its author; it bears the *imprimatur* of scientific authenticity, as given to it by the editor and the referees he may have consulted." There is an ambiguity in Ziman's explanation of public knowledge in this context, signaled by the problematic notion of "imprimatur," a term that improperly suggests that the report acquires its worth by virtue of the approval of some official authorizing agent.

Later in his discussion, Ziman shifts metaphors but retains the idea of officiating agents when he discusses some of the specific aspects of the reviewer role, comparing referees to police officers on traffic duty, whose job is to keep traffic flowing in a smooth, orderly manner, by enforcing the general rules of the road. Reviewers, like traffic cops, should not be lax or kind, he explained, for after publication an article is unlikely to be given such a thorough check, but neither should they be moral censors; their goal should be to insist that new ideas—and any discovery worth publication is new, after all—be expressed as "accurately, clearly, and plausibly" as possible. No one would argue with these specific elements of good scientific driving, but we need to disambiguate different ways they might be understood.

Sociologists Harriet Zuckerman and Thomas Merton quote and interpret Ziman in their discussion of the advent of scientific journal publication, which they analyze in sociological terms. They focus on such features as its function for establishing priority of discovery to the researcher and, indirectly, to the nation, thereby resolving a tension between this reputational issue and what Merton had called the value of "communism" as an element of the scientific ethos, namely, the norm that scientific findings were not owned but were to be openly shared by the scientific community. Publication of findings in a journal broadened their reach, and evaluating material helped ensure their authenticity so that dissemination would not bring discredit to the society.

Such institutional factors are interesting but secondary to our topic. Our concern here are the more basic epistemic values that peer review serves; after all, reputation in science is not created whole cloth, but is based on making real discoveries. Recognizing these values allows one to judge the degree to which a practice succeeds or fails for science's central guiding purpose, providing a basis for assessment and potential improvements for institutional structures and other professional practices. To clear the way for a positive account, it will be useful to highlight three things that peer review is not.

Peer review is not the same as certification by an authority. Peer review does not guarantee that the reported results are true. Peer review is not what makes something actual science.

What Peer Review Is Not

First, peer review is not the same as certification by an authority. Such a notion of justification goes against the fundamental principle of science, which rejects appeal to authority. Rather, it is an assessment by another expert of a report of the evidence. Though often conflated, there is an important conceptual difference between authorities and experts. Both make judgments, but on different bases. The former is like a judge issuing a ruling in a trial, which has official status by virtue of the judge's legal authority to rule on matters of law. Facts are a different matter entirely, even in court. Expert reviewers can help assess evidential reports of experimental trials, but they are not themselves the evidence.

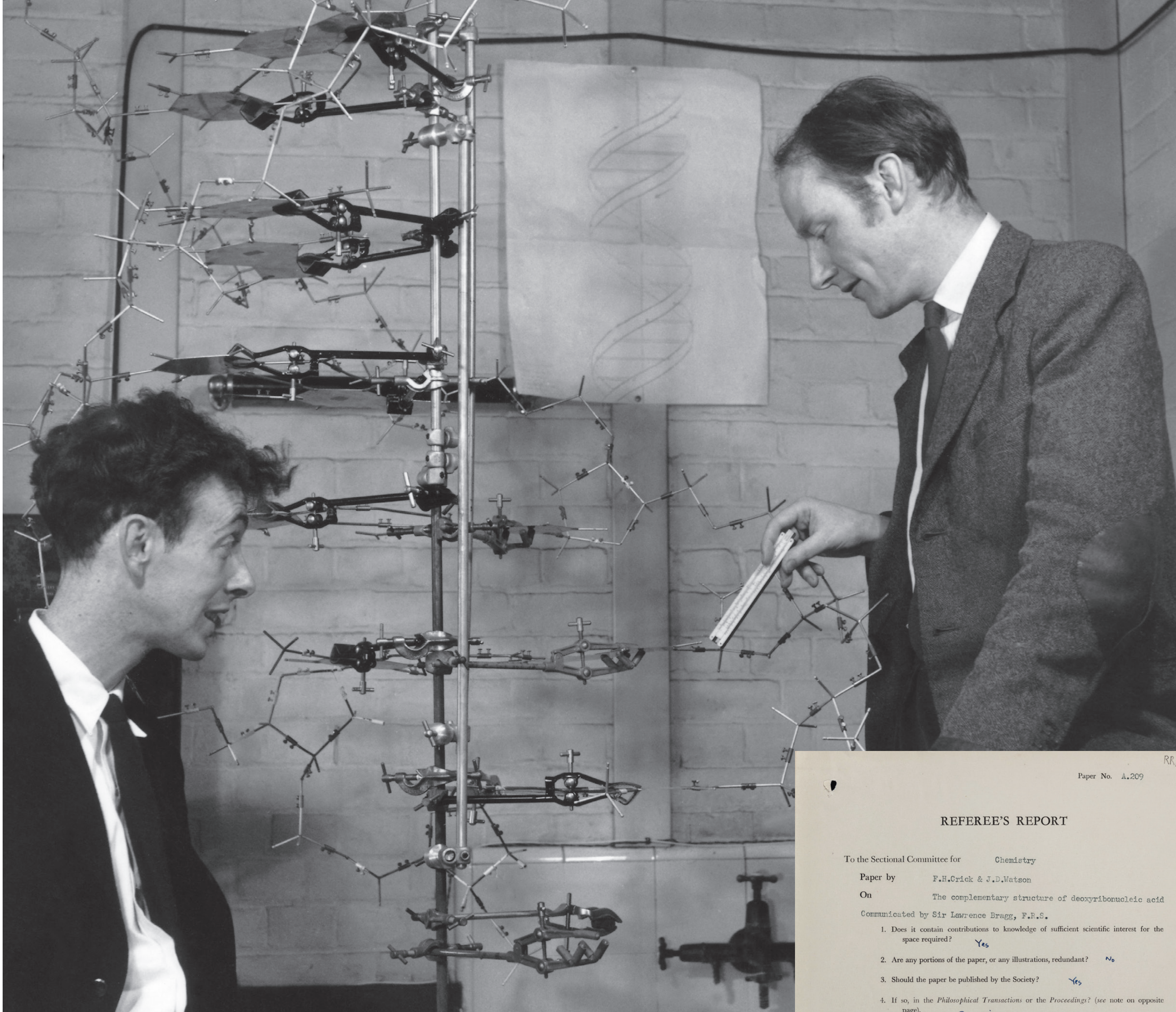
Second, peer review does not guarantee that the reported results are true. It is common to make this point by citing notorious cases of fraudulent papers that made it through review process. *Lancet* retracted the flawed 1997 Wakefield paper that linked vaccines to autism, but damage had already been done, and its fallout continues to this day. Referees usually see only a report of the research, not the research itself, so it is difficult to catch intentional fraud. However, a more basic reason for the point is that published papers are progress reports in an ongoing inductive process. Inductive evidence provides increasingly strong confirmatory support, but never guarantees of absolute certainty.

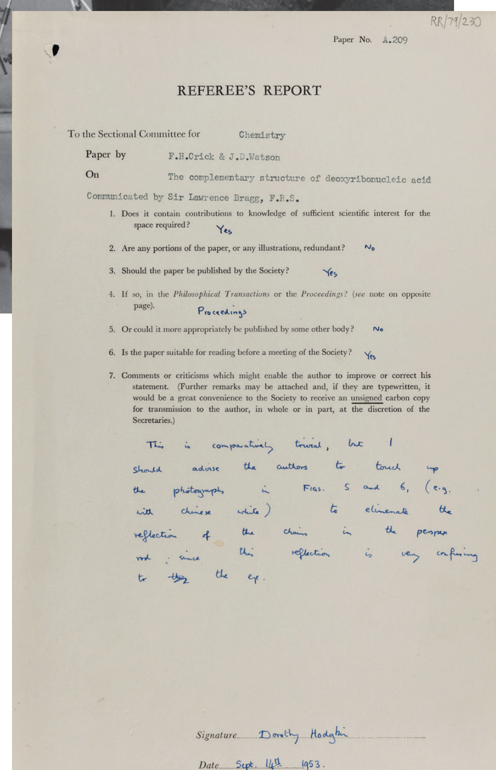
Third, peer review is not what makes something actual science. In historian Melinda Baldwin's excellent discussion of the history of the refereeing practices of the journal *Nature*, she notes how today it is common for observers to take peer review as the defining element for scientific acceptance. She mentions one journalist's article about a *Physics Letters B* publication on the detection of the Higgs boson as an example; the article headline read "CERN's Higgs boson discovery passes peer review, becomes actual science." Although this example is an amusing way to highlight the importance of peer review, it is misleading. *Nature* did not make external peer review a requirement until 1973, but science itself has of course been going on for centuries.

Getting clear about such misconceptions helps point us in the right direction for understanding the actual value of peer review.

An Extra Pair of Eyes

Published papers are reports of the state of the evidence. Peer review functions as an independent, indirect assessment of that evidence. Understanding this feature of peer review helps clarify Ziman's thesis. The relevant sense of science as public knowledge is not that it formed by community consensus; that gets the justification backward. Instead, scientific knowledge is public in the sense that anyone could in principle replicate the findings and make the same observations. That is why community consensus can be a good sign of good evidence. But, again, it is not a definitional condition for science. Indeed, it is quite possible to investigate a question scientifically and even to make a discovery by oneself. (As a thought experiment, think of the circumstance of the last person on Earth who investigates and discovers a cure for the virus that had zombified the rest of humanity. It would be the appropriate test results, not any peer reviewers of a report of the work, that made this very useful finding into actual science.) So, what is the role that referees perform?





RR/79/230

Paper No. A.209

REFEREE'S REPORT

To the Sectional Committee for Chemistry

Paper by F.H.Crick & J.D.Watson

On The complementary structure of deoxyribonucleic acid

Communicated by Sir Lawrence Bragg, F.R.S.

1. Does it contain contributions to knowledge of sufficient scientific interest for the space required? Yes
2. Are any portions of the paper, or any illustrations, redundant? No
3. Should the paper be published by the Society? Yes
4. If so, in the *Philosophical Transactions* or the *Proceedings*? (*see* note on opposite page). Proceedings
5. Or could it more appropriately be published by some other body? No
6. Is the paper suitable for reading before a meeting of the Society? Yes
7. Comments or criticisms which might enable the author to improve or correct his statement. (Further remarks may be attached and, if they are typewritten, it would be a great convenience to the Society to receive an unsigned carbon copy for transmission to the author, in whole or in part, at the discretion of the Secretaries.)

This is comparatively trivial, but I should advise the authors to touch up the photographs in Figs. 5 and 6, (e.g. with chinese white) to eliminate the reflection of the chains in the perspex rod; since this reflection is very confusing to ~~they~~ the eye.

Signature Dorothy Hodgkin

Date Sept. 14th 1953.

A. Barrington Brown / Science Source; Royal Society Library

The Royal Society of London recently released some historical documents into the public domain, including peer-review reports on famous scientific papers, such as this 1953 referee's report by Nobel Prize–winning chemist Dorothy Hodgkin on the submission by Francis Crick and James Watson about the complementary structure of DNA. Hodgkin recommends acceptance and makes what she says is a comparatively trivial suggestion to touch up the photo of the model, because the photo had a confusing reflection.

Think of referees as extra pairs of eyes. Reviewers function as peers in the sense of expert re-viewing of the evidence—they peer (view) as peers (equals) through the report with a knowledgeable eye, giving it a secondary check. This generic notion of refereeing is why it can work more or less the same way in the humanities without changing its own distinctive subject matter; humanists did not become scientists when their scholarly community adopted peer review.

As mentioned already, the original practice at the Royal Society was for researchers to demonstrate their findings in person so others could see the evidence for themselves. This ideal of direct observational replication remains the basic standard in science—it ought to be possible, at least in principle, for any other competent researcher to repeat the experiment and see the same thing. Practical constraints may limit realistic options for direct observers, but this form of reproducibility-in-principle is essential.

Reviewers are not usually in a position to see the experiment themselves in person, but they review the submission as experts who have the practical wisdom of experience—as experts who have observed such things before. Not the very thing, of course, because if it really is a discovery, then it is new to them in an existential sense, but at least the relevant methods as well as their limits and affordances. The judgment of expert peer reviewers is valuable because it links the relevant notions of reproducibility, objectivity, and humility to evidence. Reviewers' duty, as the name implies, is to double-check the report.

With this understanding of the basis of peer review, we may offer a broad checklist of some aspects of that duty.

Responsible Referees

First of all, check your expertise. Editors may have incorrectly identified

Alessandro Gottardo

The process of peer review requires a referee to ensure they are qualified in the area of the paper, then check the methods, math, and literature of the report. Peer review does not guarantee that a paper's results are true, nor does it make the results into "actual science." But expert peer review is valuable as extra pairs of eyes that hold to the ideals of reproducibility, objectivity, and humility to the evidence. Methods of research distribution that are open access but not refereed fit with the scientific ideal of transparency, but can raise dangers that can undermine trust.

you as having the relevant knowledge, so decline a request if the submission is outside your area. Let them know in your report if there is some part of the material that you didn't have the background to assess so they can be sure that another referee does. When Hodgkin was asked to review a paper by Rosalind Franklin on crystallite growth in carbons, she let the editors know that she felt "incompetent" to make a judgment and would defer to the expert view of another referee. Even Nobel Prize winners have their limits and forthrightly acknowledge them.

Next, check your biases. Editors try to pick referees who can assess a paper without improper bias for or against it, such as by being a personal colleague or known antagonist of an author. Blind review, the standard practice of anonymizing submissions, serves the same purpose. But specialists can sometimes identify who did the work even when blinded, so recuse yourself if you recognize you have such a relationship or some other conflict of interest. It may sometimes be appropriate to alert the editor of any other factor that might cloud your objective judgment. In his review of the parallel roads of Lochaber paper, Charles Darwin cautioned the editors that he had investigated the formation himself and so might overestimate the import of the research, but then gave reasons for why the subject's interest is deserved.

With these preliminaries out of the way, you can get to the primary task of the referee, which is checking the work itself. Check the methods; are the protocols designed appropriately to test the hypothesis? Check the math; were the statistics and other calculations done correctly? Check the literature; does the paper take into account and properly cite prior research, addressing known issues?

For all these points, make these assessments yourself. This statement should go without saying, but do not pass the job off to an AI tool, as some referees recently have been caught doing. Artificial intelligence is a useful tool that may aid your task, but it is no substitute for expert judgment. A peer reviewer submitting an AI review of a paper would be committing unethical conduct for the same reason it would be for an author who submitted an AI-written paper—it is an abrogation of responsibility.

Finally, keep in mind the import of what you are doing. Remember that in this work, your responsibility is not to the editor or to the journal; you are already providing free labor for them, after all (with the understanding that someone in your community will return that labor for you when it is your turn to publish). Your responsibility is to science itself.

Responsible Editors

Journal editors have an overlapping but somewhat different set of responsibilities, as they may also be employees of a journal, which can sometimes cause conflicts of interest: A journal's business model may not always be in alignment with its scientific goals. When journals were in-house publications of professional scientific societies, there was a confluence of interests. This arrangement shifted to a hybrid model as societies off-loaded more of the administrative burdens to publishing companies. No longer just in the role of printing services, publishers took on a business stake in the journal, which can be in tension with the journal's core purpose as an organ for scientific communication. Open-access journals exacerbate this conflict, as their business model is predicated upon payment for publication. This topic is relevant in thinking about the virtues needed to serve as an editor.

Mostly the same set of scientific virtues apply to both referees and editors, but the latter sometimes also need to cultivate the general virtue of courage in order to defend a journal's scientific integrity in the face of external pressures. These conflicts might come, for instance, in the form of political or social pressures to withhold or withdraw controversial findings. And although in most cases publishers maintain a hands-off policy, occasionally editors may need to stand up to directives from within, as in a recent case in which editors resigned in protest against a publisher's inappropriate intervention in journal autonomy.

Such issues may be becoming more salient as new technologies throw a wrench into the traditional business model of publishing companies. The affordances of the internet are stimulating exploration of new review and dissemination models. We can navigate these changes by paying attention to scientific values and virtues, which provide a framework for assessing and improving institutional and other social structures. Applying virtue theory to help sort out the current chaos in the world of scientific publishing deserves separate treatment, but it is worth briefly considering alternative models that are currently being explored.

Alternative Review Models

Reviewing practices have evolved and continue to do so. Different models come with their own advantages and disadvantages, so one should not expect a single, perfect approach. Two examples illustrate some of the considerations.

Review and publication in journals is not the only way to check and disseminate scientific research. The internet makes it easy to distribute material without the need for publishers, so it is possible to bypass journals and their reviewing process entirely. One recent approach has been to directly post preliminary reports (also called *preprints*) on an open forum such as *arXiv*, which was launched in 1991 and describes itself as an open-access "research sharing service." It does not peer review materials, but "curates" them using volunteer "moderators" who simply check for scholarly value and classify submissions into one of the site's subject categories. Circulating preprints within the research community has long been a common practice with a variety of benefits, and *arXiv* fulfills and improves many of them.

However, *arXiv*'s openness and scale adds risk; there is a big difference between mailing copies of typed drafts to colleagues and posting the same text on an open preprint server on the web. The former was mostly limited to a small circle of experts who understood the nature of preliminary reports, could judge their weaknesses, and might even offer suggestions for improvement before publication, whereas the latter is purposefully open to anyone, including nonscientists who are less able to assess their merit and for whom they may appear as equivalent to published reports. I once had a colleague alert me to an *arXiv* article from an industry AI group that had a significant flaw. This problem would likely be caught in regular peer review, but the paper had already received positive reporting in the news media with no recognition of the error and no mention that *arXiv* articles come with an asterisk. (Science journalism comes with its own set of norms, but now is not the time to get into a discussion of those.) Open access as a general idea fits with the principles of scientific transparency, but the term covers very different models of implementation, and those without—or with a lesser form of—review may raise dangers that can undermine rather than support trust. Perhaps a more prominent "nutritional label" on posted articles is warranted.

At the other end of the spectrum of new models are what have come to be called *registered reports*, which incorporate peer review in a two-stage process. Journals that use this model review a proposed study's methods and analysis before any data are collected, and either reject it then, or issue an "in principle acceptance." A second review is conducted after the results are written up, to confirm that the plan was followed. The results are published whether or not they confirmed the hypothesis. The basic aim of science is to get a truer understanding of the world, so disconfirmations of hypotheses can be as interesting and important as confirmations, and this model has the advantage of not improperly incentivizing only positive results. Registered reports were proposed as a means to reduce so-called "HARKing" (hypothesizing after results are known), which is said to raise the chance that findings are not reproducible or generalizable. Advocates argue that registered reports are needed for "confirmatory" research, which they differentiate from "exploratory" research. However, HARKing is not a general sin in the way it is often made out to be, and old evidence can often be valuable for confirmation, as well as exploration without sacrificing reproducibility. So, although the double-review model may be useful as a belt-and-suspenders precaution for certain kinds of research, the extra time and effort may not make it worth general adoption.

Open-access preprint servers as a general idea fit with the principles of scientific transparency, but those without review may raise dangers that can undermine rather than support trust.

These are just two of a variety of alternative or supplementary peer review models. Being clear about the purpose they serve can help us assess and improve this important practice.

Trust Me—I'm Peer Reviewed

In 2024, perhaps in recognition of peer-review week, the American Association for the Advancement of Science produced a celebratory T-shirt with *Trust Me* written in large letters and *I'm Peer-Reviewed* in a smaller font beneath it. It would be fun garb for another March for Science. Its geeky humor arises from the winking redeployment of the concept of peer review from papers to persons. There is of course a difference, but the scientific virtues of researchers and reports are interconnected, and peer review is an important point of contact.

As a re-viewer, you are performing a check with a second pair of eyes that are similarly focused on the ideals that make it more likely that science will achieve its purpose of discovering a truer understanding of the natural world. If, as a referee, you are trustworthy in applying the virtues of the researcher to the evaluation of research papers, you help increase not only their own trustworthiness but also that of the research community overall.

Bibliography

Baldwin, M. 2015. Credibility, peer review, and nature, 1945–1990. *Notes and Records* 69:337–352.

Pennock, R. T. 2017. Reaffirming the norms of science. Philosophy in the Contemporary World series. *Blog of the American Philosophical Association*. tiny.cc/f4dwzz

Royal Society. Science in the Making: Referee Reports. tiny.cc/i4dwzz

Ziman, J. M. 1968. *Public Knowledge*. Cambridge University Press.

Zuckerman, H., and R. Merton. 1971. Patterns of evaluation in science: Institutionalization, structure, and functions of the referee system. *Minerva* 9:66–100.

