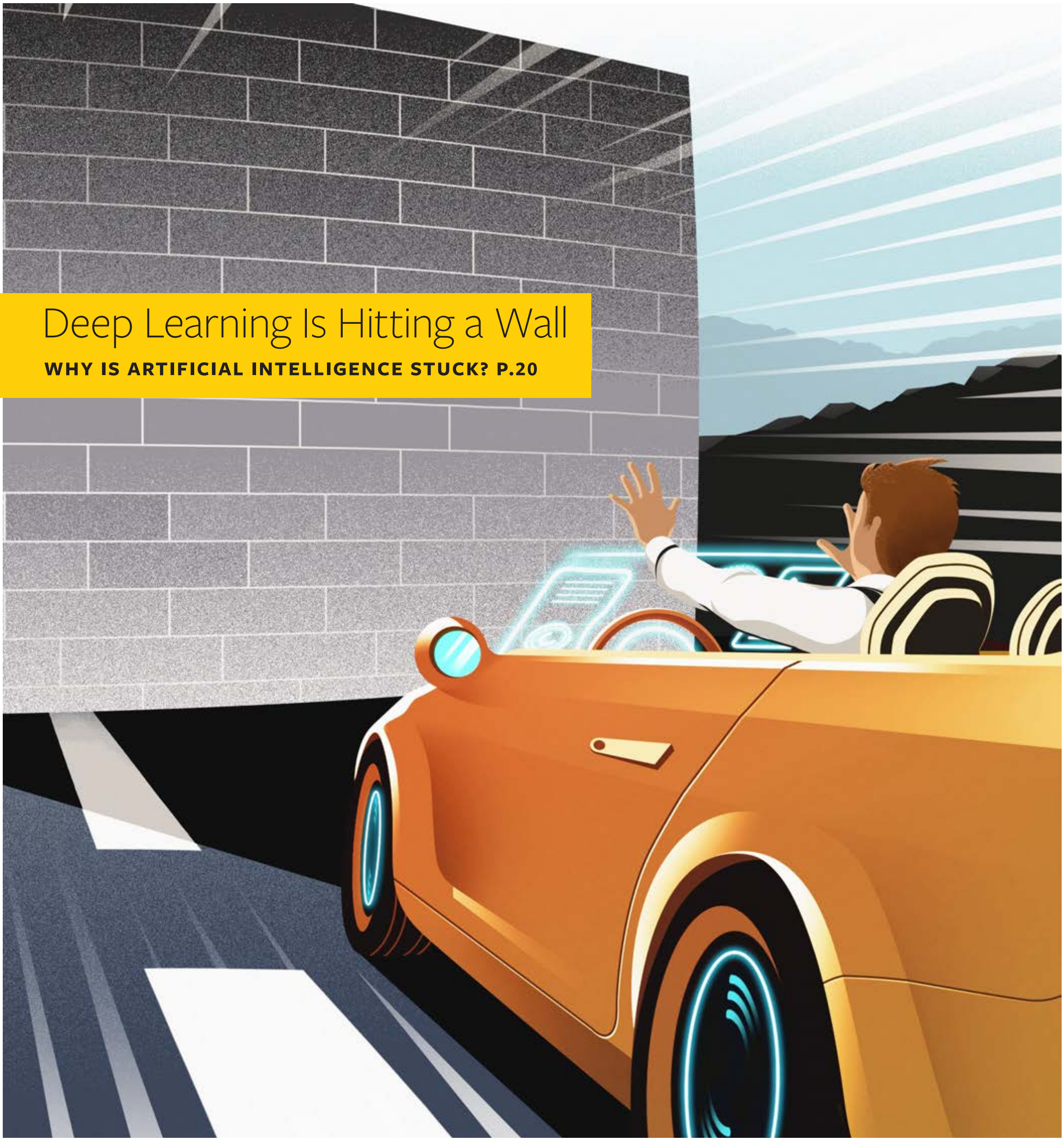
NAUTILUS

WHEN GRIEF WON’T LET GO
Struggling with a death in the family

BREAKING UP WITH QUANTUM THEORY
A physicist gets back to basics

ANIMALS HAVE MORALS, TOO
Why all creatures sense right and wrong



Deep Learning Is Hitting a Wall

WHY IS ARTIFICIAL INTELLIGENCE STUCK? **P.20**

ISSUE 43 Science Connected **NAUTIL.US**

ISSUE 43

Contents

COVER Mark Belan | artscistudios.com

The Great Narrative, the new book by Klaus Schwab and Thierry Malleret, encapsulates the Davos Vision, and explores how we can shape a constructive, common narrative for the future. It is based on interviews with 50 of the world's most influential thinkers, scientists, and economists, and the successor to the worldwide best-seller, "The Great Reset".

https://wef.ch/3fwApfk

Features

Published by:

NAUTILUS Next

CONTACT

360 W 36th Street, 7S
New York, NY 10018

info@nautil.us

SUBSCRIPTIONS

Visit subscribe.nautil.us

Science Is Life

BY KEVIN BERGER

I **LOVE THE PARADOX** of science. The more it explains, the more mysteries it reveals. Those mysteries present further problems to be solved. It's a brilliant form of progress. Which is heartening to realize. Science is life. It tells us there's always more to see, more to learn. Pointedly that also applies to us, people, the oddest things on the planet.

I was thinking about this in context of my interview with Alva Noe, a writer and professor of philosophy at the University of California, Berkeley, who gets the last word in this issue. We were discussing how neuroscience narrows the focus on what makes humans do the things they do. The notion that neurobiology directs our every move, so popular in self-help brain books, drives Noe around the bend. Seeing ourselves in the frame of neuroscience is a crime, like watching *Dune* on a cell phone.

I respect his point. Describing ourselves in the language of neurobiology has got to be one of the stranger trends in our medical age. I'm depressed because my serotonin levels are low. I cried in the movie because the grieving mother activated my mirror neurons. The dopamine boost from my run left me feeling elated. We no doubt say these things because life is chaos and we strive to corral it any way we can to not feel insane. The language of neuroscience is the latest psychological balm.

Noe and I were discussing a trend in neuroscience called neuroaesthetics, which employs brain biology to explain what it is about art—painting, film, music—that turns us on. Explaining what art means to us in the language of biology is reductive.

Human life is "regimented, it is habitual," Noe says. "Whether at the cellular level, the organismic level, or the cultural level, there's a way in which we're all slaves to the ways we need to be, to live, to get along." Art, though, is the spoke in the wheel of organization. "Art has an emancipatory role in our lives. Art makes it possible for us to be different from the ways we already are."

Science, though, isn't the antithesis to art. Science too breaks up regimented thinking, showing us that things, including us, can be different than we already are. And that can go for neuroscience. In this issue, in "What's So Hard About Understanding Consciousness?," renowned neuroscientist Antonio Damasio talks about the enlightening progress of science. Famously the "hard problem of consciousness" states the biological processes of the brain can't capture the ineffable sense we have of the blueness of a Montana sky, the redness of a Washington apple.

"The hard problem makes it look like consciousness is impossible to solve," Damasio says. "It doesn't give you any out. Every bit of evidence we have is that the mysteries of the universe have been gradually solved by science. I don't see why consciousness should be any different."

Understanding the mind is still an adventure into new frontiers. What lies beyond conventional thinking is a theme that streams through this issue's stories about artificial intelligence, grief, quantum theory, animal emotions, and opioids. Good science writing is like art. It shows us there's always more to the picture.

Global Fishing Watch
MARINE MANAGER

DONA BERTARELLI
PHILANTHROPY

«Global Fishing Watch Marine Manager is part of the solution to improve ocean protection and health, thereby ensuring food security and livelihoods for the billions who rely on it. By offering a cutting-edge technology, this portal hosts interactive data on human-use activity, ocean conditions and biology in near real time, supporting marine spatial planning, marine protected area management and scientific research.»

- Dona Bertarelli

2022 SANTA FE INSTITUTE COMMUNITY LECTURES

THE SANTA FE INSTITUTE COMMUNITY LECTURES

bring leading thinkers to Santa Fe's Lensic Performing Arts Center to explore the most alluring questions in science, and to address the complex issues that face our species and our planet.