Robert T. Pennock is Sigma Xi Senior Fellow for Science and Engineering Values. He is a University Distinguished Professor at Michigan State University, where he is on the faculty of Lyman Briggs College, the departments of philosophy and computer science and engineering, and the ecology, evolution, and behavior program. His research involves empirical and philosophical questions that relate to evolutionary biology, cognitive science, and the scientific character virtues. Email: pennock5@msu.edu

People Are Not Peas

The decades out-of-date genetics curricula taught in most U.S. schools stokes misconceptions about race and human diversity.

Elaine Guevara

On a lengthy bus ride in the early 1970s, University of Chicago geneticist Richard Lewontin passed the time by doing some novel math.

Lewontin usually kept to the laboratory, studying proteins derived from ground-up fruit flies. Because DNA encodes proteins, this research addressed a fundamental question: How much do individuals of the same species vary genetically?

On the bus, Lewontin turned his attention to humans. Using available data, he computed how protein differences mapped across people around the globe. Contrary to what scientists assumed at the time, he found that most differences existed in every population—meaning the underlying genetic variation was shared across humanity, not sorted by geographic region or prevailing racial categories.

Lewontin published his calculations in a short paper that was included in the sixth volume of the book series *Evolutionary Biology* in 1972. He ended the paper with this definitive conclusion: "Since . . . racial classification is now seen to be of virtually no genetic or taxonomic significance either, no justification can be offered for its continuance." His results have been replicated time and again over the past 50 years, as datasets have ballooned from a handful of proteins to hundreds of thousands of human genomes.

But despite huge strides in genetics research—leaving no doubt about the validity of Lewontin's conclusions—genetics curricula taught in U.S. secondary and postsecondary schools still largely reflect a pre-1970s view.

This lag in curricula is more than a worry for those in academia. Increasingly, genomics plays a leading role in health care, criminal justice, and our sense of identity and connection to others. At the same time, scientific racism is on the rise, reaching more people than ever thanks to social media. Outdated education fails to dispel this disinformation.

From the basic genetics taught in K–12 schools to university courses, biology curricula desperately need an overhaul.

How DNA Differs

I am a biological anthropologist who uses genomic data to answer questions about primate and human evolution. When I began my doctoral studies a decade ago, we learned about Lewontin's paper for its historical significance, but his findings were old news.

Prior to his calculations, many scientists expected to find substantial genetic differences between people from different geographic regions or races. For example, Indigenous people in Africa would carry marker A, whereas Indigenous people in the Americas would have marker C.

Lewontin found a quite different result: The vast majority (more than 85 percent) of genetic differences existed among individuals from the same geographic region. This equates to some Indigenous people in Africa and some Indigenous people in the Americas carrying the DNA letter (molecular base) A, while other Africans and Indigenous people in the Americas carry the C. Most human genetic variation is shared across all the continents—or the racial groups invented during and since European colonial expansion.

Equivalent calculations done over the past two decades—based on genome-wide data from thousands of individuals—have reached the same conclusion: High genetic variation exists within geographic regions, and little variation distinguishes geographic regions.

Most common genetic variants—those carried by more than 5 percent of humans—appear across all continents. Only a small portion of these variants are exclusively found on one continent, and those continent-specific variants tend to be rare among members of a population, where they are found.

Genomic Insights

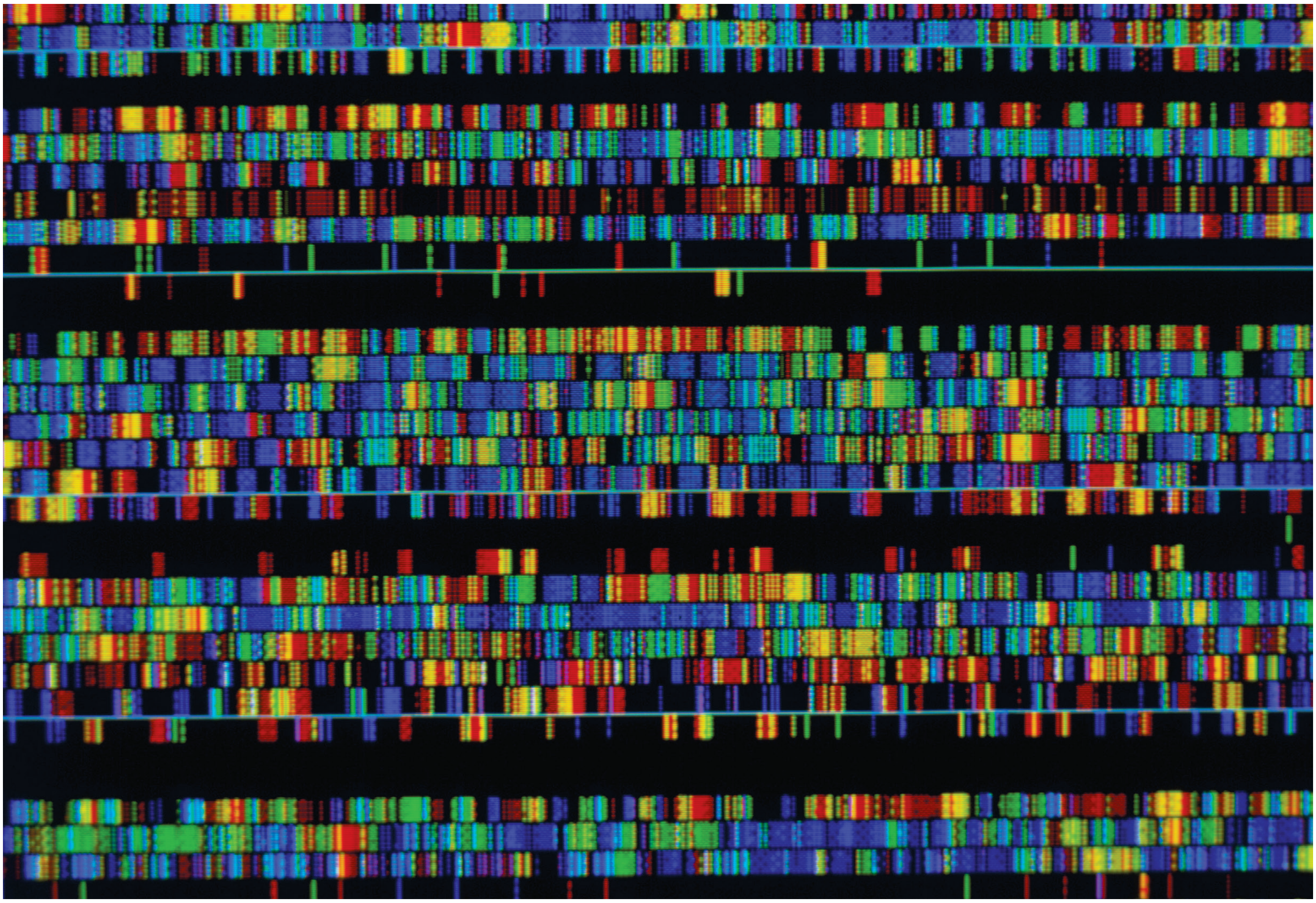
In addition to genomes from living humans, DNA extracted from ancient humans over the past two decades has revealed incredible insights. Across time, past humans frequently migrated and

QUICK TAKE

Researchers have known for decades that human genetics cannot be reduced to Mendelian inheritance, but many U.S. science curricula still include this outdated information.

Oversimplifying human genetics perpetuates the false idea that historical racial categories are inherently biological rather than social constructs.

Researchers have found that including in science curricula information about the global distribution of genetic variations effectively dispels misconceptions about racial differences.



James King-Holmes/Science Source

A human DNA sequence—represented as a series of colored bands on this computer screen at the Wellcome Sanger Institute in Cambridge, England—is incredibly complex. Each color represents a specific nucleotide base, which together compose the genetic code for an individual. In many primary and secondary school classrooms, this complex system is simplified into outdated genetics lessons that perpetuate false, and even racist, ideas about human differences.

mated with or displaced people they encountered in other regions, resulting in a tangled tree of human ancestry. The ancient DNA results refute any notion of deep, separate roots for humans in different geographic regions.

Also, contemporary researchers better understand how DNA variation contributes to differences in human traits. Scientists now know that most of our biological attributes are influenced by many genetic variants, and their effects vary in response to assorted environmental factors. For example, thousands of genetic variants influence height, and their effect is modified by childhood nutrition and infections.

As for race, researchers have shown conclusively that historical racial categories are not based in any inherent aspect of our biology. But that doesn't mean these racial categories and biology don't affect people's lived experiences.

As laid out by a major professional association for biological anthropologists, race is a social reality that affects our biology. For the past several hundred years in the United States and other colonized lands, racism has influenced people's access to nutritious food, education, economic opportunities, health care, safety, and more. As a consequence, and precisely because of the environmental influence on most traits, the social construction of race is a risk factor for many health conditions and outcomes, including maternal and infant mortality, asthma, and COVID-19 severity.

Physicians and researchers are increasingly recognizing that racial health disparities are not innate racial differences but are instead a result of racism.

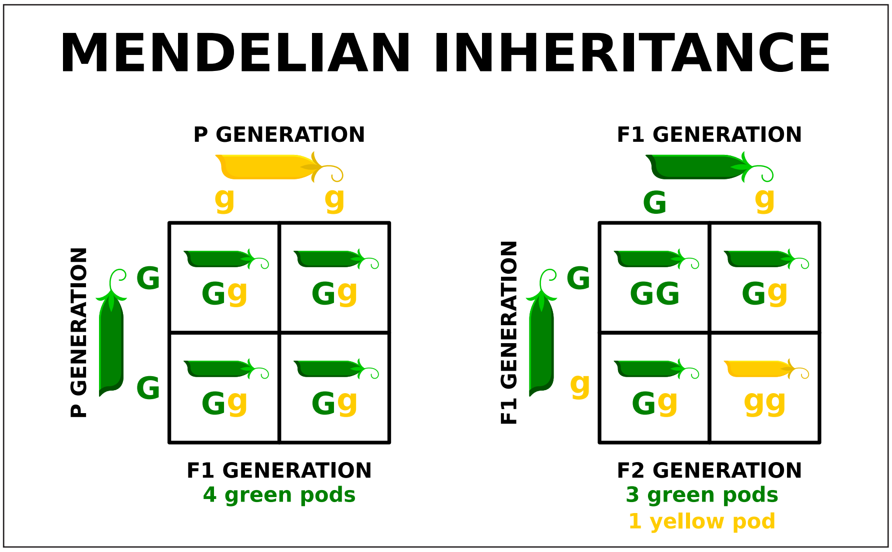
Lagging Lessons

When I started teaching at Duke University five years ago, I assumed most college students would have received a basic genetics education—one that reflected fundamental updates in genetics research over the past 50 years.

Not so. I quickly learned most undergraduates in my classes still hold the pre-Lewontin belief that human genetic variation predominately sorts geographically. Many students also thought race was based in genetic differences and that single mutations could explain complex traits in humans, such as risk for most diseases.

I doubt the students in my classes were unique. Studies have shown inconsistent and ahistorical presentations of genetics likely contribute to students' confusion about the nature of genes and their role in our lives.

Standard U.S. high school textbooks give little attention to human biological variation. Instead, most books focus on topics such as Gregor Mendel, the 19th-century Austrian priest who derived "laws" of inheritance from tracing observable traits when crossing pea plant varieties. (Remember those Punnett squares with green and yellow peas, or wrinkly and round ones?)



Tasko / Alamy Stock Vector

U.S. biology textbooks often begin their unit on genetics with the story of Gregor Mendel, the 19th-century Austrian monk who experimented with breeding peas. Simple tools such as Punnett squares are used to demonstrate how heritable traits, such as a pea's color, pass on through generations of the plant. This idea, and the Punnett square as a tool, is then extended to human characteristics such as eye color. However, domesticated species have little relevance to human genetics, and these tools give the false impression that people, like peas, can be categorized into discrete types.

I, along with others, am concerned that this focus instills and reinforces a false pre-Lewontin view that humans, like Mendel's peas, come in discrete types. In reality, early studies of peas and other inbred, domesticated species have little relevance for human genetics.

When U.S. high school, college, and medical school classes do cover human diversity, the lessons focus primarily on disease prevalence—and abound with racialized terminology. For example, students often learn that sickle cell anemia primarily affects African Americans, but sickle cell anemia is neither unique to nor characteristic of people with African ancestry. Rather, the genetic variant that causes sickle cells occurs more frequently in people with recent ancestry in parts of Africa, Europe, and South Asia—regions where malaria is or recently was endemic.

This distinction may seem like splitting hairs. But it turns out such distinctions are consequential.

Scholars such as biologist and educator Brian Donovan have tested how these simplified examples influence students' thinking. In multiple studies, he compared classrooms using standard textbooks with those incorporating more updated and accurate content on human biological variation. Students who received the typical—outdated—genetics education were more likely to think race is inherently biological and that genetic differences among races explain differences in life outcomes. The dated material also decreased students' support for efforts meant to redress racial inequity.

Students who received the typical—outdated—genetics education were more likely to think race is inherently biological and that genetic differences explain racial inequity.

On the flip side, that research also showed these measures are reversed by content that includes the global distribution of most genetic variation and the complex, multifactored basis of most human traits.

Educators can either perpetuate or dispel misconceptions, depending on how they teach genetics.

Slaying a Zombie Idea

I consider the notion that historical racial categories are based in biology to be a *zombie idea*, an idea that perpetually reanimates despite repeated empirical falsification. Zombie ideas of biological race are most likely to persist when deeply held views, particularly those important to our social identities, undermine the rational appraisal of evidence. As a result, some have argued that it is futile to combat racism with scientific evidence.

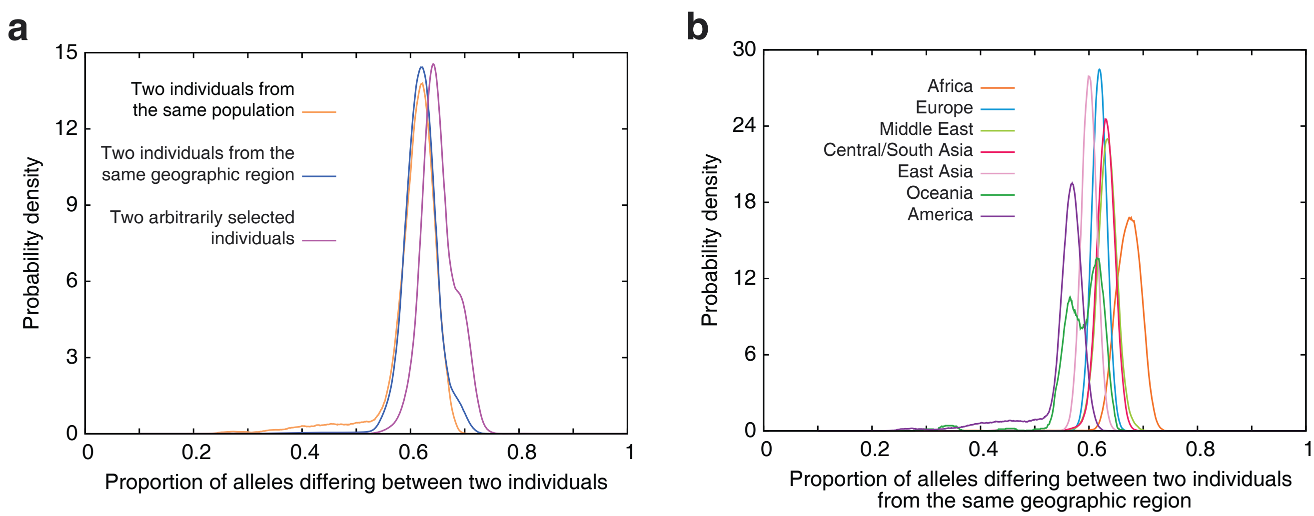
Direct-to-consumer genetic tests, such as those offered by 23andMe and AncestryDNA, can reinforce misconceptions about human variation, thereby helping the zombie idea persist. These services have become many people's primary reference point for human genetics information. To be marketable, the companies must communicate their results in simple, familiar ways that also appear meaningful and reliable. This approach usually entails simplifying genetic ancestry to bright, high-contrast colors, pinned definitively to geographic regions.

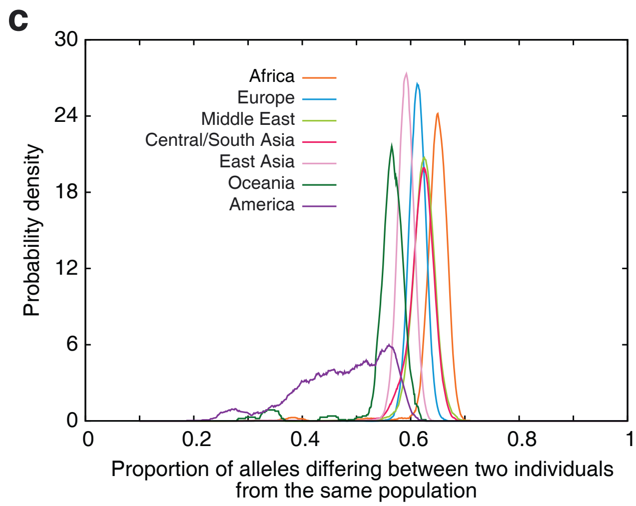
Even so, the research by Donovan and others suggests it's possible to weaken this zombie: Reaching young students via biology curricula can alter their views on race and human variation.

However, few secondary and undergraduate textbooks offer updated content. Pea plant genetics still fill the pages. Adopting new curricula, which complicate material already challenging to teach, is daunting. Implementing more accurate high school genetics curricula will require support from school administrators, parents, and entities such as the College Board, which administers the Advanced Placement biology exam.

In the meantime, widespread integration of modern genetics into college and university courses is essential. Higher education does not have the same reach as middle and high school, but college instructors have more agility in adjusting their course content. Plus, instilling up-to-date understanding in future secondary teachers and physicians can have ripple effects.

These changes aren't easy, but they are possible and worthwhile. In addition to thwarting the spread of racist worldviews, the next generation will be better informed about tricky health care and reproductive decisions. Revised curricula that do not implicitly promote a biological basis for historical racial categories are also less likely to alienate students from underrepresented groups. This change could in turn increase diversity in the scientific





In 1972 University of Chicago geneticist Richard Lewontin proved mathematically that most genetic variation is shared across all humans regardless of race or geography. An individual is statistically more likely to have greater genetic similarity to a randomly selected person on a different continent than to their neighbor (*a*). High genetic variation is common within a region (*b*) and within a population (*c*), and it is especially prevalent in Africa, where there is the greatest human genetic diversity. Nonetheless, more than 50 years after Lewontin published his findings, many U.S. students learn outdated lessons that reinforce false ideas about genetic differences between races.

N. A. Rosenberg, 2011.

workforce, leading to better, healthier science and greater trust between researchers and the public.

Lewontin died at age 92 in 2021. His work was instrumental in demonstrating that race is not based on genetic differences. Many others, such as geneticists and gifted communicators Joseph L. Graves Jr. at North Carolina Agricultural and Technical State University, Charmaine DM Royal at Duke University, and Graham Coop at the University of California, Davis, have tirelessly continued to carry this torch.

Educators and families can help by demanding their schools replace curricula focused on 19th-century peas with 21st-century human genetics.

References

Biddanda, A., D. P. Rice, and J. Novembre. 2020. A variant-centric perspective on geographic patterns of human allele frequency variation. *eLife* 9:e60107.

Carvalho dos Santos, V., L. M. Joaquim, and C. N. El-Hani. 2012. Hybrid deterministic views about genes in biology textbooks: A key problem in genetics teaching. *Science & Education* 21:543–578.

Donovan, B. M. 2017. Learned inequality: Racial labels in the biology curriculum can affect the development of racial prejudice. *Journal of Research in Science Teaching* 54:379–411.

Donovan, B. M., et al. 2024. Human genomics education can reduce racism. *Science* 383:818–822.

Graves, J. L. 2015. Great is their sin: Biological determinism in the age of genomics. *The Annals of the American Academy of Political and Social Science* 661:24–50.

Lewis, A. C. F., et al. 2022. Getting genetic ancestry right for science and society. *Science* 376:250–252.

Lewontin, R. C. 1972. The apportionment of human diversity. In *Evolutionary Biology*, vol. 6, eds. T. Dobzhansky, M. K. Hecht, and W. C. Steere, pp. 381–398. Springer.

Piel, F. B., et al. 2010. Global distribution of the sickle cell gene and geographical confirmation of the malaria hypothesis. *Nature Communications* 1:104.