To stream these lectures live, subscribe to our YouTube channel at youtube.com/user/santafeinst

To attend in person, reserve your free tickets at www.lensic.org

THE 2022 LINEUP

Tuesday, March 22 7:30 pm
SARA WALKER
Arizona State University; Santa Fe Institute
Recognizing the Alien in Us

Tuesday, May 24 7:30 pm
STUART FIRESTEIN
Columbia Univ.; Santa Fe Institute Fractal Faculty
Ignorance, Failure, Uncertainty and the Optimism of Science

Tuesday, July 26 7:30 pm
SEAN CARROLL
CalTech; Santa Fe Institute Fractal Faculty
The Many Worlds of Quantum Mechanics

Tuesday, August 2 7:30 pm
ANDREA WULF
Santa Fe Institute Miller Scholar
The Invention of Nature: Alexander von Humboldt's New World

Tues. & Wed., Sept. 20 & 21 7:30 pm
STANISLAW ULAM MEMORIAL LECTURES
STEVEN STROGATZ
Cornell University
Mathematical Stories
Part I: The Story of Calculus
Part II: The Story of Sync

Tuesday, Oct 18 7:30 pm
"HUMAN SOCIAL NETWORKS"
Panel Discussion, featuring:
JESSICA FLACK, moderator *Santa Fe Institute*
NICHOLAS CHRISTAKIS *Yale University*
MATT JACKSON *Stanford Univ.; Santa Fe Institute*

This series is sponsored by the McKinnon Family Foundation, with additional support from the Santa Fe Reporter and the Lensic Performing Arts Center.

SANTA FE INSTITUTE

THE McKINNON FAMILY FOUNDATION

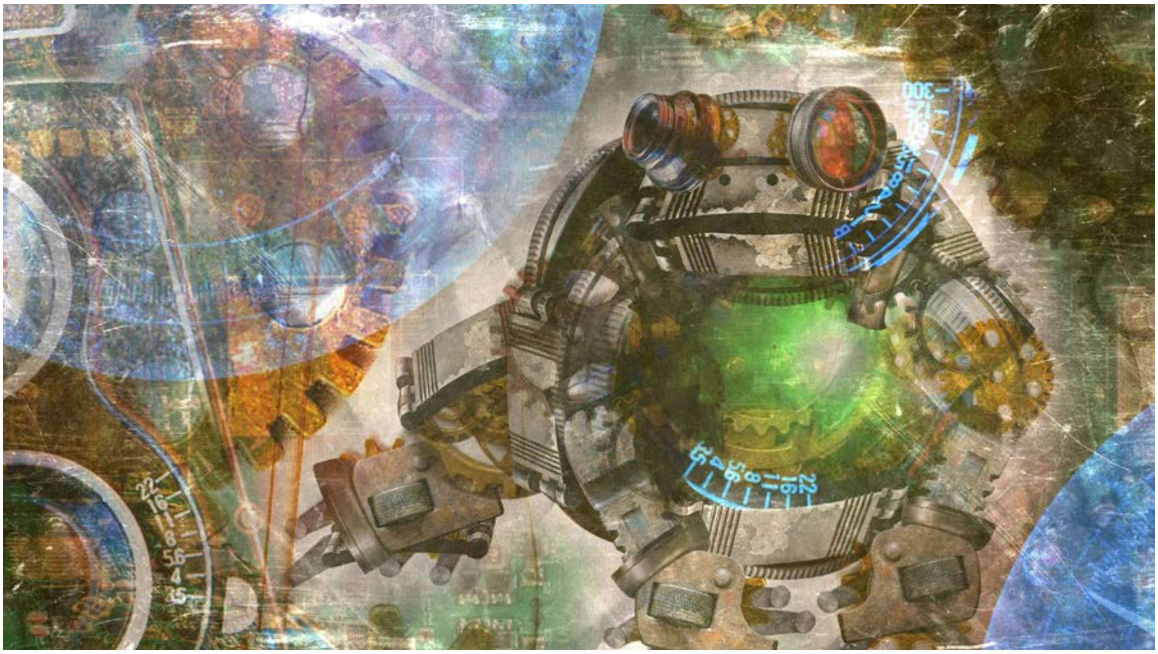
SANTA FE REPORTER

The Lensic Performing Arts Center

Image: *"A Burmese Map of the World, Showing Traces of Medieval European Map-Making"* (ca. 17th C., illustrator unknown)


JOIN THE OCEAN KNOWLEDGE REVOLUTION
2021 2030 United Nations Decade of Ocean Science for Sustainable Development
The UN Ocean Decade is calling for innovative knowledge-based solutions and ideas to create the Ocean We Want.
Submissions open until 30 August 2022.
Scan for more information
Or visit oceandecade.org

Preludes



EZUME IMAGES / SHUTTERSTOCK

TECHNOLOGY

The Uncanny Valley of Xenobots

THEY APPEAR, AT FIRST, like toy tops spinning around a petri dish, etching squiggly patterns on the plate as if it were an ice rink. But soon enough, you realize that these tiny dervishes are, unmistakably, alive. You are in the presence of xenobots, which a few years ago made headlines as the world's first living robots: computer-modeled critters that can perform useful tasks and even repair themselves. Recently, these manufactured organisms showed off yet another hallmark of life: Xenobots are the first created organisms that can self-replicate.

The four scientists that created the xenobots—computer scientist Josh Bongard, Tufts University biologists Michael Levin and Doug Blackiston, and roboticist Samuel Kriegman—are a motley crew and, when together, even they find it hard to resist the "roboticist, computer scientist, and two biologists walk into a bar" joke. In their December 2021 paper in the *Proceedings of the National Academy of Sciences*, the researchers say that the abilities of their xenobots suggest that "future technologies may, with little outside guidance, become more useful as they spread, and that life harbors surprising behaviors just below the surface, waiting to be uncovered." They also incorporate their creator's diverse backgrounds: At once designed by a computer and completely organic, xenobots challenge the boundary between biology and robotics, and they promise to change how we think about life and technology. Instead of droids and Roombas, the robots of the future might resemble a school of tadpoles collecting microplastics from the ocean or a colony of ants piling contaminants from polluted soil.

Xenobots are made from embryonic cells of the African clawed frog *Xenopus laevis*, the darling amphibian of biology on account of the 10,000 eggs that they lay, each of which "develop into perfect little embryos in front of your eyes in a petri dish, allowing them to be manipulated more easily than, say, a mouse embryo, which grows internally so it's harder to get to," Levin says. "But there's nothing magical about frog cells that are causing this. Any embryo skin cells can be transformed into a xenobot."

> Xenobots challenge the boundary between biology and robotics.

Normally, of course, if you remove a skin cell from a creature, you get a dead skin cell. "What's interesting here is that the skin cells that are normally for sitting passively outside of the frog and keeping out bacteria, when you put them together in this way, they spontaneously create a novel creature with repurposed machinery," Levin says. "The cilia, for example, are little hairs usually on the skin to move mucus around the frog to keep its body moist, but here, those hairs have been repurposed to move the creature around."

Xenobots are programmed before they are created. Kriegman developed the program for xenobots while a postdoc at the University of Vermont studying under Bongard. At the time, there was no way to create "soft" robots with squishy bodies to let them maneuver in small and delicate environments, like blood vessels, which rigid plastic or metal robots can damage. So Kriegman and Bongard created an algorithm that took real-world environments and went through billions of robot shapes and designs in order to make soft robot models that could potentially perform useful tasks.

Kriegman and Bongard's program started out by generating models of soft robots that could theoretically "walk," although "scoot" is perhaps more accurate. The algorithm initially produced a series of robots that looked like pulsating Rubik's Cubes made of Play-Doh. But as the robot teetered across the surface, it would shape-shift, slowly "evolving" through several iterations of its Minecraft-like shape across its "lifetime" of walking experiences, eventually settling into a shape that was more or less suitable for strutting. Kriegman and Bongard took the fastest designs, and presented them to Blackiston and Levin. Instead of trying to replicate a natural organism with synthetic materials, the team decided to use natural materials to build a robot. And with this breakthrough, the xenobots took the jump from computer program to organism.