Redfield, R. J. 2012. "Why do we have to learn this stuff?"—A new genetics for 21st century students. *PLoS Biology* 10:e1001356.

Rosenberg, N. A. 2011. A population-genetic perspective on the similarities and differences among worldwide human populations. *Human Biology* 83:659–684.

Sella, G., and N. H. Barton. 2019. Thinking about the evolution of complex traits in the era of genome-wide association studies. *Annual Review of Genomics and Human Genetics* 20:461–493.

Sparks, R. A., K. E. Baldwin, and R. Darner. 2020. Using culturally relevant pedagogy to reconsider the genetics canon. *Journal of Microbiology & Biology Education* 21:21.1.30.

Stoneking, M., L. Arias, D. Liu, S. Oliveira, I. Pugach, and J. J. R. B. Rodriguez. 2023. Genomic perspectives on human dispersals during the Holocene. *Proceedings of the National Academy of Sciences of the U.S.A.* 120:e2209475119.

Winston, A. S. 2020. Why mainstream research will not end scientific racism in psychology. *Theory & Psychology* 30:425–430.

Elaine Guevara is a biological anthropologist who studies the evolution of life history and aging in primates. She received her Ph.D. from Yale University and is a lecturer at Duke University. This article was originally published on Sapiens *(sapiens.org). Email: elaine.gomez.guevara@duke.edu*

The Science of Hi-Fi Audio

Despite great advances in quantifying sound quality, engineers are still struggling to satisfy the subjective ways listeners respond to music.

John G. Beerends and Richard Van Everdingen

The swell of the orchestra reaches a crescendo, all of the instruments together creating a swirling field of sound that fills the concert hall and surrounds the listener. Anyone who has ever attended a classical music concert has probably encountered that joyous feeling of being completely immersed in sound. But most of us don't have an orchestra at home, and a large orchestra probably would not fit in there, anyway.

Engineers have been seeking ways to re-create the immersive experience of a live music performance ever since 1877, when Thomas Edison made the first crude recording of himself reciting "Mary Had a Little Lamb." The ultimate goal has been high-fidelity audio, or hi-fi: the reproduction of sound without audible noise and distortion, based on a flat frequency response within the human hearing range. In terms of technology, that goal might now seem easily attainable. Even moderately priced consumer equipment can process sound accurately; given that humans only have two ears, a simple stereo setup with two speakers would seem sufficient for the job. Yet modern designers of hi-fi audio systems keep adding more speakers with more audio channels without ever quite managing to recapture the sensation of musical immersion.

We have spent our careers pursuing a scientific, perception-based approach for assessing audio devices, so we are keenly aware of the obstacles to attaining hi-fi sound. Above all, every person has different ears, a different brain, and unique, personal preferences. It is therefore difficult to separate facts from opinions and fake claims when discussing the quality of recording and playback.

One of us (Beerends) memorably experienced the subjectivity of sound while attending a hi-fi trade show, where a small company demonstrated a very expensive audiophile amplifier. During that demo, a soft hum was audible to me in the silent intervals of the music. At first, the man running the equipment could not perceive the hum. Only after I suggested that he listen to the loudspeakers at a closer distance could he, too, perceive the hum. Nobody appreciates a humming amplifier, so presumably multiple engineers at the company failed to notice the sound that was obvious to me.

For recordings of speech, at least, test subjects largely tend to agree in their assessments of reproduction quality, especially when they are listening to familiar voices. But for music, individual preferences tend to dominate, greatly complicating the situation. Whether people are listening through headphones, earbuds, Bluetooth speakers, home stereo, automotive audio, or any audio system you can dream of, their judgements of musical sound quality show large differences from individual to individual.

The upshot is that audio engineers can achieve high quality rather easily for the recording and playback of speech, but the recording and playback of hi-fi music remains elusive. Even multi-channel systems cannot consistently and satisfactorily re-create most listeners' experiences of, say, the rich, diffusive sound of a large classical orchestra. In fact, such complex audio setups miss the most important subjective aspect of listening to music: being immersed in the sound. We argue that there is a better and simpler solution.

A Search for Transparency

An essential quality of hi-fi audio is what's called *transparency*. For a well-designed audio device—regardless of whether it is for music recording, compression, storage, streaming, or playback—there should be no discernible difference between the input and the output, as if the device itself were transparent and invisible. Using that audiophile amplifier as an example, we could take a sample of the input signal and compare it with a sample of the output signal. If we then subtract the output from the input (after aligning the amplitude and compensating for a possible delay), we should get an overall zero signal.

If the subtracted signal is not exactly zero, the difference between the input and output might still be so small that it is not audible, making the device transparent from a perceptual point of view. But if the device is not perceptually transparent, we then want to have an interpretation algorithm that can quantify the extent to which the system falls short of the transparency

QUICK TAKE

The goal of high-fidelity audio is to capture the feeling of a live musical event. Doing so requires more than just reproducing sound accurately, without audible distortion or noise.

Perceptual measurement techniques provide an effective way to evaluate sound quality for speech. But the techniques cannot fully capture subjective impressions of music.

A sense of immersion is crucial for a satisfying musical experience. Most commercial systems fail in that regard; the authors propose a new solution, using both direct and diffuse sound.

Keith Jefferies/Stockimo/Alamy Stock Photo

A live musical experience depends on many factors. The instruments and the acoustics of the performance space affect the sounds that reach the listener. But the ways that listeners respond also depend heavily on each individual's unique characteristics, both physiological and psychological. A satisfying hi-fi system should do more than reproduce sounds accurately; it should also re-create the feeling of immersion produced by a live event.

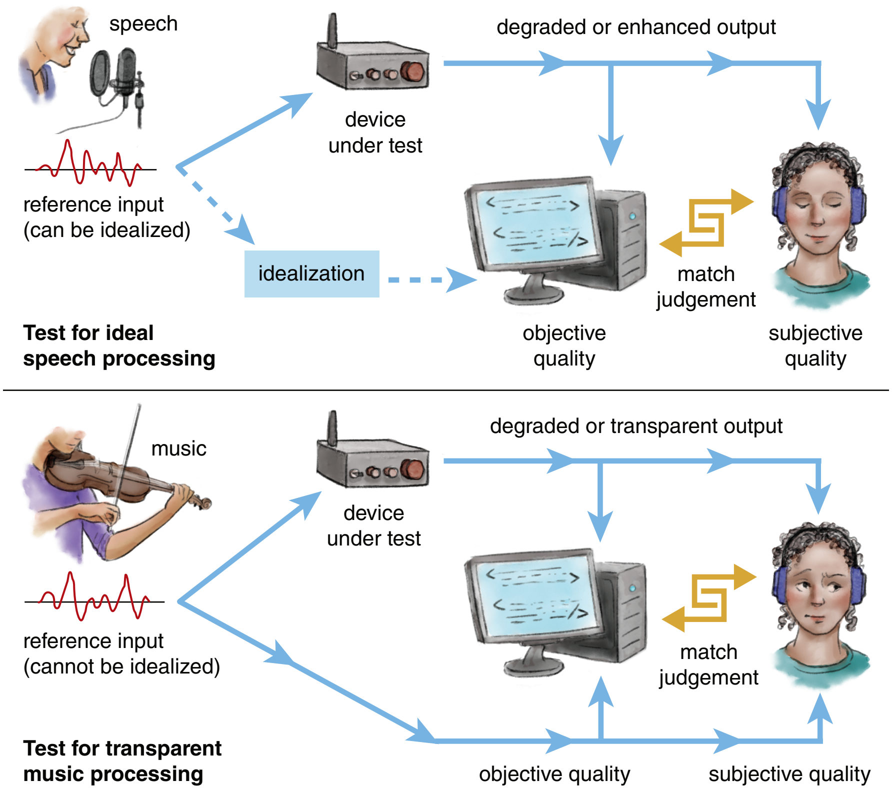
ideal. Following this approach, audio engineers have designed perceptual measurement systems that assess audible degradations of perceived audio quality (*see illustration on page 34*).

An effective perceptual measurement method was developed in the early 1990s by one of us (Beerends) in collaboration with Jan Stemerdink at KPN Research NL, the research arm of the biggest Dutch telecom company. The initial version of this method, called Perceptual Speech Quality Measure or PSQM, could assess the software used to code and decode *narrowband* speech, the kind commonly used for telephone communications; PSQM demonstrated high correlations between subjective evaluations and objective measurements of speech quality. In 1996, the International Telecommunication Union (ITU) endorsed PSQM as a worldwide standard ("Recommendation P.861 PSQM"). An improved version of PSQM, which also allowed for the assessment of wideband speech (used for high-definition communication), was developed in 2001 by KPN Research and British Telecom and accepted by the ITU as "Recommendation P.862 PESQ."

In 1998 the ITU adopted a similar perceptual measurement technique for assessing the quality of music encoded using common digital formats, such as MP3, AAC, WMA, and OGG ("Recommendation BS.1387 PEAQ"). However, assessing the quality of coding-decoding systems, or *codecs*, is far more difficult when dealing with music than it is with speech—especially assessing how much the sound quality has been degraded when the codecs behave non-transparently.

Listeners have more widely divergent opinions on the effect of degradations on music than they do on speech. Furthermore, the varied ways that people perceive and process sound (due to both innate physiological differences and subjective, psychological ones) are far more important when listening to music than they are when listening to speech. Even simple differences in *perceptual threshold*, the level at which certain frequencies become audible, can lead to large differences in listeners' quality assessments. In particular, degradation that occurs at high frequencies, above roughly 8 kilohertz, has limited impact on how people perceive speech but can have a large impact in the way they perceive music. Because of these complicating factors, perception-based measurements of audio quality show significantly poorer correlations with subjective evaluations when the experiments use music rather than speech.

A fundamental obstacle to developing a more accurate objective perceptual quality assessment method is that typical listeners, who simply want to enjoy their audio system, generally do not have access to an ideal reference signal. Instead, they judge the sound quality of their system against their own subjective, internal ideal.



Stephanie Freese

Perceptual measurement techniques are used to assess devices that code, decode, store, or transmit sounds. A reference input signal is fed into the device being tested, such as an audio amplifier. The reference signal and the output signal from the device are then played back for listeners, who evaluate the subjective quality of the resulting sounds. Objective computer models attempt to simulate how the listeners will respond. For speech (*top*), engineers can construct an ideal that allows them to assess the quality even if the speech has been enhanced, such as if some noise has been removed. For music (*bottom*), such idealization is not possible, and quality can be assessed only in terms of audio transparency.

In principle, if we had access to a listener's ideal sound, we could design a processing method that delivers a personalized perfect audio quality. For speech, we can do something quite close to that, because test subjects largely agree about how ideal speech should sound. Such consensus means that it's possible to create a perceptual measurement technique to assess the end-to-end quality of any voice connection, such as a video meeting or a cell-phone call. One of us (Beerends) was the main developer of yet another speech-quality standard known as P.863 POLQA, adopted by the ITU, which compares such connections against an average, ideal speech representation derived from a large database of speech-quality assessments. No such standard exists for music.

Another obstacle to objectively assessing the quality of music processing is that our ears hear sounds, not digital signals. Subjective audio tests therefore require a transduction device—headphones, a loudspeaker, or set of loudspeakers—to assess an audio signal. The device that we use has to be of superior quality for a listener to hear small degradations in the audio output, especially if we are evaluating high-quality devices that are designed to come close to perceptual transparency. When we are testing such devices, we will allow subjects to directly compare the reference input with the output, making it easier for them to detect small degradations in the output signal. For instance, we might let them hear what the audio sounds like before and after it passes through an amplifier or through a Bluetooth streaming system.

The situation becomes trickier still if we want to assess the quality of headphones and loudspeakers using perceptual modeling, because we run into the problem that the output is an acoustic wave that we need to transform back to data that we can feed into a perceptual measurement model. Accurately recording the output of a loudspeaker or a headphone is difficult, and it can be carried out in a variety of ways that each lead to different assessments of the device under test.

This is the core problem in the science of hi-fi audio quality assessment: Subjective tests of microphones, headphones and loudspeakers are all based on judgments that use an unknown internal ideal. Developing an objective perceptual measurement of listeners' subjective and diverse responses is exceedingly difficult.

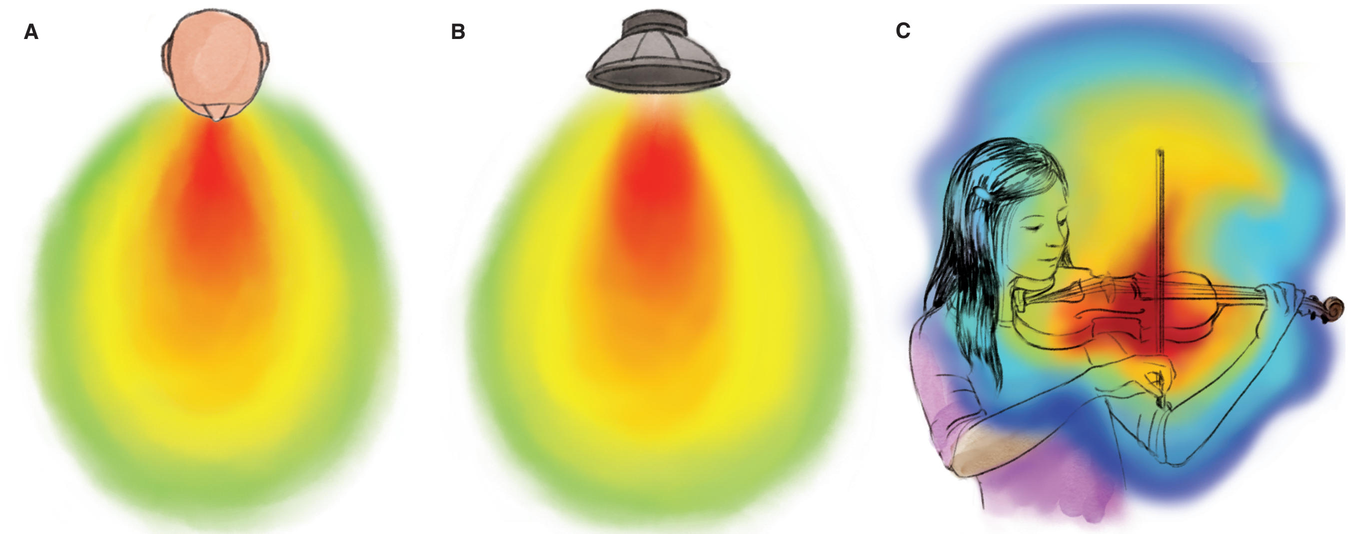
From Recording to Playback

What we really want to do is create objective perceptual measurements that can assess the complete life of a piece of music from recording to playback. That process includes everything from transduction, in which recording microphones convert sound into electronic signals, to reproduction, in which headphones or loudspeakers convert the final versions of those signals back into sound that the listener can hear.

At this point, the acoustic environments in which the recording and the playback occur become important. When you listen to a recorded sound, the room where the recording was made has a significant effect on the audio quality. Listen, for instance, to a voice recorded in a bathroom and you will hear that acoustic reflections from the room dominate the audio quality. The way we reproduce the recording also has a significant impact on the audio quality.

Audio engineers often use reference-standard headphones when asking test subjects to make audio quality judgments. Unfortunately, headphones produce an unnatural auditory effect: They make it seem as if sound is localized in the center of your head, whereas in real life the sound will be localized at some external source. When you move your head, your perception of that source will change; when you move your head while wearing headphones, everything stays the same. To make headphone playback more realistic, we therefore add a set of personalized corrections called *head-related transfer functions*. With the proper corrections applied, the sound localization will seem to move along with the listener's head movements.

Listening to audio playback over loudspeakers presents its own challenges, because the setup of the reproduction room has a significant effect



Stephanie Freese

Simple, directional sounds are relatively easy to reproduce accurately. A human voice (*a*) and an individual loudspeaker (*b*) have similar, directional properties and so produce similar sound fields. We also normally listen to spoken voices one at a time. In contrast, a single musical instrument such as a violin (*c*) produces a sound field with wildly varying directional properties. Combining multiple instruments makes the situation even more complex.

on the perceived audio quality. The advantage of the loudspeaker approach is that the room degrades the playback in the same way that it would have degraded the live source in that room. We can therefore make a monophonic recording of an acoustic source in an anechoic room (which prevents sound from reflecting) and play it back through a single loudspeaker with the same directional properties as the source, such that there is a transparent relationship between recording and playback. In contrast with the headphone experience, there is no need for a head-related transfer function correction. You could go into that listening room, rotate your head in any direction, and move around freely while maintaining full transparency between the recording and the original live sound.

The drawback of using a loudspeaker for playback is that it requires that the loudspeaker's directional properties match those of the original acoustic source. For a single voice, made by one person and coming from one direction, we can easily do that. But if we try to make a loudspeaker match up with the sound field radiating from a musical instrument, we run into trouble (*see figure above*).

What we really want to do is create objective perceptual measurements that can assess the complete life of a piece of music, from recording to playback.

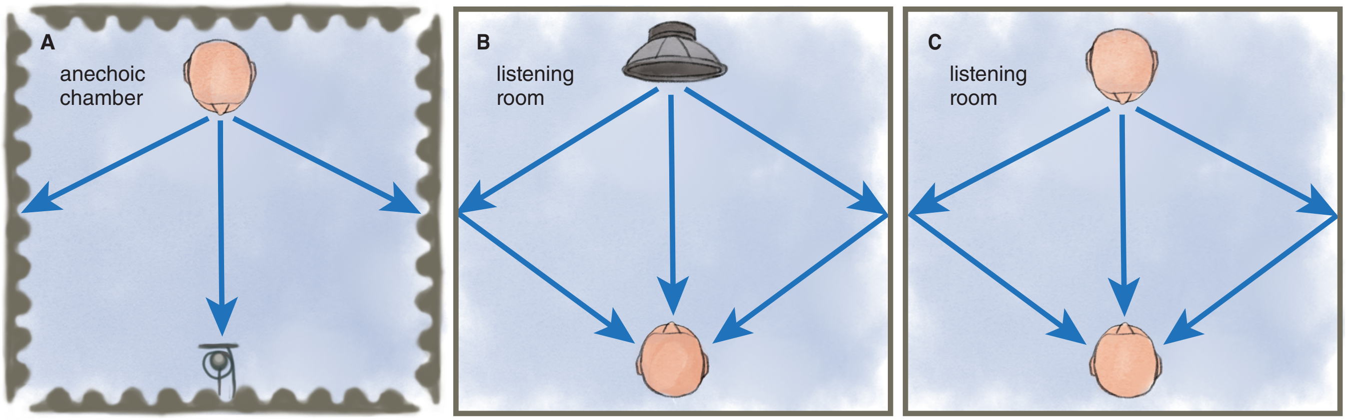
Musical instruments can have complex directivity patterns, with some frequencies more likely to reach the listener directly but others more likely to arrive via reflection, so recording them in an anechoic room will result in an unbalanced sound. Many modern recordings use electronic instruments that lack a natural reverberation, which introduces another issue. Audio engineers often add artificial reverberation to electronic instruments and to recordings made in sound-dampened studios; such reverb will also become imbalanced when applied to an anechoic room recording.

The situation becomes even more complicated if we apply a "dry" recording approach, with no added reflection or reverb, to a performance with multiple acoustic sources, such as an orchestra. To reproduce those sound locations, we would need a large (possibly very large) number of anechoic mono recordings played back over at least the same number of correctly placed loudspeakers. It's a rather impractical approach for a large orchestra that cannot be contained within a recording studio or a living room.

For recording live music, we strive to capture an immersive feeling similar to the experience of the original event. Ideally, the acoustics of the recording room would provide proper acoustic integration of all the instruments, including their directional patterns. In the room where we play back the recording, we want to reproduce the sound field as it would have been experienced live, taking into account the crucial feeling of immersion.

Re-creating the Immersive Experience

We now run into a dilemma, because we have arrived at two distinctly different approaches to the recording and playback of hi-fi sound. One is focused on transparency in the "here and now," optimizing the sound from a single, simple directional source. The other is focused on transparency in the "there and then," attempting to re-create the experience of a complex, multi-source, diffuse live event. The two approaches require completely different, incompatible recording and playback techniques.



Stephanie Freese

Recording and playback of a spoken human voice can be carried out effectively in an anechoic recording room (*a*), where the sound-damped walls mean that the microphone picks up only the direct sounds. On playback, a loudspeaker (*b*) that re-creates the voice produces the same direct and reflected sounds as does a human speaker (*c*) at the same location; in audio terms, there is a transparent relationship between recording and playback.

If we are aiming for the illusion of "there and then," we need to figure out the minimum number of audio channels required for hi-fi quality loudspeaker reproduction. We've known for a long time that one is not enough. The invention of stereophonic sound by British electronics engineer Alan Blumlein in the 1930s significantly improved the perceived loudspeaker reproduction quality of music events compared with mono. In stereo recordings, we can use time and intensity differences between the two channels to allow the listener to hear different musical instruments in different locations.

The reproduction of music is seldom improved by adding more playback channels beyond the typical two.

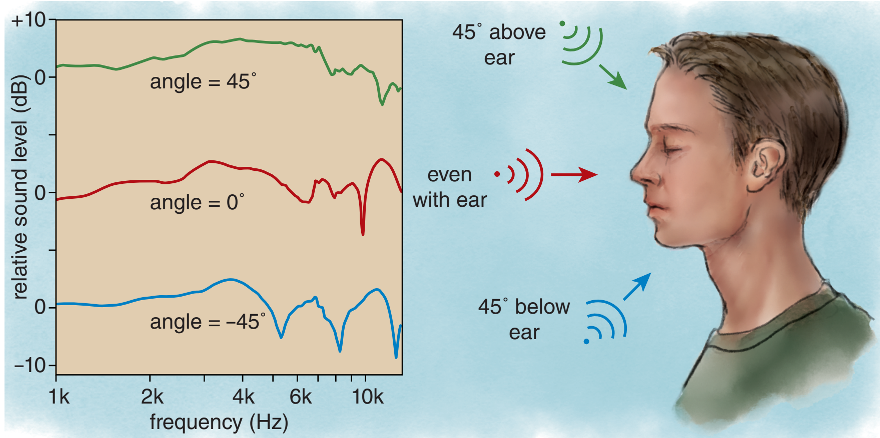
For headphone reproduction, two channels are sufficient, although they require meticulous, personalized head-related transfer function corrections. For loudspeaker reproduction, the sound quality is determined by a number of characteristics, of which the acoustics of the listening room is a dominating factor. For a high-quality audio experience, the acoustic resonances in the listening room should be damped and we should aim for a low reverberation time, preferably less than 0.5 seconds, allowing the listener to hear the longer echoes that were captured in the original recording environment. On the other hand, if the reverberation time of the listening room is too low, such as in an anechoic room, people lose the feeling of being immersed by the sound. To compensate for that effect, an extremely large number of audio channels would be required.

In a typical listening location, such as a living room, the number of audio channels needed for hi-fi quality loudspeaker reproduction of music events is not clear. Although expanding the number of recording-playback channels from one to two (from mono to stereo, that is) was a great improvement, extending that principle to four-channel "quadraphonic" sound was a commercial failure in the 1970s. The likely reason for the lack of public acceptance is that musical events seldom require localization behind the listener. In a concert hall, you seldom hear musical instruments behind you; the immersive experience of a concert performance is influenced instead by the more subtle, diffuse sound field that reaches your ears from all directions. To replicate that experience, a multi-channel system should reproduce only the diffuse field over the back channels.

The recent development of elaborate home theater surround systems with more than a dozen channels seems inconsistent with the characteristics that improve music reproduction. Commercial systems such as Dolby Atmos and DTS-X are useful mainly for watching films and playing games, media in which the sound effects require a more exact localization. The reproduction of music is seldom improved by adding more playback channels beyond the typical two. While the number of audio channels has been growing in home theater systems and high-end audio systems in vehicles, recording of music has remained mainly in stereo. In general, multi-channel systems introduce complexity in the setup and often introduce sound-localization errors that diverge from the live experience.

For music, the feeling of being immersed in a natural-sounding diffuse field is much more important than an improved sense of localization. Adding more reproduction channels can even lead to undesirable, uncontrolled degradations that people describe as "hearing things jumping around." Creating a high-quality, immersive diffuse field turns out to be quite difficult, however. Engineers have developed many complex algorithms for achieving such immersion, often using four speakers possibly with an added center one. But such so-called *two-to-five up mixing algorithms*, which extend stereo reproduction to five channels, tend to provide a poorer front sound-image quality along with only a marginal improvement in immersion. In most cases, listeners report that they prefer the original stereo reproductions, even though stereo audio cannot fully capture the feeling of immersion from a live music event.

The major reason why immersion is so difficult to attain is that it is a highly cognitive concept, one that was only recently introduced in the world of sound reproduction over loudspeakers. The feeling of being immersed is related to the perceived sound quality and is therefore difficult to define and measure. In general, when engineers discuss *quality* they are referring to two different dimensions: function and beauty. Quality optimization usually starts with the former. An excellent car, for example, should never fail in its function of transportation; it must fulfill that role with high reliability. Once function is achieved, the focus shifts toward beauty. But because sonic beauty lies in the ear of the beholder, it is difficult to quantify and optimize.



Stephanie Freese

Perception of sound depends on the location and orientation of the listener relative to the source. For instance, the ear responds differently to sounds above, at, and below the horizontal plane. The changes that occur in the horizontal plane are the ones that stereo loudspeaker setups use to localize the sound between the two speakers. When the listener moves, therefore, the perceived sounds change. To make music seem more realistic when heard through headphones, audio engineers add a set of corrections (called *head-related transfer functions*) that restore some of the sense of location.

In sound-quality research, we have therefore focused more on functional quality aspects, such as localization, and less on beauty aspects, such as immersion. The first studies related to immersion were carried out in the context of speech perception, addressing familiar problems such as the functional difficulty of understanding a single voice when you are immersed in a loud party. The goal here is to improve functional localization in order to optimize speech intelligibility. The same basic motivation inspires home theater systems that prioritize localization accuracy over auditory beauty.

In recent years, audio researchers have begun to focus more intently on the beauty aspect of immersion. In a 2019 study, Callum Eaton and Hyunkook Lee at the University of Huddersfield in the U.K. asked a group of consumers and audio professionals to rate 10 aspects of sound quality in relation to immersion. Eaton and Lee found that horizontal sound perception was more important than vertical, but they could not determine to what extent subjects prefer to be immersed by a sound. If we take a single-voice recording and play it over a standard stereo setup or over four loudspeakers, increasing the number of loudspeakers will improve the feeling of immersion but will not improve the perceived sound quality. For this reason, a single direct-radiating loudspeaker is preferable for reproducing a single-voice recording.

Many audio designers have recognized the importance of widespread directivity of loudspeakers for high-quality music reproduction. At the same time, we know that multiple direct-radiating loudspeakers are not well suited to creating an immersive diffuse field for music. To improve the feeling of immersion, those designers have used additional sound drivers that do not radiate directly toward the listener.

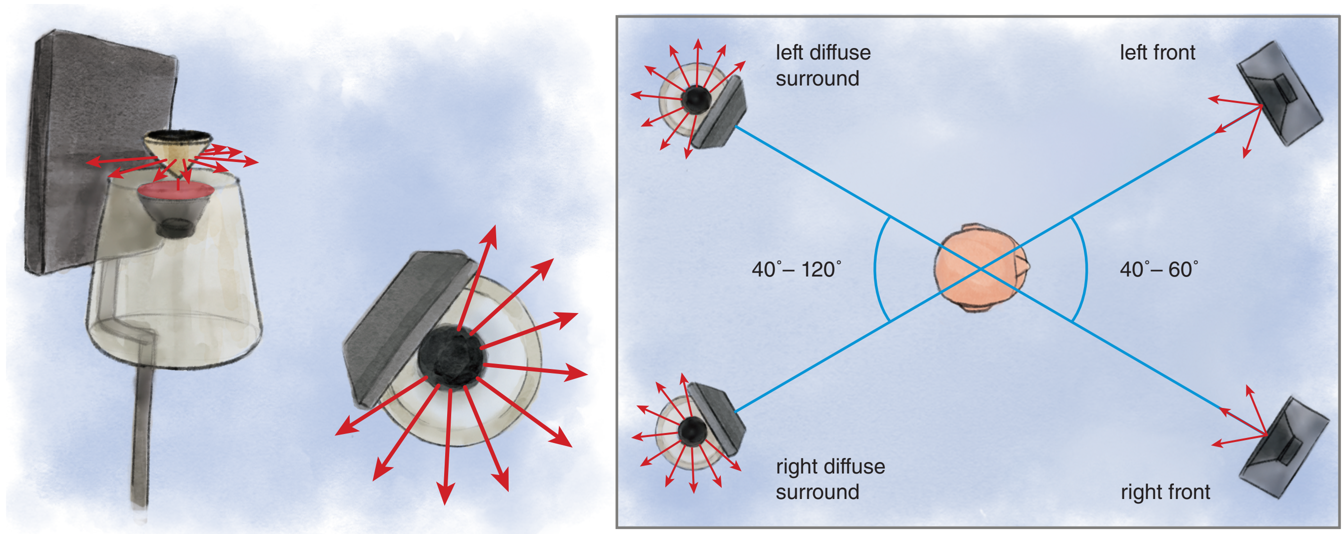
The best known of the people pursuing this direction is probably Amar Bose, founder of Bose Corporation, who in the 1960s designed a loudspeaker enclosure that has additional drivers in the back panel to produce reflections against the walls, thereby improving the balance between the direct and diffuse fields. In the 1980s, Kenneth Kantor and Alexander de Koster from Teledyne Acoustic Research in Cambridge, Massachusetts, extended the idea and developed an enclosure that uses extra backward radiating drivers to equalize the diffuse field room response independently from the direct field. One of us (Beerends) demonstrated the quality improvement from widespread directivity in 1988 for Dutch loudspeaker manufacturer BNS, using an extra set of back-radiating loudspeakers, which can be added to any regular stereo setup, to equalize the diffuse field response.

The weakness of all these setups is that they primarily create a frontally localized diffuse field. That distribution of sound does not closely replicate the diffuse field that a listener experiences during a concert-hall performance.

Once function is achieved, the focus shifts toward beauty. But because sonic beauty lies in the ear of the beholder, it is difficult to quantify and optimize.

A Simple Loudspeaker Solution

Today's home audio listening experience often falls into one of two extremes. At one end, we have a simple, tabletop Bluetooth speaker or a mono radio/television loudspeaker producing a single-source sound with one exact location, allowing excellent speech reproduction. At the other end, we have elaborate, multi-channel home theater setups producing highly detailed but mostly exaggerated localizations. In the middle of these extremes,



Stephanie Freese

The authors' experimental loudspeaker setup can produce a realistic mix of direct and diffuse sound. A cone-shaped diffuser (*left*) radiates sound in all horizontal directions (*arrows*), while a sound-absorbent block (*gray*) shields the listener from sounds that would arrive directly. Two front loudspeakers (*right*) create direct sound while two rear loudspeakers create an adjustable level of diffuse sound, mimicking the immersive experience of a live concert.

we have the traditional stereo setup that many people still use for listening to their favorite music. However, none of these designs does much to recreate a diffuse sound field that allows for a rich, immersive music listening experience.

We see a big missed opportunity, because excellent quality of immersion can be achieved using ordinary stereo recordings reproduced by a regular stereo loudspeaker setup, complemented only by two additional omnidirectional loudspeakers that project most of their sound energy toward the walls. In our experiments, we have shown that the two additional speakers can be designed to contribute only to the diffuse field, so the degree of immersion can be easily controlled without introducing localization errors. This setup also reduces undesired *comb filtering effects*, the sharp frequency peaks and dips that arise when sound waves interfere between the front and rear loudspeakers.

We have devised a simple but effective way to create a diffuse-radiating surround speaker using a cone-shaped diffuser that produces, for a substantial part, a 360-degree pattern of sound that radiates horizontally. Optimally, the speaker is designed to minimize its contribution to the direct field, for example, by limiting the actual radiation to about 300 degrees.

The basic layout of a complete loudspeaker configuration designed for an optimal sense of immersion can be adapted to one's personal preferences (*see figure above*). In our setup, the left and right diffuse speakers mainly radiate toward the walls of the listening room, as opposed to the standard surround setups in which the surround speakers radiate directly toward the listener. This approach prevents the "things jumping around" effect. Our setup can't create a full three-dimensional diffuse field because it is designed mostly to spread out sound along the horizontal plane, but the feeling of immersion is dominated by horizontal sound anyway.

The proof of the playback is in the listening, so over the past few years we have carried out a series of experiments in cooperation with a number of small hi-fi companies in The Netherlands. These experiments were conducted in four locations: three in a professional listening room, and one in a home environment. Both professional audio engineers and nonexpert listeners were asked to set the optimal playback level of the front loudspeakers, after which they were asked to adjust the level of the diffuse surround speakers for maximum perceived overall audio quality. We also adjusted the time delay between the surround speakers and the front ones, to keep the main stereo image (the sense of sound location) stable and prevent the rear speakers from creating unwanted localization from behind.

For the delay, we found that the optimal value was between 10 and 20 milliseconds, depending on the acoustic properties of the room where the recording was made. Roughly speaking, more delay could be allowed for recordings that are made in large concert halls than for dry pop-music recordings. The optimal volume level for the front speakers depended marginally on the preference of the test subject and not on the properties of the recordings, as they were equalized in loudness. The optimal level for the surround speakers depended significantly on both the test subjects and on the properties of the recording.

We were interested to learn that listeners' preferred levels for the diffuse field varied significantly. Some subjects set the level very low, close to the minimum noticeable volume, about 20 decibels below the level of the direct field loudspeaker. Others choose to set the diffuse sound level very high, even above the volume of the direct field loudspeakers. We also gave our testers the option to turn off the surround speakers entirely. Among the 24 test subjects, 23 chose to switch on the extra diffuse field speakers for most of the music samples, and 16 subjects chose to keep the speakers on the whole time. Even our least enthusiastic subject switched on the diffuse speakers for 43 percent of the samples.

Overall, our testers reported a significant increase in perceived overall sound quality when the diffuse surround speakers were switched on. Using a 5-point scale, ranging from 1 (a very small improvement) to 5 (a very big improvement), the audio experts judged the overall sound quality im-

Piano Piano!/flickr.com/photos/hansthijs/3585941755/

Quadraphonic sound was an attempt by the audio industry to create a home hi-fi experience that was more immersive than conventional stereo. Despite the wide availability of quadraphonic recordings and equipment in the 1970s, the technology flopped—probably because it failed to capture the way the people actually experience immersion and the locations of sounds.

provement around 3 on average. The nonexpert listeners judged the quality improvement even bigger, with average scores around 4.

The most encouraging aspect of these experiments is that only two small additional surround speakers were needed to produce a significant increase in overall perceived sound quality. Our diffuse field approach did not introduce the degrading localization errors that occur in many surround-sound systems. The setup we created allows for a simple "immersion control": Listeners can easily adapt the main volume, diffuse volume, and time delay characteristics of any standard stereo recording to their personal immersion preferences.

Hi-Fi in Your Life

The long quest for high-quality, widely accessible hi-fi audio is far from over. The extreme dependence of the optimal audio experience, especially perceived immersion, on personal preferences makes it difficult to design an objective system for assessing the overall sound quality of a system. For mono speech and music, and to some extent stereo music, audio engineers largely have conquered the basics. Perceptual models have been developed that show good correlation between objective measurement and subjectively perceived speech and music quality. There are also useful models for spatial audio quality, although they do not take into account personalized immersion optimization.

The major shortfalls of currently available commercial audio systems are that most of them provide only limited or ineffective amounts of immersion, and that none of them allow easy adaptation of immersion to personal preferences. Those preferences also vary strongly depending on the room in which the sound reproduction takes place. One of us (Beerends) has been experimenting with home theater systems that can generate artificial sound reflections, using algorithms to simulate the acoustics of concert halls. This approach allows listeners to optimize the feeling of immersion in rooms that sound too dry, lacking enough acoustical reflections. However, such systems do not make it easy to dial in an optimal level of immersion, and they can lead to sound localization errors.

The diffuse sound setup that we developed offers a simpler yet effective way to optimize the feeling of immersion, but for now it exists only as a prototype. Currently no company manufactures such a system. We hope that this article will encourage manufacturers to commercialize a system that can be hooked up to any standard hi-fi set, allowing for easily controlled and optimized immersion into the music.

Bibliography

Beerends, J. G., and J. A. Stemerdink. 1992. A perceptual audio quality measure based on a psychoacoustic sound representation. *Journal of the Audio Engineering Society* 40:963–978.

Beerends, J. G., and J. A. Stemerdink. 1994. A perceptual speech quality measure based on a psychoacoustic sound representation. *Journal of the Audio Engineering Society* 42:115–123.

Beerends, J. G., A. P. Hekstra, A. W. Rix, and M. P. Hollier. 2002. PESQ, the new ITU standard for objective measurement of perceived speech quality, part II—perceptual model. *Journal of the Audio Engineering Society* 50:765–778.

Beerends, J. G., et al. 2013. Perceptual objective listening quality assessment (POLQA), the third generation ITU-T standard for end-to-end speech quality measurement part II—perceptual model. *Journal of the Audio Engineering Society* 61:385–402.

Eaton, C., and H. Lee. 2019. Quantifying factors of auditory immersion in virtual reality. In: *2019 AES International Conference on Immersive and Interactive Audio*, eds. T. Tew and D. Williams. Audio Engineering Society.

Thiede, T., et al. 2000. PEAQ—The ITU-standard for objective measurement of perceived audio quality. *Journal of the Audio Engineering Society* 48:3–29.

John G. Beerends researched speech, audio, and video quality assessment for Dutch telecom operator KPN. For the past 22 years he worked for TNO (Dutch Organization for Applied Scientific Research), focusing on the idealization approach in speech quality assessment. Richard van Everdingen has worked in audio and video distribution, with an emphasis on audio quality optimization. In 2010 he founded Delta Sigma Consultancy, where he has conducted research on audio and video-related issues. Email: johnbeerends@hotmail.com

The Discovery of Nothing

Creating a vacuum on Earth led not only to cleaning tools but also to weather forecasting, light bulbs, televisions, computers, and modern medical imaging.

Mark Miodownik

I remember Felix Baumgartner jumping from a high-altitude balloon 39 kilometers above the Earth's surface. It was streamed live on the internet in 2012. Before he jumped, we saw footage of him in his capsule, on the edge of space, preparing to leap. Below, we could see the blue planet Earth in all its spherical magnificence. Baumgartner was wearing a spacesuit because his balloon had reached the stratosphere. At that height, there is very little of the Earth's gas atmosphere and almost no oxygen. The temperature outside the capsule was –57 degrees Celsius. As I waited, watching the live video feed, I envied him being up there between heaven and Earth, in this place where the gas atmosphere of our planet ends and the mysterious sublime state of nothing stretches out into the universe.

The materiality of space has puzzled humans throughout the ages. What really is it? Surely space can't actually be nothing? The ancients agreed. Aristotle declared that "nature abhors a vacuum." The heavens were thought to be filled with a sacred material. The Greeks called it *aether*, the substance the gods breathed, the fifth element, separate from the four that made up the earthly realm: earth, air, fire, and water. It allowed light from the stars to propagate, and by medieval times it was also holding planets in their orbits. Even when Isaac Newton proposed gravity as a force in 1666, it relied on aether to propagate across the Solar System. But no one could actually find a trace of this material, and as science began to rely on it more, so finding it became more urgent. The story of the search for this material starts back on Earth—the same Earth that Felix Baumgartner hurtled toward at some 1,357 kilometers per hour as he jumped from his balloon on October 14, 2012, almost certainly not thinking that the technology protecting him from the vacuum of space was in any way linked to this ancient quest for aether.

Under Pressure

An Italian and a student of Galileo Galilei called Evangelista Torricelli was one of the first to make a breakthrough in the search for aether in 1641. His experiment was simple and elegant. He took a tube of mercury and turned it upside down in a bowl of mercury. Remarkably, such an experiment shows that mercury does not rush down to the bottom of the tube pulled by gravity, as you might expect. It falls a short distance and then stops. For a meter column of mercury, roughly 76 centimeters of it stay up the tube, defying gravity. But there is a gap at the top where 24 centimeters of mercury used to be but are not there anymore. Torricelli asked what is in the gap. It is not air, because no air could get in. So it is a vacuum, just like the vacuum of space, and presumably filled with aether. Could this invisible aether be responsible for the mysterious force holding up the mercury against the force of gravity?