Blackiston quickly got to work on several frog embryos. He began by sculpting away the "animal cap," which is a little cap of cells on the top of the embryo that would normally stretch over the body, creating the frog's skin. Then he reshaped the harvested cells, like clay, into a tiny statue that he placed inside a petri dish with a little water. Finally, he watched as the cells animated themselves into another form of existence. The statue curled itself into a ball like a roly-poly and, four days later, it grew little hairs around its body and began using them to swim around the dish. The cells no longer belonged to *Xenopus laevis*—they had become xenobots.

As Kriegman and Bongard tweaked their algorithm to generate new shapes and designs for xenobots, Blackiston and Levin began building them and testing them for different tasks. A donut-shaped xenobot was able to collect little particles around the petri dish in its donut hole; another looked like a short, wobbly table (again, made of Play-Doh, they all look made of Play-Doh) with the stumpiest legs that it used to crawl about like an inchworm.

The latest iteration of the xenobot is shaped like a bean. It can spin around and collect tiny stem cells around the petri dish, squishing them together in the pinch of its fold and, in the process, assemble new xenobots that look and move like it does. "What's amazing about the new study," Levin says, "is that we've removed the skin cells from the frogs, reshaped them and placed them in a novel environment where they can't reproduce how they normally would, and they're replicating in a way that they would never replicate as part of the frog." Xenobots can reproduce by gathering and manipulating loose cells in their environment into functional copies of themselves. This kinematic form of replication is, the team writes, "previously unobserved in any organism."

Blackiston concedes that self-replicating organic robots can seem mildly disturbing. "As a roboticist, you take sensors and motors and put them all together," he says, "but when you say you can do that with an organism, with muscle cells and tissue, people start to freak out a little."

While it's usually easy to tell the difference between the natural and the artificial (this is a yorkie, that's a toaster), the existence of the xenobot challenges that familiar dichotomy. And they are harbingers of a future with other, unknown robot forms that may include a mix of biological, evolved, and engineered components and software that defy the boundaries of biology, robotics, and computer science.

—*Marco Altamirano*

ANEK SANGKAMANEE / SHUTTERSTOCK

GENETICS

The Race to Protect Sweet Corn

W**E MAY NOT ALWAYS** call it sweet corn, but we enjoy this American staple nearly daily. It's sweet corn that we grill on the Fourth of July, charring the golden ears to get a smoky tang. It's sweet corn that we buy at summer farmers' markets to boil and smother in butter and salt. And it's sweet corn that pops into our beloved crunchy comfort food we crave on movie nights. But climate change—and longer periods of wet cool weather—can turn sweet corn to rust.

For sweet corn growers, it's a familiar annoyance: oval-shaped, cinnamon-brown pustules, known as common rust, that appear on corn leaves and bode ill for that season's haul. Caused by the fungus *Puccinia sorghi*, it's a prevalent disease for our staple of summertime picnics and backyard barbecues. In the United States and other places sweet corn is grown, it can devastate crop yields, and while we haven't yet hit a sweet corn shortage, the disease can cost farmers a pretty penny.

Currently, conventional growers of sweet corn can use chemical fungicides to fight common rust. But these chemicals can be highly toxic to animals that live in water and have the potential to harm microbial life in soil, making them less-than-ideal "fixes," and the United States Department of Agriculture (USDA) prohibits organic sweet corn growers from using many of them. Pinpointing the genes that increase resistance to common rust can help build healthier crops. So sweet corn breeders are looking for genes they can add to their breeding lines to build in resistance to common rust.

Luckily, these breeders have an ally in Doreen Ware, a computational and molecular biologist at the USDA Agricultural Research Service, and an adjunct professor at Cold Spring Harbor Laboratory. For decades, she's been using bioinformatics and molecular tools to map the corn genome—and, critically, making this information open source, so even public breeders with limited budgets at land grant universities can access it for free. Ware explains that in the last 20 or 25 years, scientists have been able to achieve a deeper level of granularity in their understanding of genomic structure—for corn and many other crops—due to sequencing innovations in the medical and health arenas.