The answer is no. Torricelli showed that there was a much simpler explanation. The air we breathe forms atmosphere on our planet, and despite being a gas, it has weight. It pushes down on us and everything it surrounds. It is this air pressure that pushes down on the bowl of mercury, pushing the mercury up the tube. At the same time, the column of mercury is pulled down by gravity. When those two forces are equal determines the height of the mercury. This balance is why the height of the column changes depending where you are on the Earth's surface. At sea level, the height of the mercury column is 760 millimeters. If you go up a mountain, the column gets smaller. This change is because there is less air above you, less air pushing down on the surface of the mercury in the bowl, so less pressure pushing the mercury up against the force of gravity. If Baumgartner had done this experiment in his balloon 39 kilometers above sea level, he would have found that the tiny amount of atmosphere above him pushed so feebly that the height of the mercury column would have been 3 millimeters. So the vacuum doesn't do anything; its role is not to push back. What Torricelli had done was to find a way to create a vacuum on Earth. It had lots of technological implications, some of which would end up being the creation of TVs, computers, and vacuum cleaners. But before that, a more immediate invention beckoned, a way to measure atmospheric pressure: the barometer.

The barometer turned out to be able to predict the weather—or at least some aspects of the weather. It could detect invisible changes in air pressure associated with different weather patterns, because they changed the height of the column of mercury in the glass tube. Those analyzing weather

QUICK TAKE

The heavens were once thought to be filled with a sacred material that the Greeks called *aether*, which allowed light from the stars to propagate across the universe.

Experiments with columns of mercury, and evacuated spheres and glass globes, led to an understanding of atmospheric pressure and the properties of a vacuum.

The use of a vacuum is not limited to cleaning, but played a vital role in the invention of light bulbs, televisions, and early computers, as well as x-ray machines and silicon chips.

Abaca Press/Alamy Stock Photo

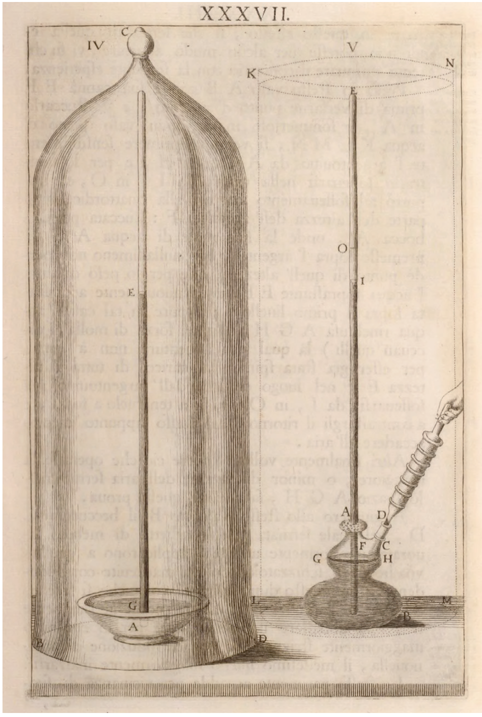
In October 2012, skydiver Felix Baumgartner took a balloon to the edge of space and jumped out, becoming the first person to break the sound barrier unaided by vehicular power. At that height, some 39 kilometers above the Earth's surface, the air pressure is so low that standard phenomena we see on the ground don't occur, such as the movement of a mercury barometer. Centuries ago, experiments that involved air pressure demonstrated that a vacuum really is filled with nothing, which over time led to a number of important technological advances.

patterns realized that high pressure was often associated with clear skies and sunny weather, whereas low pressure (a small column of mercury) accurately preceded rain and storms. The phrase "the mercury is sinking'" started to become used by sailors. It meant that the height of the column of mercury in their barometers was decreasing, indicating a low-pressure weather system was approaching and potentially a storm. Now they didn't have to pray to the wind gods or leave them offerings in order to know when was a good time to set sail. To this day, air pressure is still measured in millimeters of mercury, denoted mmHg, as a result of this invention 400 years ago.

The development of this weather-forecasting tool was an unexpected bonus of exploring nothingness, but scholars of the time still had the puzzle of vacuums. Surely a region of the glass barometer with absolutely nothing in it was impossible: It had to be filled with something, even if that something was not air. Light travelled through the space at the top of a barometer, just like it travelled through outer space. So, they argued, they both should be filled with aether. They considered it a fundamental element of the universe, a perfect substance, but one that could perhaps be chemically isolated.

The materiality of space has puzzled humans throughout the ages. What really is it? Surely space can't actually be nothing?

So the quest to isolate and distill aether began. It was led by the alchemists, who called it *quintessence* (the fifth element) and thought it could be used as a medicine to cure disease. Illness, at the time, was thought to be something that came from within a person, an imbalance of the four humors: black bile, yellow bile, blood, and phlegm. Quintessence, the perfect substance, it was argued, could balance these humors and thus cure a person of illness. Others came to believe that quintessence was the fabled philosopher's stone that could turn base metals into gold. Once again it was a question of balance: Lead had an imperfect balance of the fundamental substances sulfur and mercury, and was thus a base metal that was soft and corroded easily. Quintessence could adjust the balance and so make this substance into perfect gold. Success in distilling quintessence would bring fame and wealth, but more importantly, com-



Courtesy of Linda Hall Library of Science, Engineering & Technology; Wikimedia Commons (*right*)

In 1641, Evangelista Torricelli (*above*) found that the mercury in a glass tube turned upside down into a bowl of mercury falls a short distance and stops (*left*). The empty area at the top of the tube has to be a vacuum. Torricelli showed that air in the planet's atmosphere presses down on everything, including the mercury in the bowl, which balances with the force of gravity to determine the height of mercury in the tube. This result was later used in mercury barometers (*above right*) to determine air pressure and the chance of rain.

plete their quest to become close to God by studying and understanding God's creation. And so the search for quintessence became a holy quest.

Stuck Together

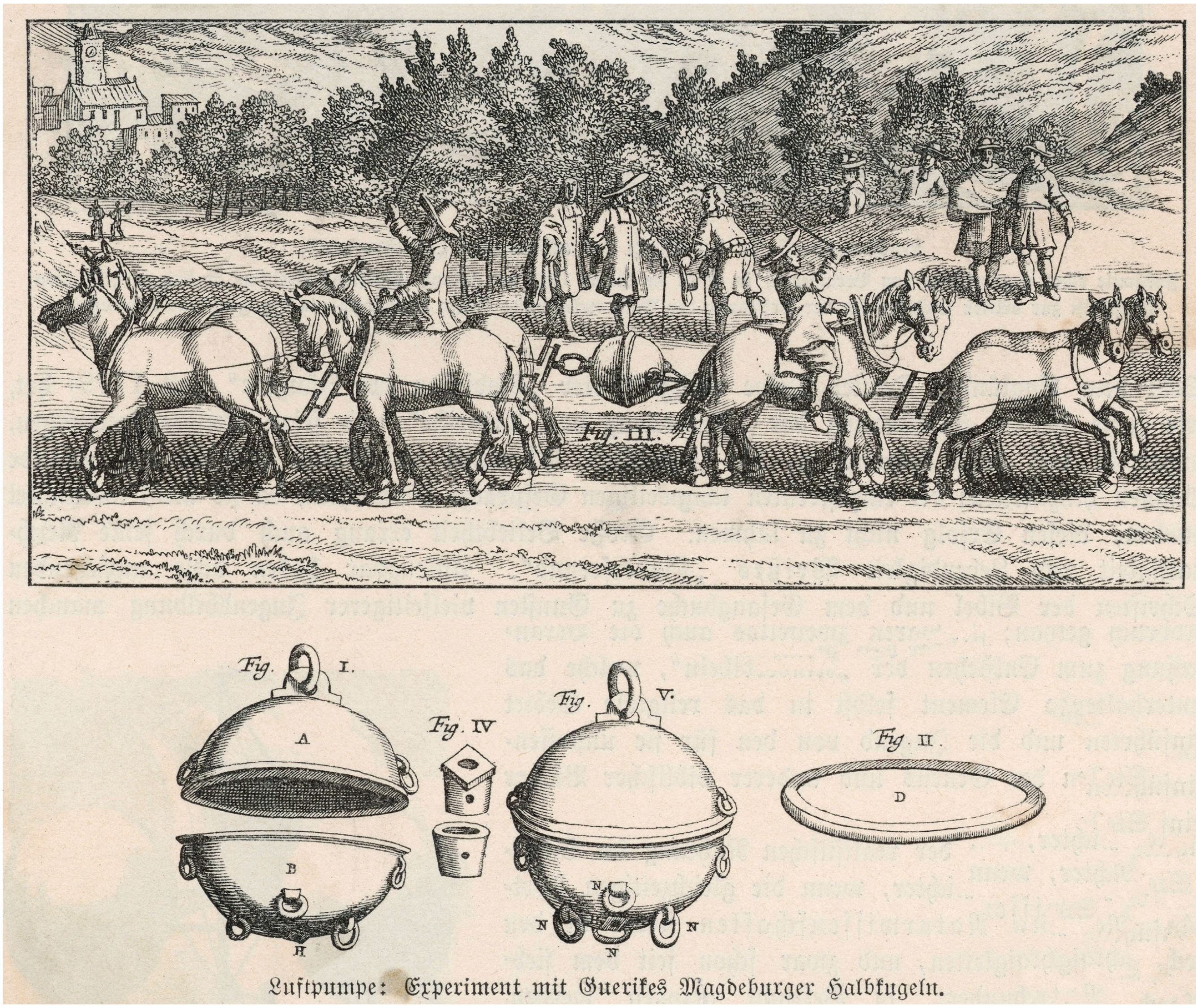
The person who made the next big breakthrough was not an alchemist, though, but the mayor of the German town of Magdeburg, Otto von Guericke. As a politician, he traveled across Europe, which meant he was exposed to new ideas and the big scientific problems of the day. A devout man, he got to hear that quintessence might be the substance that filled the vacuum at the top of a barometer. Not being an alchemist turned out to be an advantage, because he did not have preconceived ideas of the right way to obtain quintessence. While alchemists were using all sorts of methods of chemical distillation, von Guericke did something completely different: He decided to isolate nothingness mechanically.

To do so, he invented an air pump. It is a device we would recognize today as similar to a bicycle pump, except that the valves are reversed, so that each stroke of the cylinder removes air from whatever it is connected to, and then on the return stroke prevents the air from coming back. The mechanism is simple, but the execution is not. Whenever you remove air from a container, the air pressure outside the vessel creates a force driving air back into the container. This force gets bigger the more air you remove. Any leak in the valves or the fabric of the container destroys the vacuum. So to make it work requires precision engineering.

We take the accuracy and intricacy of screws, gaskets, and valves for granted today. In the 17th century, such precision engineering was just beginning: For instance, the mechanical clocks in city centers were only able to keep time to an accuracy of 10 minutes in a day. Nevertheless, through ingenuity, perseverance, and many failures, von Guericke succeeded in constructing an airtight pump. Despite this engineering success, he probably wouldn't have been credited as being pivotal to the understanding of vacuums if he hadn't also been a bit of a showman. He showed the power of his air pump with a demonstration that would blow the minds of everyone who saw it.

Von Guericke made two hemispheres of bronze that were machined so accurately that when they were placed together, they fitted to each other exactly. One had a small pipe incorporated to allow von Guericke's vacuum pump to be fitted. Then he assembled the important people in the land, including the king of Prussia, to witness something incredible. He showed everyone the two hemispheres. They were just two pieces of not-very-interesting metal. Then he put them together to create a hollow sphere of metal. Next, he used his air pump to remove the air from this internal space and create a vacuum. Now there was nothing physical holding the hemispheres together: no bolts, no straps, no welding, no glue. Everyone could see that. Nothing. Then he assembled two teams of eight horses. The first team of horses was harnessed to one-half of the now-joined Magdeburg sphere (as it came to be called) and the second team to the other, the

Whenever you remove air from a container, the air pressure outside the vessel creates a force driving air back into the container.



Chronicle/Alamy Stock Photo

Two halves of a bronze Magdeburg sphere (*above, bottom*) were machined precisely to fit together exactly, and then an airtight pump was used to evacuate all the air inside. The suction from the vacuum inside the sphere, and the atmospheric pressure pushing on it with nothing pushing back, held the hemispheres together so strongly that a team of horses was unable to pull them apart (*above, top*). This 1654 demonstration launched a trend of public experiments with air pumps and glass vessels that showed, among other things, that sound could not travel through a vacuum, but magnetism and electricity could.

two teams facing in opposite directions. Presumably, the horses neighed and stamped their hooves, not knowing what was going on. Perhaps the wind dramatically ruffled their manes. Then von Guericke drove the two teams away from each other, trying to make them pull the two halves of the sphere apart. They pulled against suction, but they could not defeat it. A pump, and some precision engineering, had created a suction that could defy 16 horsepower. But it wasn't a force from the vacuum inside. Just as with the barometer, atmospheric pressure was pushing the two hemispheres together, and, without air inside, nothing was pushing back.

Investigating Nothingness

Soon, engineers and instrument makers across Europe were building their own air pumps and using them to explore the anatomy and properties of vacuums. As with von Guericke's demonstration, part of the magic was the public nature of the experiments. Famous scientists of the day such as Robert Boyle started using air pumps to evacuate glass vessels, so that anyone who cared to look could see what was going on inside. These demonstrations became public entertainment as well as pushing forward the science.

Does a bell ring in a vacuum? Answer, no: Sound waves need air as a medium to travel. Does a candle burn in a vacuum? Answer, no: But oxygen had not been discovered yet, so there was no good explanation. Can an insect fly in a vacuum? Answer, no: Wings need a gas to create lift. Can a snail survive in a vacuum? Answer, no: It dies. Can a mouse survive in a vacuum? Answer, no: It dies. Can a bird fly in a vacuum? Answer, no: It flutters and then dies in agony. What happens if you put a compass in a vacuum; does it still point north? Answer, yes: Magnetism is unaffected by a vacuum. Does electricity flow in a vacuum? Answer, yes: And light travels through it without a hitch too. Ah-ha, you're thinking: A clue! And yes, you're right, this result is exactly why the scientists of the day were so excited about these discoveries.

Wikimedia Commons/The National Gallery, London

A 1768 painting by Joseph Wright, *An Experiment on a Bird in the Air Pump*, captures the trend at the time of public demonstrations that used evacuated glass vessels. A small Magdeburg sphere sits on the table as a reference to the earlier experiment. This demonstration uses a live bird, to the distress of some and the intent fascination of others, to display novel information about how the universe works.

So it was that von Guericke's air pump was crucial to build the evidence that although some things, such as sound, needed the medium of air to travel, others, such as light, magnetism, and electricity, did not. Perhaps they were special in some way, or perhaps they were connected to whatever there was in a vacuum that allowed them to travel, not just through a vacuum but across space and time. The potential role of quintessence was expanding.

In 1768, the spectacle of the popular and mysterious air pump experiments was captured in a painting by Joseph Wright of Derby in the United Kingdom. Called *An Experiment on a Bird in the Air Pump*, it now hangs in the National Gallery in London. There is wonder and sorrow in that painting. The central figure conducting the experiment is a man looking out toward the viewer with an impartial expression, as if to say, "This is how to understand the world." Some of the onlookers are covering their eyes, distressed at the cruelty of experimenting on live animals. Others are staring intently at the demonstration, utterly fascinated by this insight into how the universe works. On the table is a small Magdeburg sphere, a reference to the origin of these pumps and the quest to understand air, vacuums, and quintessence.

I wish I could say that this was one of the paintings I remember as a kid. I wish I could say that I stood transfixed in front of this painting on one of our many visits to the National Gallery, where my mom's relationship with the gods of parking allowed us to access the museum with ease. But unfortunately, I don't remember seeing this painting as a child, even though my mom almost certainly would have shown it to us—yes, because it is a masterpiece, but also because it's a tangible connection between her and my dad. He was a renowned metallurgist and very much involved in exploring how the world works through experiment and philosophy. She would have appreciated the mystical and ceremonial quality of the painting, with the candlelit setting in particular lending the scene a spiritual air; this effect was all lost on us boys, who were probably running amok in the gallery. I was perhaps like the boy in the painting who is not looking at the experiment but instead fiddling with the window blind, and in doing so letting moonlight into the room. This part of the scene is an intentional reference by the painter to the Lunar Society and the questions being asked at the time about how light travels from the Moon to Earth through space. One of the reasons why something like quintessence had to exist was because it was thought light waves needed a medium by which to travel through space. Sound waves traveled through air, sea waves traveled on water—what was the equivalent medium for space? Scientists called it *luminescent aether*—renamed because they couldn't find quintessence.

Brilliant Moments

Meanwhile, the engineers, who had been spending a lot of time making vacuums in glass containers, were getting annoyed at having to continuously pump out the container every time they wanted to do an experiment. What if, they reasoned, once the glass vessel contained a high-quality vacuum, the glass was melted to seal the vacuum inside the chamber. This seal produced a permanent vacuum inside the glass on which to experiment. Of course, you could not move things in and out of the container once it was sealed, so you had to decide what you were going to experiment on and leave it in there. Metal wires could be used, for instance, connected at either end of a glass tube, or glass bulb as it was called. When a voltage was applied, electricity would flow through the tube and the wires would grow very hot, which caused them to glow red. It was the birth of the electric light bulb, an invention deemed so ingenious that the universal symbol for having a brilliant idea is a light bulb.

Early versions in the 1800s emitted light only for a short time, after which the hot glowing wires, called *filaments*, would then break. Scientists realized that for electric light bulbs to replace candles or gas lamps, the electricity would need to heat up the filament to temperatures exceeding 1,500 degrees. But there was a problem: This temperature exceeds the melting point of most of the metals used to conduct electricity. By the time British chemist Humphry Davy had a go in 1802, the metal platinum was the leading contender, with a melting point of 1,768 degrees. But white-hot platinum vaporized at that temperature quite quickly and so the filaments didn't last long. They were also very expensive. A cheaper conducting material was needed with a high melting point. (*For later filament developments, see "Tungsten's Brilliant, Hidden History," March–April 2020.*)

The British chemist and inventor Joseph Swan used graphite, which seemed perfect because solid carbon doesn't melt at all. You have to increase the temperature to 3,642 degrees before it gets so hot it evaporates into a gas, a process called *sublimation*. Swan took out a patent in 1860, and that should have been the beginning of a bright future. But the difficulty was that carbon reacts very easily with oxygen in the air. Of course, the vacuum inside the bulb should have meant that there was no oxygen. But the early mechanical air pumps did not produce perfect vacuums. They could reduce the air pressure enough to suffocate a bird, kill a mouse, or prevent an insect from flying, but there was still a small amount of air left in the glass bulbs. The oxygen in that air reacted with the carbon filaments and destroyed Swan's early electric light bulb.

Science History Images/Alamy Stock Photo

It was not Thomas Edison but Joseph Swan who first patented a light bulb, in 1860. The British chemist and inventor needed an inexpensive filament that didn't melt and lasted a significant time. Graphite seemed perfect, but vacuum technology didn't advance enough until 1875 to produce a vacuum that removed the oxygen that reacts with carbon, which would allow the filament to last 40 hours. One of Swan's early light bulbs is shown here.

It was not until 1875 that vacuum technology improved enough to create an electric light bulb with a carbon filament that could glow white-hot in a vacuum, providing ample light for 40 hours. Swan started with his own house in Gateshead, then lit a whole street in Newcastle-upon-Tyne, and then the Savoy Theatre in London. It was the future. American Thomas Edison is often credited with inventing the light bulb. He didn't. What he did was to see that the future of lighting was electric. He perfected the production and marketing of lighting systems, including the bulbs. He is famous for stating that an idea is only a small part of invention: "What it boils down to is 1 percent inspiration and 99 percent perspiration."

The perspiration in the case of the electric light bulb was the enormous number of experiments Edison performed on different designs of bulb. Most of them failed. But proof of the importance of perspiration and systematic testing was that one of his experimental light bulbs, which seemed to have no use, turned out to be the beginning of electronics and computers.

Mind the Gap

In essence, it was just a light bulb with a broken filament. What Edison's engineering team noticed was that you could still get electricity to flow through the vacuum, but only if the filament was hot. The electrons would jump across the gap between the broken filament from the negatively charged end to the positively charged end, but not the other way. This discovery was the birth of a component that would kick-start the electronics industry, and it was called a *vacuum tube*. These vacuum tubes acted as valves, the equivalent of the taps in your kitchen that control water flow. These valves allowed electric signals to be turned on and off by another electric signal (which heated the filament). This design was a programmable tap that

Stefan Riepl (Quark48)/Wikimedia Commons

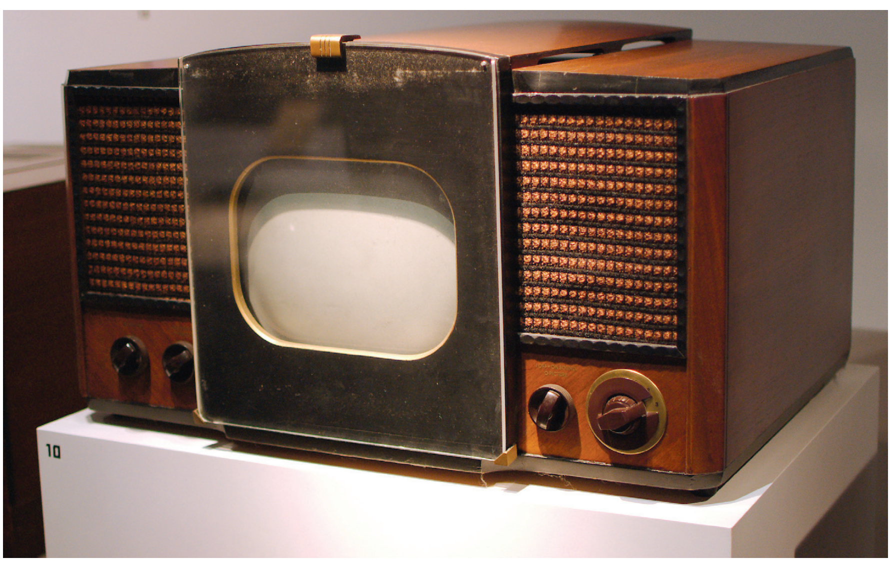
Many types of vacuum tubes came into existence as the electronics industry blossomed in the early 20th century. Vacuum tubes shared their origin with light bulbs, but basically had a broken filament. If the filament is hot, electrons flow through the vacuum across the gap, but only from the negatively charged end to the positively charged end. Thus the vacuum tube acted as a switch for the flow of electricity, controlled by an electric signal that heated the filament. Later, it led to the development of loudspeakers, radios, television, and early computers.

could tune and amplify electricity. It led to the development of the loudspeaker, the radio, and the television, the last-mentioned having at its heart one giant vacuum tube, called a *cathode-ray tube*.

Cathode-ray tubes have Edison's hot filaments at one end and a high voltage at the other, where there is a screen. The cathode ray is not a ray of light but a ray of electricity. It literally flies across the vacuum tube, but the only reason it reaches the screen is because there is no gas in the way for the ray of electrons to bump into. When the electricity hits the screen, it lights up because of a special coating called a *phosphor*. Now there is a bright spot on the screen. To make these TVs work, the ray is scanned across the screen very fast, row by row, so that each part of the screen is hit by the electricity 25 times every second. You would observe this scanning dot if you could see that fast, but you can't, so instead you see a continuous image of, say, a wizard casting a spell, or a tornado transporting a house through the air. I still remember these TVs from my childhood: We watched films such as *The Wizard of Oz* on them. They resembled vacuum laboratory equipment because that's exactly what they were. When you turned the TV off, there was a click and the screen suddenly went blank, except for a single dot in the middle. This dot was the place where the last electrons had hit the screen. The place still glowed for a second before fading to nothing. It was always a sad moment for me and my brothers. The appearance of that dot meant we were going to bed.

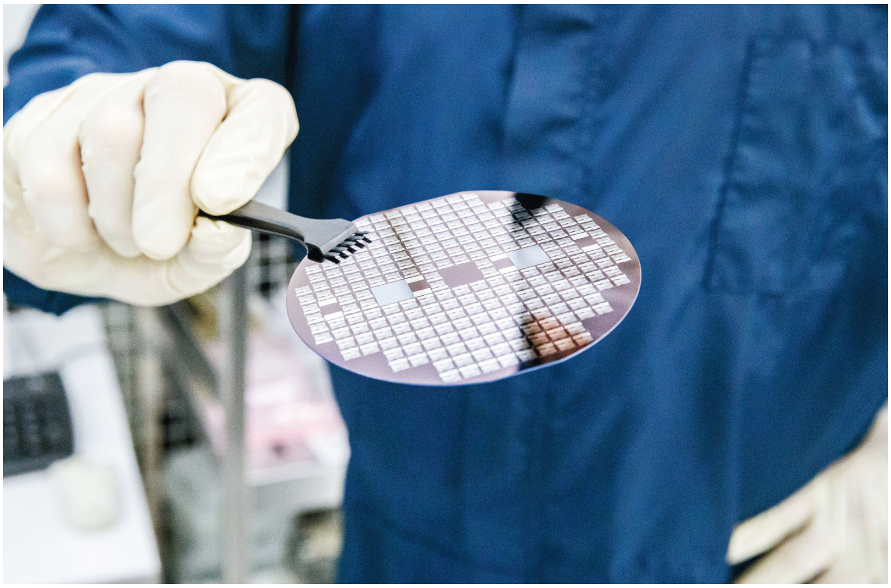
TVs in those days were huge, heavy things. They were weighty because the cathode-ray tube was made of glass, and it was not just ordinary glass. A by-product of accelerating electricity to create that dot on the screen is the creation of x-rays, the same x-rays that are used in hospitals to detect broken bones and cancer tumors (and yes, hospital x-ray machines are also vacuum tubes). To protect TV viewers, these x-rays had to be stopped before they escaped from the vacuum tube and radiated everyone watching the TV programs. That meant adding lead to the glass, which absorbed the x-rays. This process worked, but lead, being a very heavy element, increased the weight of the TVs, which were the size of armchairs in my childhood.

Most of these enormous TVs are gone now, freeing up a lot of space in



Fletcher6/Wikimedia Commons

Early televisions resembled laboratory equipment. At their heart was one giant vacuum tube, called a *cathode-ray tube*. These tubes had a hot filament at the back and a high voltage at the screen in front. A ray of electricity flew across the vacuum and lit up a coating, called a *phosphor*. The ray scanned across the screen 25 times a second to create the moving image.



dpa picture alliance/Alamy Stock Photo

Although modern computers do not use vacuum tubes to operate, a vacuum still comes into the process during the manufacturing of the silicon chips that power their computational abilities. Silicon wafers must be processed under an ultrahigh vacuum so that their surfaces do not become contaminated with impurities from the air, which would make the wafers unusable.

our living rooms but leaving me with a feeling of nostalgia for the simplicity of when we only had three TV channels to watch. They have been replaced by liquid crystal flat-screen technology controlled by silicon chips, with hundreds of TV channels. This materials science invention of silicon chips from the 1950s created the revolution in computing, replacing glass vacuum tubes. Silicon chips are a core technology in every computer, mobile phone, car, washing machine, and piece of hospital equipment. These silicon chips need to be manufactured in ultrahigh vacuums; otherwise, they become contaminated with impurities from the air, which render the chip worthless. Thus a thousand-year-old quest to create the purest "nothing" still continues. And there is still plenty to do, because we as yet can't even make a vacuum as pure as that found in outer space, which is millions of times purer.

A thousand-year-old quest to create the purest "nothing" still continues, because we as yet can't even make a vacuum as pure as that found in outer space.

Cleaning Up

For most people, the holy grail of vacuum technology is not their mobile phone, despite its importance and much as they might love it. It's not the vacuums used in medical technology to produce x-rays, much as they care about its importance for diagnosing illness and tooth decay. It's not the vacuums used in the scientific equipment in every lab in the world, without which scientific research would come to a standstill. These uses are all too remote and hidden from view to be of daily concern to citizens of the world. No, for most people the most important vacuum in their life is inside their vacuum cleaner. These machines, like the early steam engines, harness atmospheric air pressure created by the hundred kilometers of air above our heads to clean our homes. They create a vacuum inside the machine, which causes air to rush in to equalize the pressure, and in doing so it sucks up dust as the vacuum cleaner kisses the floor. It is so simple, and yet so marvelous. It has made all of our homes less filthy, especially homes with fitted carpets, which would otherwise be dirty, dusty, and smelly. The vacuum cleaner is the stalwart of the home, creating order and cleanliness. It has even played its part in creating more equality between the sexes, making cleaning faster and more effective—freeing time for other things, such as careers and hobbies—and also lowering the barriers to those reluctant to contribute to cleaning the home.

This point brings us back to the search for luminescent aether, the perfect substance, said to inhabit space. By 1905, Albert Einstein's special theory of relativity banished the need for aether to explain how gravity works and how light travels through space. According to this theory, there is no need for aether, and "nothing" really does exist. It is the creation of nothing inside a vacuum cleaner that harnesses atmospheric air pressure to clean our homes. It's the nothing inside a light bulb that allows light to emerge. It's the nothing inside an x-ray tube that helps doctors diagnose illness. It's the nothing in vacuum chambers that allows us to test the safety of space suits, enabling Felix Baumgartner to safely jump from a balloon on the edge of space. The purer the nothing, the more effective it is. Less is quite literally more, when it comes to a vacuum.

Bibliography

Bridle, S. L. 2020. *Food and Climate Change without the Hot Air.* UIT Cambridge Ltd.

Connor, S. 2010. *The Matter of Air: Science and Art of the Ethereal.* Reaktion Books.

Crump, T. 2007. *A Brief History of the Age of Steam: From the First Engine to the Boats and Railways.* Running Press.

Fisher, D. E. 2010. *Much Ado about (Practically) Nothing: A History of the Noble Gases.* Oxford University Press.

Johnson, S. 2009. *The Invention of Air.* Penguin Random House.

Kean, S. 2017. *Caesar's Last Breath: The Epic Story of the Air around Us.* Doubleday.

Lane, N. 2016. *Oxygen: The Molecule That Made the World.* Oxford University Press.

Tomory, L. 2012. *Progressive Enlightenment: The Origins of the Gaslight Industry, 1780–1820.* MIT Press.

Mark Miodownik is a professor of materials and society, and director of the Institute of Making, at University College London. This article is excerpted and adapted with permission from It's a Gas: The Sublime and Elusive Elements That Expand Our World *(Mariner, 2024). Online: markmiodownik.net*

Kicking Cocaine

Once lauded as a cure-all, by the 20th century the drug's reputation soured to that of a societal scourge.

Douglas Small

In the winter of 1886, William A. Hammond—a famed neurologist and the former Surgeon General of the United States Army—took an enormous amount of cocaine. A reporter from the New York paper *The Sun* who interviewed him waggishly observed that the doctor had been "on a terrific spree for science." Hammond had experimentally worked his way through as many different ways of taking the drug in as many different quantities as he could devise: He tried fluid extracts of coca (the plant from which pure cocaine is extracted), mixed grains of cocaine hydrochloride into purified wines, and eventually began injecting the drug hypodermically. The injections, he said, gave him "a delightful, undulating thrill." On cocaine, everything felt "refined" and "softened." Hammond became intensely talkative: When he was alone, he would talk to himself at great length. "I became," he said, "rather sentimental and said nice things to everybody. The world was going very well, and I had a favorable opinion of my fellow men and women . . . I enjoyed myself hugely."

Hammond went on taking the drug in increasing amounts until "the sensations became rather painful than agreeable." He eventually pushed his tests as high as 18 grains (just over 1 gram) in a single dose, which caused him to become "oblivious" to his own actions. He woke up in bed the next day with no memory of how he got there, and quickly discovered that he had, at some point in the night, decided to thoroughly wreck his own library. After this experience (and after recovering from a "most preposterous headache that lasted two days") he called a halt to the experimentation.

He might have been unusually enthusiastic in his experiments, but Hammond's fascination with cocaine was far from uncommon for a medical professional of his time. In early 1885, *The Lancet* laconically observed that "The medical press is full of cocaine just now." By the end of the year, the sheer volume of publications dealing with the substance had become "so extensive and so many sided that it is difficult to deal with it summarily." Cocaine had been chemically isolated decades before, but it had mostly been seen as a scientific curiosity—an "obscure" and "useless alkaloid," as one medical journalist later put it. The substance's sudden ascent from near-total obscurity to worldwide celebrity was due to a single, remarkable innovation: the discovery that cocaine could be employed as the world's first local anesthetic.

Thanks to cocaine, it became possible for the first time to eliminate pain without resorting to more powerful (and dangerous) general anesthetics

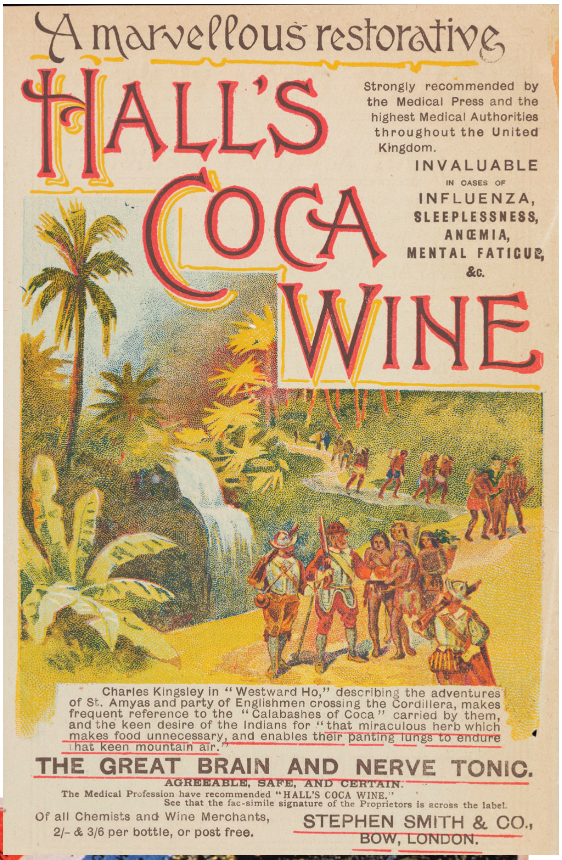
In the late 19th century, cocaine was a key ingredient in lozenges, tonics, and various other medications. Many of these products were marketed to women and children; for example, Iron Bitters claimed in one advertisement that it would help children grow "healthy and strong" (*facing page, lower left*), and in another that it would cure "female infirmities" (*above top*). These claims stemmed from the drug's efficacy as a local anesthetic, which could mask many symptoms. By the turn of the century, however, public perception of cocaine changed as widespread use took the shine off of the previously elite treatment and as the drug's dangerous and addictive side effects became apparent.

QUICK TAKE

Physicians had long recognized cocaine's potential as a pharmaceutical, but its true value became apparent in the late-19th century as the world's first local anesthetic.

Cocaine became a sign of modernity tied to technological progress. It was used to sharpen the mind and to cure an array of ailments, from sore throats to malaria.

The craze for cocaine transformed the drug's image from miraculous to poisonous. Early-20th-century attitudes toward cocaine reflected social and racial anxieties.



National Library of Medicine; Wellcome Collection

Library of Congress

The National Library of Medicine

Both William A. Hammond and Karl Koller studied the effects of cocaine by taking the drug themselves. Hammond (*left*), who had served as Surgeon General of the U.S. Army during the Civil War, recorded his experience taking increasing doses of the drug. Koller (*right*), an Austrian ophthalmologist, was introduced to cocaine by his friend Sigmund Freud. Koller's experiments led to a breakthrough when he discovered that cocaine could be used as the world's first effective local anesthetic.

such as chloroform. This breakthrough captivated the public imagination in a way that few substances have, before or since. For many, cocaine seemed to convey the promise of the modern, technologically dynamic 19th century: a quickening new age of scientific revelations, new inventions, and marvels on an industrial scale. The story of cocaine between the end of the 19th century and the start of the 20th is one of the slow change from a technological wonder to a dangerous drug of addiction. It is also a story that illustrates the ways in which individual substances can become loaded with ideological meanings, how those meanings can change as they spread through society, and how our perceptions of particular drugs are intimately bound up with our feelings about the people who use them.

For many, cocaine seemed to convey the promise of the modern, technologically dynamic 19th century.

Coca Koller

Karl Koller was never to become as personally famous as his friend Sigmund Freud, but he did manage to make cocaine very famous indeed. In 1884, Koller was 27 years old and working as an intern in the eye surgery department of Vienna General Hospital. He was professionally ambitious and hoped that an important-enough discovery might allow him to apply for a position at one of the city's large and prestigious eye clinics. To this end, he began doing laboratory work on experimental anesthetics. Both ether and chloroform (the two anesthetics primarily in use at the time) had side effects that made them awkward to employ in eye surgery, and Koller hoped that he might make his name by finding a better alternative. It was Freud who introduced Koller to cocaine. Freud—then a similarly young and ambitious medical man—had been toying with the idea of using the alkaloid as a stimulant and a treatment for heart disease and nervous exhaustion. He asked his colleague to help with his experiments, and so—as Koller recollected years later in the journal *Anesthesia and Analgesia*—the pair began "taking the drug by mouth" and recording its various effects.

Toward the middle of the year, Freud left on a month-long visit to see his fiancée while Koller continued the work on his own. He noticed that cocaine had a numbing effect when applied directly to the tongue, and it occurred to him that it might work similarly on the surface of the eye. After successfully anesthetizing first the eye of a frog, then a guinea pig, and then finally his own eye, Koller wrote up an account of his results and handed them to a colleague to present at an upcoming conference in Heidelberg (Koller was too poor to afford the trip there himself). The public reaction was electric. Decades later, Koller recalled that "knowledge of the new remedy spread quickly, and in looking over the medical and the lay press of the time, one will encounter a perfect flood of communications on cocaine and local anesthesia." On hearing the news, one of the presidents of the British Medical Association asserted: "In the discovery of cocaine, a new era seems to have dawned."

As for Koller, though he certainly achieved international renown from his discovery, the dawn of cocaine's new era coincided with the arrival of a less fortunate interval in his own career. In January 1885, while medical papers were still full of news of his success, he got into an argument with a man named Friedrich Zinner, another surgical intern at Vienna General. Beginning with a technical disagreement over a patient's injured finger, matters escalated until Zinner called Koller an "impudent Jew" (or possibly a "Jewish swine" according to Freud's recollection) and Koller replied by punching Zinner in the face. Both men were medical lieutenants in the army reserve, so Zinner challenged Koller to a duel. When they met five days later, Koller emerged from the duel unharmed, but he left his opponent with two deep wounds. The Vienna public prosecutor was obliged to bring charges against both men, and a criminal case meant that Koller was bound to resign his position at the hospital. He spent the next few years living in the Netherlands before emigrating to New York in 1888. In the United States, he had better luck capitalizing on his renown, opening a thriving ophthalmological practice and becoming the first winner of the Lucien Howe Medal for outstanding achievements in eye medicine.

When news of Koller's discovery had first begun to spread, Freud had jokingly invested his friend with the nickname "Coca Koller." As the 19th century wore on toward the 20th, however, cocaine was destined to dramatically outgrow the relatively specialized applications Koller had envisioned for it.

A Safer Anesthetic

Part of the reason cocaine captured the Victorian imagination so vividly was because of how well it performed in comparison with existing anesthetics. Chloroform had been in use for almost 40 years by the time that cocaine appeared, but doctors and patients alike were still ambivalent about its use in surgery. Accidents and deaths under general anesthesia were comparatively rare but consistent enough that, as the Scottish practitioner William Semple Young observed: "There are many people who dread chloroform so much, that they decline to take it, unless practically coerced."

Cocaine looked like an obvious answer to this problem, at first. "With a solution of cocaine at hand," wrote a reporter for the London newspaper the *St James's Gazette*, "chloroform and ether may be dispensed with." In this light, cocaine appeared to be the ideal anesthetic—safe, effortless, and effective. The discovery of the drug came to be seen as almost epoch-making: It appeared to mark the advent of a new age where modern, technological medicine would sweep away pain and poor health altogether. Reporting on the discovery, a correspondent writing to the *Chambers's Journal* called cocaine "a wonder of the age." The writer continued, "Cocaine has flashed like a meteor before the eyes of the medical world, but, unlike a meteor, its impressions have proved to be enduring."

As cocaine's meteoric ascent continued, its price also began to rise. The more that enterprising medical practitioners rushed to obtain samples of the alkaloid, the more demand began to outstrip supply until, in late 1884, the value of cocaine reached £32 an ounce (around £3,300 or $4,000 an ounce today). In the United States, some suppliers in major cities could ask as much as $300 an ounce. For a while, the white powder was more valuable than gold.

Once supply caught up with demand and the cost of the drug leveled off, cocaine quickly found itself applied to all manner of uses, both exotic and everyday. Outside the operating room, the most common use of cocaine was as a cold and flu remedy. By the 1890s, Burroughs, Wellcome & Co supplied a portable cocaine nasal spray for congestion that was "so small as to be easily carried in the waistcoat pocket." For those who preferred to mix their cold medicines at home, newspapers provided recipes compounded of cocaine, ground coffee, menthol, and powdered sugar, which were finely ground together and "used like an ordinary snuff." Cocaine lozenges were regularly advertised as the best thing an anxious traveler could get for seasickness. The same lozenges were also frequently touted as the ideal treatment for "the sickness of pregnancy." Hay fever and tickly coughs also yielded to tablets of cocaine, while tubes of cocaine toothpaste promised to remedy the pain of toothache and bleeding gums. And for those who suffered from less definitively physical maladies, there were products like "Neurogene": a "compound syrup of cocaine" that offered relief to "Speakers, Singers, Athletes, Business Men, and all who suffer from Brain Fog, or Nervous Debility"—price 2 shillings and 9 pence, or just over £11 ($15) today.