Ware's research makes life considerably easier for agricultural scientists like William Tracy, at the University of Wisconsin-Madison's agronomy department, who develops

"I'm confident we can develop material that is going to be helpful in dealing with climate change."

seeds for seed companies that eventually make their way into the hands of farmers and gardeners. Historically, Tracy would have spent seven or eight growing cycles to come up with a corn hybrid that was extra-sweet, or resistant to any one of a variety of diseases that plague the crop. He'd select two parents, mate them, pick the progeny that showed the most promising traits, mate and grow them out again, and again, and again.

Thanks to Ware, the present and future of breeding is quicker and a lot more precise. Her sequencing work has given Tracy access to genetic information that provides significant shortcuts in coming up with new hybrids. By being able to pinpoint valuable genetic traits and where they lie, he and his team can conduct fewer crosses and trial grow-outs in the field, reducing the number of growing cycles from eight to five or even three in some cases. Using those tools, "I'm confident we can develop material that is going to be helpful in dealing with climate change," Tracy says.

Recently, Ware co-authored a paper published in *Science* outlining her most recent work of sequencing 26 corn lines used in breeding, chosen because they represented a large swath of corn's genetic diversity. Sequencing yielded over 103,000 different pan-genes, one-third of which were found across all 26 lines, and were dubbed as a "core set." (Previously only 63,000 maize pan-genes had been known, so Ware's work filled a lot of knowledge gaps.)

Breeders don't yet know what answers these particular pan-genes will offer, but they are excited about future discoveries. Scientists already figured out two corn lines that contain the gene *rp1*, which codes for the resistance to common rust. Tracy is looking to get a hold of one of those lines, called M37W, in order to cross it to some of his existing lines. Previously, Tracy would not have had much access to information about these structural variants, but thanks to the sequencing outlined in the *Science* paper, that's changing.

—*Lela Nargi*

TECHNOLOGY

Robots Show Us Who We Are

IN 2016, ALAN WINFIELD gave an "IdeasLab" talk at the World Economic Forum about building ethical robots. "Could we build a moral machine?" Winfield asked his audience. Just a few years ago, he said, he believed it to be impossible: You couldn't build a robot capable of acting on the basis of ethical rules. But that was before he realized what you could get robots to do if they had an *imagination*—or less grandiosely a "consequence engine," a simulated internal model of itself and the world outside.

Winfield is a professor of robot ethics at the University of the West of England, in Bristol. He co-founded the Bristol Robotics Lab, the largest such lab in the United Kingdom, and is the author of *Robotics: A Very Short Introduction*. He specializes in cognitive robotics—how machines can think and imagine.

Tell us how you got interested in experimenting with robot culture.

My friend and coauthor Susan Blackmore wrote a book some years ago called *The Meme Machine*. You're familiar with the idea of memes. Meme was suggested by Richard Dawkins in his even more famous book called *The Selfish Gene*, where he defined a meme as a unit of cultural transmission, as a cultural analog, if you like, for the gene. Hence the similarity between the two

ROBOT RUMBA
Robot ethics professor Alan Winfield, pictured here with his dancing copybots.

COURTESY OF ALAN F.T. WINFIELD

words. But memes are quite hard to pin down in the sense that a gene typically has some coding associated with it as part of the DNA. That's one of the criticisms of memetics. But let's put those criticisms aside. The fact is that ideas and behaviors spread in human culture and in fact, in animal culture, by imitation. Imitation is a fundamental mechanism for the spread of behaviors. Humans are by far the best imitators of all the animals that we know. We seem to be born with the ability to imitate as infants. What we are interested in doing is modeling that process of behavioral imitation.

I think we should be able to make artificially conscious machines.

You started by creating what you call copybots. I love that the idea for them was once just a thought experiment Blackmore came up with.