Cocaine became well established in its role as both a technological and fashionable drug *à la mode*. The substance was encircled with an aura of newness and transformative potential.

Wikimedia Commons

Sherlock Holmes was renowned for his intellect and his embrace of science—including his experimentation with cocaine. William Gillette adapted Arthur Conan Doyle's stories into a play in which he starred as the famous detective (*right*) alongside Bruce McRae as Dr. Watson (*left*). This photograph of Holmes preparing to inject cocaine was included in a souvenir album commemorating the initial production in 1899. Within a few years, as the public image of cocaine became associated with addiction rather than with innovation, Doyle had Watson help Holmes kick the habit.

One unforeseen consequence of cocaine was to accelerate the fashion for tattoos. Both Edward VII and the future George V had been tattooed on overseas visits to Jerusalem and Japan, respectively, which sparked off something of a craze for the practice among the British public. Tattooing was a somewhat fraught process by the standards of the time, though. The willingness to endure pain in the interest of nothing more substantial than personal decoration was often thought to betray something coarse—even brutal—in the temperament of the tattooed. But cocaine offered an effective solution to this issue. One paper covering the trend wrote: "Some years ago [tattooing] was a very painful operation, but the discovery of cocaine has made it a painless one." In making tattooing painless, cocaine had also managed to make it seem refined enough for polite society.

Armed with cocaine, a new breed of celebrity tattoo artists began to emerge. One of the most famous was Sutherland MacDonald of Jermyn Street in London, who—when a journalist asked if his clients were required to suffer much pain in the execution of his designs—confidently responded: "Not at all, because I inject cocaine under the skin at the part upon which I am going to operate, and use more cocaine directly the effects of the first injection have passed away." For those desirous of learning the art themselves, it was possible to purchase a full home tattooing kit, comprising "a complete set of tattooing instruments, needles mounted in ivory handles, non-poisonous inks of various colures, and a tiny bottle of cocaine to render the operation painless," all neatly enclosed in a handsome "Russia leather case."

Miracle Cure to Dangerous Drug

In the years after Koller's discovery of the use of cocaine as an anesthetic, the drug had become well established in its role as both a technological and fashionable drug *à la mode*. The substance was encircled with an aura of newness and transformative potential. Journalists rhapsodized over the way in which the drug seemed—like a modern Athena—to have "sprung into existence fully armed": It was at once a vital tool in "the armoury of the modern scientific surgeon" and "the prized possession of millions."

Cocaine even found its way into the hands of one of the most famous fictional characters of the age: Arthur Conan Doyle's Sherlock Holmes. *The Sign of Four* (1890), the second of the Holmes novels, begins with its hero rolling up his shirt sleeve and injecting himself with a "seven per-cent solution" of cocaine. Doyle had trained as a doctor himself and was well aware of cocaine's popular associations with modernity and innovation. For Doyle, giving his character a cocaine habit was a way to quickly convey to his readers that Holmes was a thoroughly modern man—energetic, specialized, and scientifically knowledgeable. Strange as it might seem to us now, more than a century after cocaine's criminalization, the first Victorian reviewers were fascinated by this facet of Holmes's personality. *The Graphic* newspaper thrilled at the detective's "genius and energy." Holmes was, the reviewer enthused, a "first-class" detective, "who must either be engaged in unravelling a first-class mystery, or in consoling himself for the want of one with cocaine."

As time wore on, however, and Doyle's detective became ever more popular, his cocaine use was to become a focus for new anxieties building around the drug. In 1901, John Wyllie, a professor at Doyle's old medical school, the University of Edinburgh, described how he had one day been called to see a sick young man. As Wyllie entered the house, his patient's sister ran to him, crying pitifully: "It's all that horrid book!" Wyllie went on: "Inquiry elicited the fact that the patient's favorite reading was Sherlock Holmes. The young man was in a very low state, and his tell-tale arm was dotted with hypodermic punctures. His admiration for the most popular of paper detectives had betrayed him into the cocaine habit." Wyllie's experiences were published in both popular and medical papers. *The British Medical Journal*'s report wound up with the vague but nevertheless pointed suggestion that authors who cavalierly encouraged their readers into the "baleful spell" of the drug habit might have "much to answer for."

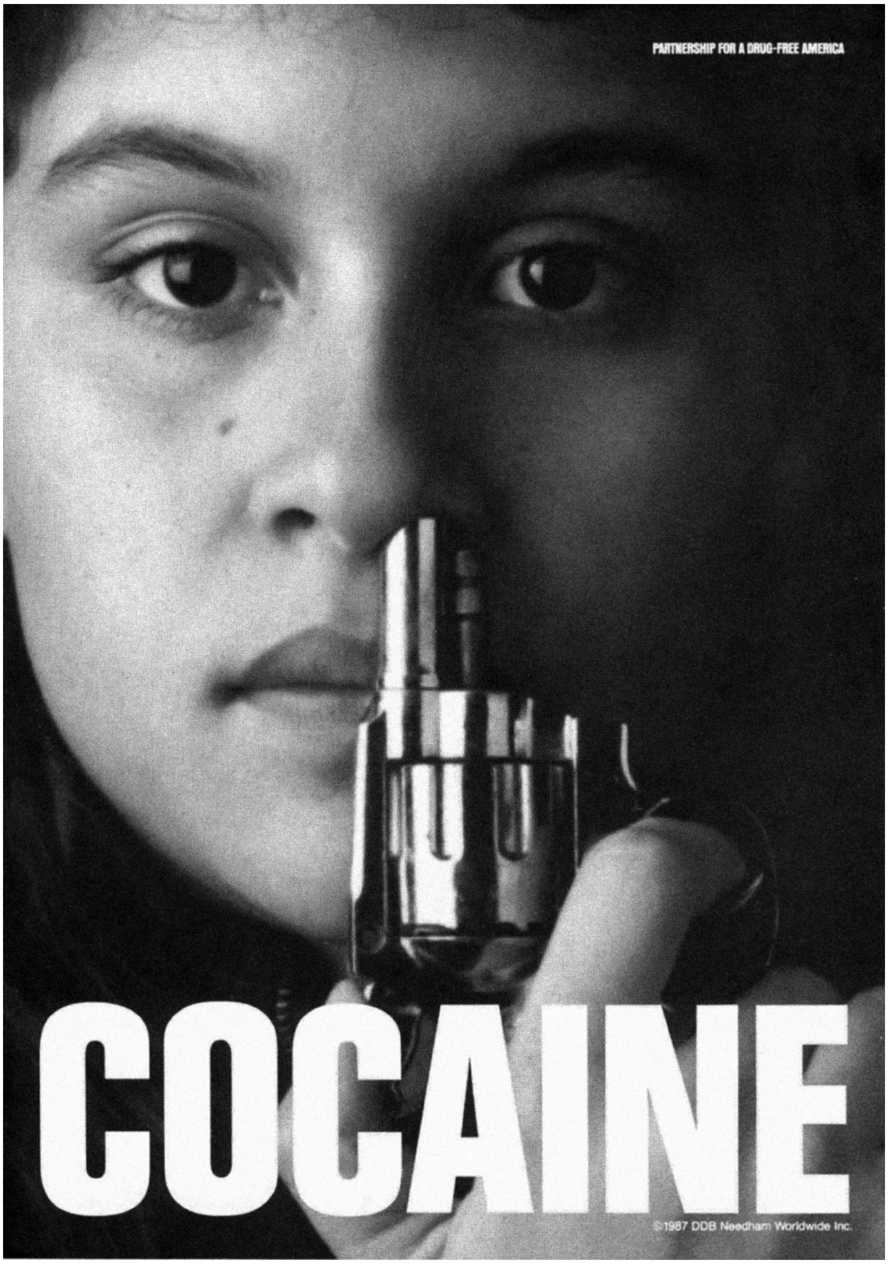
With the coming of the 20th century, popular perceptions of cocaine began to shift in subtle but important ways. With the passage of the years, cocaine was still seen as a technological triumph, but this was balanced against a wider sense of the dangers that might come from its overuse. Writers who had at first been intrigued and excited by Holmes's "seven per-cent solution" were increasingly cautious of being seen to advocate for a "dangerous drug." Doyle decided to dispense with Holmes's cocaine habit for good. "The Adventure of the Missing Three Quarter" (1904) begins with Watson recalling how he had worked "for years" to gradually wean his friend off the "drug mania which had threatened once to check his remarkable career." This passage marks the final word on cocaine in the Holmes canon. Henceforward, the drug was to be emphatically relegated to the detective's past—a tragic and dangerous misadventure from which he had been rescued by Watson's conscientious intervention.

Doyle's reframing of Holmes's cocaine use illustrates the ways in which perceptions of cocaine were changing in the early years of the new century. When Doyle had first conceived of his detective, cocaine had been regarded as a "modern panacea"—tangible proof that science could better the human condition—and its use marked Holmes as a modern, technologically engaged individual. By the early 20th century, though, these associations had begun to shift as the substance became much more directly tied to the threat of addiction and degradation.

As Sherlock Holmes grew more popular, his cocaine use became a focus for new anxieties building around the drug.

Granger

In the mid-20th century, cocaine became associated with danger and criminality. Movies, such as the Italian film *Una lettera all'alba* (1948, released in English as *Cocaine: The Thrill That Kills*), depict the drug as a corrupting force. These films played off of public anxiety about cocaine's spread from white, affluent communities, where its use was seen as scientific and therapeutic, to poor communities of color, where it was associated with drug abuse and addiction.



Contraband Collection/Alamy Stock Photo

In 1987 the Partnership for a Drug Free America released a series of advertisements that equated cocaine use to suicide. This anti-drug campaign marked a reversal of attitudes about cocaine from miracle drug to health and safety threat.

Cocaine as a Societal Threat

Legislative controls on cocaine first started to appear in the 1910s. In Britain, the Defence of the Realm Act was introduced in response to the demands of the First World War, and in 1916 it was used to restrict the sale of the drug to specific "authorized persons" such as doctors, surgeons, and dentists. Under the act, cocaine could now be obtained by members of the public only with a doctor's prescription. These restrictions remained in place until after the war, when they were permanently codified into law through the Dangerous Drugs Act of 1920. In part, these laws were a response to the genuine dangers of cocaine addiction—to the risks that cocaine might pose to naive or overenthusiastic individuals like Wyllie's unfortunate patient. But they also reflected the broader social and racist prejudices of their time.

As cocaine use became more widespread, it began to permeate out of the relatively closed circle of the affluent, white middle and upper classes. For a white, wealthy, socially established man like William A. Hammond to experiment with cocaine and its pleasures was acceptable. At worst, it might be a little comical. The same experiences could be made to seem much more threatening when they were taken up by the poor, by women, or by people of color.

In 1914, *The Lancet* and various popular newspapers in the United Kingdom republished an article by the doctor Edward Huntington Williams on cocaine in the southern United States. Williams claimed that any Black man who took up the cocaine habit was "absolutely beyond redemption. His whole nature is changed by the habit. Sexual desires are increased and perverted; peaceful men become quarrelsome and timid ones courageous." Sidney Felstead, author of *The Underworld of London* (1923), claimed to be similarly appalled by how often "some pleasure-sated girl dies from an overdose of cocaine or morphia, supplied to her by some black or yellow parasite."

These remarks illustrate the double standard that developed around cocaine in the period leading up to its criminalization. The sense of newness, of transcendent modernity that cocaine imparted to William A. Hammond or to Sherlock Holmes, was not extended to everyone. The stigmatization of the drug as an agent of danger and perversity reflected already existing prejudices against those minorities, and the laws that were enacted to control it reflected (at least in part) the desire to control those same people.

Over the decades, cocaine had transitioned from a wonder of the newly technological and industrial Victorian age to a frightening and corrupting source of addiction. The story of cocaine illustrates not only how much our perceptions of specific drugs can shift over time but how readily drugs can capture and condense our emotions. Cocaine was always a drug peculiarly surrounded by fantasies: hopes, fears, optimism, and anxiety. Describing its history reveals the degree to which our fantasies and fears about drugs shape, and are shaped by, our fantasies about their users.

Douglas Small is a historian of medicine in the United Kingdom. He is the author of Cocaine, Literature, and Culture, 1876–1930 *(Bloomsbury, 2024). This article originally appeared on* Aeon *(aeon.co). X (formerly Twitter): @DrDouglasSmall*

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 online, which explores how science intersects with other areas of knowledge, entertainment, and society.*

ALSO IN THIS ISSUE

THE UNIVERSE IN VERSE: 15 Portals to Wonder Through Science & Poetry. By Maria Popova.
page 58

ONLINE

On our *Science Culture* blog:
www.americanscientist.org/blogs/science-culture

A Window into the Origins of Brain Surgery

Margot Kelly-Hedrick reviews Theodore H. Schwartz's *Gray Matters: A Biography of Brain Surgery*.

Some Assembly Required: A Bold New Vision of Life

Michael L. Wong reviews Sara Imari Walker's *Life as No One Knows It: The Physics of Life's Emergence*.

Survivors of the Sea

Melissa Cristina Márquez

THE SECRET HISTORY OF SHARKS: The Rise of the Ocean's Most Fearsome Predators. John Long. 480 pp. Ballantine Books, 2024. $35.00.

For anyone fascinated by the mysteries hidden beneath the ocean waves and the ancient creatures that once roamed our oceans, John Long's *The Secret History of Sharks: The Rise of the Ocean's Most Fearsome Predators* offers an exhilarating journey through time. Long brings to life the incredible story of sharks: predators that have not only survived but thrived for millions of years.

Long, a world-renowned paleontologist, writes, "The origin of sharks is one of the last great unsolved mysteries in the five-hundred-million-year-old evolution of the backboned animals we call 'vertebrates' (fish, amphibians, birds, reptiles, and mammals)." He goes on to meticulously trace the evolutionary journey of these incredible creatures, guiding the reader through millions of years with a narrative that is scientifically rigorous, introducing us to the earliest protosharks and eventually to the formidable predators we know (and love or fear) today.

Although the Megalodon is a natural climax to the shark's history, some of the most intriguing parts of this book were the detailed portrayals of various lesser-known sharks and their evolutionary branches. Long describes astonishing creatures such as collared catsharks, who have the "amazing ability to change their body color to match the seabed to blend in, to hide from predators or ambush prey" or the extinct genus *Ptychodus*, whose "jaws were lined with hundreds of mostly wrinkly, dome-shaped crushing teeth." When describing various types of sharks that have long since gone extinct, there are clear distinctions between what are established facts, informed speculation, and Long's narrative license. This transparency is a departure from some other popular science books, where the lines between fact and the author's speculation are often blurred.

Long presents recent, cutting-edge science made possible by new technological advancements, which makes the book feel current while simultaneously underscoring the dynamic nature of paleontological research. He discusses how these technologies have allowed scientists to reassess existing fossil collections, leading to new discoveries and interpretations. For example, through techniques such as computed tomography (CT) scans and advanced imaging, scientists have captured fine details that have previously eluded them, allowing them to create more precise digital models, which have facilitated detailed analyses and comparisons with today's shark species. This advance has led to the identification of new species and a deeper understanding of the evolutionary pathways that resulted in the diversity of sharks we see today.

Interspersed throughout the book are biographical sketches of scientific figures in shark science, including anecdotes that bring important people in the field to life and provide quirky and interesting details about them to the reader. For example, Chas Eastman, who "expanded our knowledge of buzz saw sharks, the Solnhofen sharks from Germany, the great Cretaceous lamniforms, the complete Eocene sharks and fishes of Bolca, Italy, and Cenozoic sharks from several U.S. and other localities around the world," was indicted for the murder of his brother-in-law, did significant paleontological research while behind bars waiting for

Jonas Gruhlke

The grey reef shark (*Carcharhinus amblyrhynchos*) is a common reef shark in the Indo-Pacific. They can be identified by the dark area on the edge of the caudal fin. They're social creatures, often curious about divers and congregating in groups during the daytime, and at night hunting on their own.

trial, and ended up winning the case and was released. Reading details often left out of textbooks and memoirs of these scientific giants is refreshing, especially during a time when scientists are being asked to take a close look at their predecessors' prejudicial viewpoints. Long's recounting of his own experiences and relationships in the research community adds another layer of intimacy. Although some might perceive this information as boasting, it can also be seen as justifiable pride in his contributions and appreciation of his connections in the field.

Along with photographs and diagrams that complement the text, readers will find themselves immersed in the life of a paleontologist, sharing Long's experiences on global expeditions, from the fossil-rich beds of North America to surviving emus in Australia to the remote cliffs of Antarctica. Each adventure is pulse-pounding, giving readers a front row seat to the discoveries that have reshaped our understanding of shark evolution.

Parts of the book are much denser than others, because of the technical depth of the research and writing. The section on the Cretaceous period, for instance, is densely packed with biological and biographical information; all of that is fascinating, but it can feel overwhelming if the reader does not have a background in paleontology. However, the book largely succeeds, thanks to Long's ability to translate complex scientific findings into understandable language. This text is complemented by a thorough index, which is particularly helpful for those looking to delve deeper into specific topics. Long peppers the text with unexpected language that helps make scientific literature entertaining, such as "how placoderms made sharks sexy" and "sharky-sharks." With an engaging writing style, his ability to weave personal anecdotes and experiences into the broader narrative highlights his enthusiasm for the subject matter. This enthusiasm is infectious, making *The Secret History of Sharks* a page-turner even for less shark-obsessed readers.

This book is not just a recount of the past; it is a poignant reminder of the current threats these survivors face. Long brings attention to the numerous urgent threats facing sharks today, from overfishing and habitat destruction to the wide-reaching effects of climate change. He passionately argues for sharks' conservation, highlighting that the survival of sharks is intricately tied to the health of our oceans and the balance of marine ecosystems, and reminding us that saving these apex predators isn't just about protecting one species—it's about safeguarding our planet's most vital ecosystem. But Long doesn't stop at sharks' ecological importance; he explores the surprising ways they contribute to human advancement. From the potential antibiotic applications of their unique immune systems to the breakthroughs in medical research their study has enabled, sharks are providing novel insights into fields far beyond marine biology.

The Secret History of Sharks offers more than just a journey through millions of years of shark evolution; it is a call to action for a deeper appreciation and preservation of these extraordinary predators. By showcasing their adaptability, resilience, and even their unlikely contributions to human innovation, Long underscores that sharks are not just relics of the past but invaluable allies in our planet's future. This accessible and engaging book is not only an enlightening read for those curious about marine life, but also a compelling case for why their survival is essential to our own.

Melissa Cristina Márquez is a marine biologist, science communicator, and shark researcher known for her work in marine conservation and her dedication to increasing diversity in STEM. She engages global audiences through public speaking, writing, and educational programs, inspiring others to protect our oceans.

The Poetry of Science

THE UNIVERSE IN VERSE: 15 Portals to Wonder Through Science & Poetry. Maria Popova and Ofra Amit. 112 pp. Storey Publishing, 2024. $22.00.

The relationship between writers and science is fascinating. Writers gain inspiration, solace, and a sense of wonder from the natural world, as well as areas of science such as physics, biology, mathematics, and astronomy. At the same time, scientific ideas can be illuminated in unexpected ways through poetry. Poetry can be especially adept at capturing aspects of science, with unconventional formatting and styles that mirror the complexity of the sciences.

One collection of poetry that was published in 2024 highlights this relationship in a variety of ways: *The Universe in Verse: 15 Portals to Wonder Through Science & Poetry*, edited by Maria Popova and illustrated by Ofra Amit, pairs poems with brief glimpses into different science histories.

Poets highlighted in *The Universe in Verse* include contemporary and classic poets such as Sylvia Plath, Emily Dickinson, Maya Angelou, Adrienne Rich, and Marie Howe. The science histories touch on topics such as Stephen Hawking and singularity, the concept of entropy, radioactivity and Marie Curie, and the cosmos.

The first poem excerpted here, "We Are Listening," by Diane Ackerman, is paired with a brief discussion of Carl Sagan, the discs known as the Golden Records encoded with music and images and sent into the cosmos aboard the two *Voyager* spacecraft, and Dr. Jill Tarter, an astronomer known for her work in the search for extraterrestrial intelligence/life (known as SETI). Sagan supported SETI research, and in the film adaptation of his novel *Contact*, the lead role, astronomer Ellie Arroway, is largely modeled after Tarter.

The second poem is Howard Nemerov's "Figures of Thought." In the book, it is paired with a short exploration of mathematician Emmy Noether, the first woman to give the plenary address at the International Congress of Mathematicians in 1932 and the namesake of the 1918 theorem proving that conservation laws rely on symmetry.

Braiding poetry, science, and history, the book lives up to its subtitle, inspiring wonder and piquing curiosity in its readers.

WE ARE LISTENING

Diane Ackerman

I.
As our metal eyes wake to absolute night,
where whispers fly
from the beginning of time, we cup our ears to
the heavens.
We are listening

on the volcanic lips of Flagstaff and in the fields
beyond Boston, in a great array that blooms
like coral from the desert floor, on highwire
webs patrolled
by computer spiders in Puerto Rico.

We are listening for a sound beyond us,
beyond sound,

searching for a lighthouse
in the breakwaters of our uncertainty,
an electronic murmur
a bright, fragile I am.

Small as tree frogs
staking out one end
of an endless swamp,
we are listening
through the longest night
we imagine, which dawns
between the life and time of stars.

© Ofra Amit

II.
Our voice trembles
with its own electric,
we who mood like iguanas
we who breathe sleep
for a third of our lives,
we who heat food
to the steaminess of fresh prey,
then feast with such baroque
good manners it grows cold.

In mind gardens
and on real verandas
we are listening,
rapt among the Persian lilacs
and the crickets,
while radio telescopes
roll their heads, as if in anguish.

With our scurrying minds
and our lidless will
and our lank, floppy bodies
and our galloping yens
and our deep, cosmic loneliness
and our starboard hearts
where love careens,
we are listening, the small bipeds
with the giant dreams.

FIGURES OF THOUGHT

Howard Nemerov

To lay the logarithmic spiral on
Sea-shell and leaf alike, and see it fit,
To watch the same idea work itself out
In the fighter pilot's steepening, tightening turn
Onto his target, setting up the kill,
And in the flight of certain wall-eyed bugs
Who cannot see to fly straight into death
But have to cast their sidelong glance at it
And come but cranking to the candle's flame—

How secret that is, and how privileged
One feels to find the same necessity
Ciphered in forms diverse and otherwise
Without kinship—that is the beautiful
In Nature as in art, not obvious,
Not inaccessible, but just between.

It may diminish some our dry delight
To wonder if everything we are and do
Lies subject to some little law like that;
Hidden in nature, but not deeply so.

© Ofra Amit

Statement of ownership, management and circulation (required by 39 U.S.C. 3685). 1. Publication title: *American Scientist*. 2. Publication number: 2324-0. 3. Filing date: October 1, 2024. 4. Issue frequency: Bimonthly. 5. No. of issues published annually: 6. Annual subscription price: $30. 7. Complete mailing address of known office of publication: P. O. Box 13975, Research Triangle Park, NC 27709-3975. 8. Complete mailing address of headquarters or general business office of publisher: P.O. Box 13975, Research Triangle Park, NC 27709-3975. 9. Full names and complete mailing addresses of publisher, editor, and managing editor: Jamie Vernon, publisher, P.O. Box 13975, Research Triangle Park, NC 27709-3975; Fenella Saunders, editor, P.O. Box 13975, Research Triangle Park, NC 27709-3975; Stacey Lutkoski, managing editor, P.O. Box 13975, Research Triangle Park, NC 27709-3975; 10. Owner: Sigma Xi, The Scientific Research Honor Society, P.O. Box 13975, Research Triangle Park, NC 27709-3975. 11. Known bondholders, mortgagees, and other security holders owning or holding 1 percent or more of total amount of bonds, mortgages, or other securities: None. 12. The purpose, function, and nonprofit status of this organization and the exempt status for Federal income tax purposes: Has not changed during preceding 12 months. 13. Publication Title: *American Scientist*. 14. Issue Date for Circulation Data: September–October 2023—July–August 2024. 15. Extent and nature of circulation: science. A. Total no. copies: Average no. copies each issue during preceding 12 months, 41,474; no. copies of single issue published nearest to filing date, 42,479. B. Paid circulation: B1. Mailed outside-county paid subscriptions stated on PS Form 3541: average no. copies each issue during preceding 12 months, 24,108; no. copies of single issue published nearest to filing date, 25,199. B2. Mailed in-county paid subscriptions: average no. copies each issue during preceding 12 months, 0; actual no. copies of single issue published nearest to filing date, 0. B3. Paid distribution outside the mails including sales through dealers and carriers, street vendors, counter sales, and other paid distribution outside USPS: average no. copies each issue during preceding 12 months, 4,211; no. copies of single issue published nearest to filing date, 3,800. B4. Paid distribution by other classes of mail through the USPS: average no. copies each issue during preceding 12 months, 0; no. copies of single issue published nearest to filing date, 0. C. Total paid distribution: average no. copies each issue during preceding 12 months, 28,320; no. copies of single issue published nearest to filing date, 28,999. D. Free or nominal rate distribution: D1. Free or nominal rate outside-county copies as stated on PS Form 3541: average no. copies each issue during preceding 12 months, 259; no. copies of single issue published nearest to filing date, 260. D2. Free or nominal rate in-county copies as stated on PS Form 3541: average no. copies each issue during preceding 12 months, 0; actual no. copies of single issue published nearest to filing date, 0. D3. Free or nominal rate copies mailed at other classes mailed through the USPS: average no. copies each issue during preceding 12 months, 0; actual no. copies of single issue published nearest to filing date, 0. D4. Free or nominal rate outside the mail: average no. copies each issue during preceding 12 months, 271; no. copies of single issue published nearest to filing date, 271. E. Total free or nominal rate distribution: average no. copies each issue during preceding 12 months, 530 no. copies of single issue published nearest to filing date, 531. F. Total distribution: average no. copies each issue during preceding 12 months, 28,850; no. copies of single issue published nearest to filing date, 29,530. G. Copies not distributed: average no. copies each issue during preceding 12 months, 12,625; no. copies of single issue published nearest to filing date, 12,949. H. Total: average no. copies each issue during preceding 12 months, 41,475; no. copies of single issue published nearest to filing date, 42,479. I. Percent paid: average no. copies each issue during preceding 12 months, 98.2 percent; no. copies of single issue published nearest to filing date, 98.2 percent. 16. Total circulation includes electronic copies: no. copies of single issue published nearest to filing date: A. paid electronic copies, 201. B. Total paid print copies + paid electronic copies, 29,200. C. Total print distribution + paid electronic copies, 29,731. D. Percent paid: 98.21 percent. 50% of all distributed copies (electronic and print) are paid above a nominal price.

Sigma Xi Distinguished Lecturers 2025–2026

For the 87th year, Sigma Xi presents its panel of Distinguished Lecturers as an opportunity for chapters to host visits from outstanding individuals who are at the leading edge of science. These experts have agreed to visit chapters and share their insights and excitement on the topics detailed below.

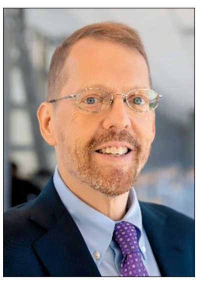
The Distinguished Lecturers are available from July 1, 2025, to June 30, 2026. Each speaker has consented to a modest honorarium together with full payment of travel costs and subsistence.

Local chapters may apply for subsidies to support expenses related to hosting a Distinguished Lecturer. Subsidy applications must be submitted online by March 1, 2025, for funds to be available the next fiscal year.

Additional support for the program comes from the American Meteorological Society. Lecturer biographies, contact information, and additional details can be found online at sigmaxi.org/lectureships or by sending an email to lectureships@sigmaxi.org.

—Committee on Lectureships

Chapter Subsidy Application Deadline:
March 1, 2025
sigmaxi.org/lectureships



David A. Bader, Distinguished Professor, New Jersey Institute of Technology

Solving Global Grand Challenges with High Performance Data Analytics (P, G, S) • Predictive Analysis from Massive Knowledge Graphs (P, G, S) • Interactive Data Science at Scale (P, G, S)

Matthew Baum, Marvin Kalb Professor of Global Communications and Professor of Public Policy, Harvard University

Misinformation: How Big a Problem and What Can Be Done? (G, S) • Soft News, Satire, and the Blending of Politics and Entertainment: Why It Matters (P, G, S) • Media Bias: Perceptions, Reality, Consequences (G, S)

Joseph J. Biernacki, Professor Emeritus, University Distinguished Faculty Fellow, Tennessee Technological University

What Do Artificial Intelligence, Synthetic Live-Chemistry and Nuclear Fusion Have to Do with Portland Cement (P, G, S) • The Looming Housing Crisis and Hope on the Horizon (P, G, S) • Who's the Biggest Maker of Them All? (P, G)

Lynn Cominsky, Professor, Physics and Astronomy, Director, EdEon STEM Learning, Sonoma State University

Gravitational Waves: The Discovery That Won the 2017 Nobel Prize (P, G) • High Energy Visions of the Universe (P, G) • Science of War and Peace (P, G)

James N. Druckman, Martin Brewer Anderson Professor, University of Rochester

Partisan Hostility and American Democracy (P, G, S) • (Dis)trust in America (P, G, S) • The Polarization and Politicization of Trust in Scientists (P, G, S)

Eduardo Fernandez-Duque, Professor, Department of Anthropology and School of the Environment, Yale University

Fatherhood: From Molecules to Society (P, G) • The Evolution of Pair-Bonds and Monogamy (P, G) • Cause and Effect in Biological Anthropology (G, S)

P (Public), G (General), S (Specialized) **Details available at sigmaxi.org/lectureships**

Efi Foufoula-Georgiou, American Meteorological Society, Distinguished Professor, Civil and Environmental Engineering, University of California, Irvine

Precipitation in the Earth System: Global Estimation, Precipitation Extremes and Climate Change (P, G, S) • A Life in Science: A Few Lessons Learned and My Professional Journey (P, G) • The Challenge of Rainfall Estimation and Prediction across Scales: Learning from Patterns (P, G, S)

John R. Jungck, Professor of Biological Sciences and Mathematical Sciences, Inaugural Fellow Honors College, Associate Director, Institute for Transforming University Education, Delaware Environmental Institute; Computational Biology and Bioinformatics, Delaware Biotechnology Institute

Mathematics Saves Lives! (G) • Citizen University (G) • Biomimetic Design Principles of Self-Assembling, Self-Folding, and Origami (G)

Haagen Klaus, Professor of Anthropology, George Mason University

Human Sacrifice in the Ancient Andes: Connecting Skeletal Trauma, Archaeology, and the Meanings of Ritual Killing (P, G, S) • Surfacing from the Wake of Contact: A Bioarchaeology of Indigenous Creativity, Resistance, Resilience, and Suffering in Colonial Peru (P, G, S) • Ancient Skeletons and Violence: A Global Reconstruction of the Origins and Causes of Human Conflict in the Past and Present (P, G, S)

Dante Lauretta, Regents Professor of Planetary Science and Cosmochemistry, University of Arizona

Life in the Cosmos: The Search for Biology in the Universe (P) • OSIRIS-REx: NASA's Sample Return Mission from Asteroid Bennu (G) • Journeys on the Asteroid Frontier: The Engineering behind NASA's OSIRIS-REx Asteroid Sample Return Mission (S)

June Pilcher, Alumni Distinguished Professor, Clemson University

Science behind Mindfulness Practices and Meta-Awareness (P, G, S) • Lifestyle Matters: Impact of Sleep and Physical Activity (P, G, S) • Diving into Difficult Discourse (P, G)

Anne Savage, Executive Director, Proyecto Tití, Inc.

Proyecto Tití: Saving Colombia's Critically Endangered Cotton-top Tamarin (P, G) • Teens, Tamarins, and Teamwork: Successful Efforts to Engage Communities in Conserving Cotton-top Tamarins in Colombia (P, G) • Cotton-top Tamarins: Studies in Captive Care Have Informed Conservation Actions (P, G)

Karen C. Seto, Frederick C. Hixon Professor of Geography and Urbanization Science, Yale University

Urbanization in the 21st Century: Problem or Panacea for the Environment? (P, G) • How Will Urbanization Change Food Systems? (P, G) • Are Cities the Solution to Climate Change? (P, G) • Revealing Patterns of Urbanization with Remote Sensing (S)

Ramteen Sioshansi, Professor, Department of Integrated Systems Engineering, Department of Electrical and Computer Engineering; Director, EmPOWERment National Science Foundation Research Traineeship Program, Associate Fellow, Center for Automotive Research, The Ohio State University

Technology Pathways to and Economic and Technical Challenges with Decarbonizing Electricity Systems (P, G, S) • How Regulatory Choices Impact the Sustainability, Reliability, and Resilience of Energy Supply (P, G, S)

Karen Strier, Vilas Research Professor and Irven DeVore Professor of Anthropology, University of Wisconsin–Madison

Saving the World's Most Peaceful Primate (P, G) • Primates and Conservation in a Rapidly Changing World (P, G, S) • Primate Behavioral Flexibility and the Limits of Resilience (P, G, S)

P (Public), G (General), S (Specialized)

Details available at sigmaxi.org/lectureships

Sigma Xi Today

A NEWSLETTER OF SIGMA XI, THE SCIENTIFIC RESEARCH HONOR SOCIETY

Call for Nominations

Sigma Xi, The Scientific Research Honor Society, is seeking nominations for qualified candidates to fill positions for President-elect, Board of Directors, Committee on Nominations, and Associate Directors for representation of regions and constituencies. The Board of Directors is principally responsible for managing the activities, property, and affairs of the Society in accordance with the policies established by the Assembly of Delegates. Sigma Xi seeks diverse and inclusive participation in all its elected and appointed positions. Sigma Xi's elected positions are voluntary.

Nominations should be submitted to **elections@sigmaxi.org**. The submission deadline is April 1, 2025, for president-elect candidates and June 30, 2025, for all other positions. All positions carry three-year terms. Beginning July 1, 2025, president-elect will serve one year each as president-elect, president, and immediate past-president, concluding on June 30, 2028. Active full members of Sigma Xi are eligible to run for office. An inactive member may become active at any time through payment of current dues. Self-nominations are welcomed.

The election will take place electronically by ballot immediately following the November 2025 International Forum on Research Excellence (IFoRE). Please visit **sigmaxi.org/elections25** to view the lists of duties and responsibilities for each position.

Sigma Xi Today is managed by **Jason Papagan** and designed by **Chao Hui Tu.**

From the President

The Call for Collaboration

In today's increasingly technology-driven world, we face more complex problems than ever, including climate change, global pandemics, and infrastructure deterioration, just to name a few. Research collaboration, we all now recognize, is the best way to address problems that cannot be tackled by a single discipline, institution, or researcher. In today's research landscape, collaboration is not just beneficial—it is essential.

Collaboration in scientific research can be interdisciplinary, multidisciplinary, or transdisciplinary. In interdisciplinary collaboration, researchers from different fields work closely together to integrate their expertise in a way that leads to new perspectives. Multidisciplinary research also involves collaboration between different disciplines, but where each retains its methodology while contributing to a common goal. Transdisciplinary collaboration goes beyond disciplinary boundaries, where researchers and nonacademic stakeholders, such as policymakers or industry leaders, cocreate knowledge and solutions.

To support this goal of greater collaboration, I see several exciting areas of opportunity:

- Resource sharing: Research, especially in complex fields like biomedical sciences and space exploration, often requires costly and sophisticated equipment, as well as extensive datasets.
- Diverse perspectives: Collaboration brings together researchers from different intellectual, experiential, and social backgrounds.
- Funding opportunities: Many funding bodies, including government agencies and private foundations, now prioritize or require collaborative projects.
- Training and development: Collaborations offer valuable opportunities for mentoring and skills development, especially for early-career researchers.

Going forward, we must continue actively engaging in building strong networks, attending conferences, joining professional societies, and utilizing online platforms. Sigma Xi membership and its IFoRE conference are excellent venues for this. We need to embrace new technologies like cloud-based platforms, user facilities, and collaborative software. We must involve institutions and governments in supporting research collaboration by funding collaborative projects, establishing frameworks for intellectual property sharing, and creating policies that encourage cross-border partnerships.

Please join me in this call to actively seek out new opportunities for collaboration, to break down traditional silos, and to be open to diverse ideas and perspectives. Only by working together can the scientific community achieve the breakthroughs necessary to ensure a better future for all.

Kathy Lu

Announcing the 2024 Cohort of Sigma Xi Fellows

Sigma Xi is proud to announce the 2024 cohort of Sigma Xi Fellows. They were recently honored on November 16 during the third annual International Forum on Research Excellence (IFoRE) in Washington, DC.

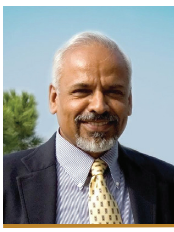
The Fellow of Sigma Xi distinction is awarded on a competitive basis to members who have been recognized by their peers. Fellows must be active (dues-paying) full members for the last 10 years continuously, or life members, with distinguished service to Sigma Xi and outstanding contributions to the scientific enterprise.

Learn more about the 2024 Fellows and how to nominate members for future cohorts by visiting **sigmaxi.org/fellows**.

Lynn Cominsky
Sonoma State University

Sonya Smith
Howard University

Nicholas Donofrio
IBM



Katepalli Sreenivasan
New York University

Sylvia Earle
National Geographic Society

Angela Hight Walker
National Institute of Standards and Technology

Eric Shirley
National Institute of Standards and Technology

Peter Kurzhals
Posthumous

Matthew Traum
University of Florida

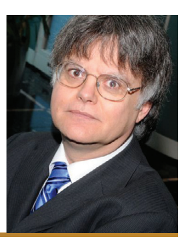
William D. Nordhaus
Yale University

Robert T. Pennock
Michigan State University

Keivan Stassun
Vanderbilt University

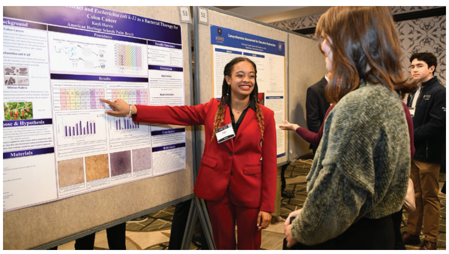
Donna Weistrop
University of Nevada Las Vegas

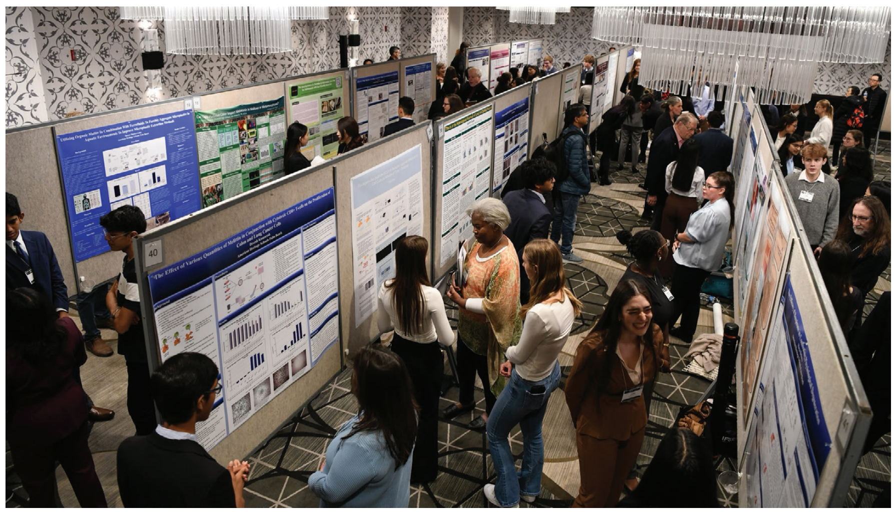
Peter Boily
Inforex, Inc.



Walter E. Massey
University of Chicago

C. Kumar Patel
University of California, Los Angeles





IFoRE
POWERED BY SIGMA XI

Sigma Xi's annual conference, the International Forum on Research Excellence (IFoRE), was held November 14–17 in Washington, DC. Scientific minds spanning multiple generations and disciplines converged to share research, make connections, celebrate awards, and discuss emerging fields such as inclusion in STEM education, artificial intelligence, and science policy. Highlights included five award-winning keynote speakers, over 200 student research presentations, Friday night networking party, Sigma Xi Fellows ceremony, museum tours at the Smithsonian, and over 50 dynamic plenary and breakout sessions. Thanks to the 400+ attendees who brought energy, ideas, and inspiration to the event. We look forward to seeing everyone in November 2025!