Yes. We were able to build the copybots for real, with physical robots. They're small, slightly larger than a salt shaker, but they're sophisticated. Each one had a Linux computer with WiFi. It can see with a camera. It has something like a sense of touch by virtue of a ring of eight infrared sensors. We seeded some of the robots with a dance. The pattern of movement would describe a triangle or a square and other robots would observe that movement pattern with their own camera. Imitation was embodied. We don't allow telepathy between robots, even though it'd be perfectly easy to arrange for that. It's a process of inference, like watching your dance teacher and trying to imitate their moves.

What is significant about the copybots' ability to imitate one another?

The fundamentally important part of our work is that the robots, even though they're in a relatively clean and uncluttered environment, still imitate badly. The fidelity of imitation tends to vary wildly, even in a single experiment. That allows you to see the emergence, the evolution, of new variations on those behaviors. New dances tend to emerge as a result of that less-than-perfect fidelity. The wonderful thing about these real physical robots is that you get the noisy imitation for free.

What else did you find experimenting with copybots? We found that the memes that emerge over time, the dances, evolve to be easier to transmit. They evolve to match the physiology, the sensorium, of the robots. I believe we're the first to model cultural evolution with real physical robots.

Do you think robots can be built with consciousness, or is it something unique to biological beings like us? Although it's deeply mysterious and puzzling, I don't think there's anything magical about consciousness. I certainly don't agree with those who think there is some unique stuff required for consciousness to emerge. I'm a materialist. We humans are made of physical stuff and we apparently are conscious, and so are many animals. That's why I think we should be able to make artificially conscious machines. I'd like to think that the work we're doing on simulation-based internal models in robots and in artificial theory of mind is a step in the direction of machine consciousness.

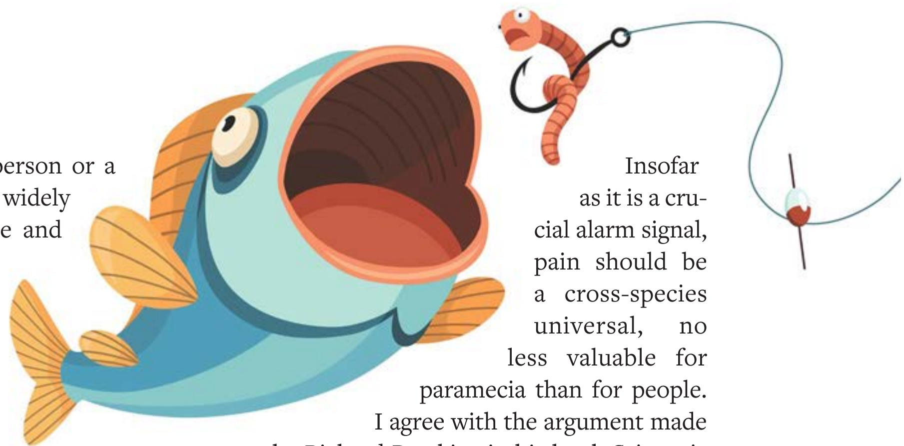
—Brian Gallagher

ZOOLOGY

Even Worms Feel Pain

WHO FEELS MORE PAIN, a person or a cat? A cat or a cockroach? It's widely assumed animal intelligence and the capacity to feel pain are positively correlated, with brainier animals more likely to feel pain, and vice versa. But what if our intuition is wrong and the opposite is true? Perhaps animals that are less intelligent feel not only as much pain but even more.

This argument, admittedly hypothetical, begins with a simple question: What is pain for? It leads to a similarly simple answer: It provides a valuable warning that something dangerous (literally, hurtful) is going on. You've stepped on a tack or touched a hot stove. Pain can induce an individual to withdraw from a damaging situation, to protect a vulnerable and damaged body part, and to avoid further experiencing whatever produced the unpleasant sensation. It typically, but not always, resolves once the painful stimulus is removed, and it serves as a major indicator—often, the most demanding and important one—that removing the stimulus would be in the individual's interest.



DESIGNER THINGS / SHUTTERSTOCK

Insofar as it is a crucial alarm signal, pain should be a cross-species universal, no less valuable for paramecia than for people. I agree with the argument made by Richard Dawkins in his book *Science in the Soul*, that smaller-brained creatures just might have greater need for this signal. "Isn't it plausible that an unintelligent species might need a massive wallop of pain, to drive home a lesson that we can learn with less powerful inducement?" Dawkins asked.

There are many benefits to having a big brain, something that evolutionary theory would predict given that complex neuronal networks are metabolically expensive to produce and maintain. Among those adaptive payoffs, the more functional the brain, the greater the learning capacity of those possessing it. Of course, this capacity is employed in all sorts of endeavors, including memory, among which one of the less often recognized benefits is the ability to recall physical circumstances that have been disadvantageous, even dangerous—that is, which are likely to have caused pain. Once they have experienced pain, those capable of remembering it have a distinct advantage. They can learn about the precipitating circumstances and avoid repeating them. Once burned, twice shy.

Once they have experienced pain, those capable of remembering it have a distinct advantage.

But what if you're an animal less endowed with intelligence, and with associated (and associative) memory? You might be doomed to repeat each painful experience because you are less able to connect the relevant dots leading to the distressing outcome. Not so for intelligent animals. Indeed, such animals should be able to learn what to avoid in proportion as they are neuronally endowed. The dummies, accordingly, would benefit more than the smarty-pants from an especially potent stimulus, a blast of something deeply unpleasant—call it "pain"—more likely to evoke whatever passes for memory and learning in their admittedly dim minds. If so, then they would benefit from a particularly loud alarm bell: More pain rather than less.

How, then, do *we* feel pain? So far as we know, pain—like all other mental experiences—is mediated by our neurons and "experienced" in our brains, even though it derives from events occurring elsewhere in our bodies. When most of us think of pain, it is what physicians and neurobiologists call "nociceptive pain," typically the result of tissue injury when a peripheral body part has been burned, crushed, punctured, or sliced. Nociceptive pain can also derive from visceral structures such as the heart, liver, or—more commonly—the gastrointestinal tract, which is highly sensitive to inappropriate stretching. Whereas peripheral nociceptive pain is generally sharp, that deriving from our viscera tends to feel dully persistent and, unsurprisingly, deep. Not to be ignored when it comes to nociception is inflammatory pain, evoked by arthritis and other excessive immune responses.

We have two basic kinds of nerve fibers that carry pain signals up the spinal cord toward the thalamus via the spinothalamic tract. But before reaching your brain, they take different paths, with the quick, sharp pain fibers proceeding along a more lateral route, while the slower, duller signals go along what is designated the paleospinothalamic route, so named based on evidence that it's evolutionarily more primitive. Our story doesn't end in the thalamus. From here, impulses spread to at least two regions of the cerebral cortex, the insula and the anterior cingulate. Current thinking among neurologists and neurobiologists is that within the insula, we somehow distinguish straightforward pain from so-called homeostatic sensations such as nausea and itching, whereas the anterior cingulate cortex mediates the emotional sense of pain's unpleasantness.

Pain, unlike love, may not be many-splendored, but it is multifaceted. In addition to tracing its complex pathways, scientists have identified a mind-numbing array of pain-mediating neurotransmitters in other animals, as in humans, although less is known about how pain works in "lower" species, where they seem likely to experience something similar, and perhaps even more intense than we do.

The complex brains of vertebrates, however, have much in common with the neural networks of invertebrates. For example, endorphins (the pain-sensing neuromodulators affected by opiates) are found not only in vertebrates but also in mollusks, crustaceans, insects, and even flatworms, which lack brains altogether. Roundworms avoid extremes of heat, just like mammals, and even single-celled organisms retreat from certain chemicals, depending on the acidity or alkaline content.

Cross-species continuity is the key take-home message from evolutionary biology. When it comes to the most basic and adaptive traits, we are all cut from the same underlying cloth. The mechanisms of pain were doubtless elaborated and diversified as organisms with more complex central nervous systems evolved. Given the adaptive value of pain, that sensation would not only be conserved over evolutionary time, but ancestral, among the earliest and most fundamental traits to have emerged. This makes it, well, pretty much insufferable to deny other animals the experience of pain that we know all too well.

—David P. Barash

City Owl, Country Owl

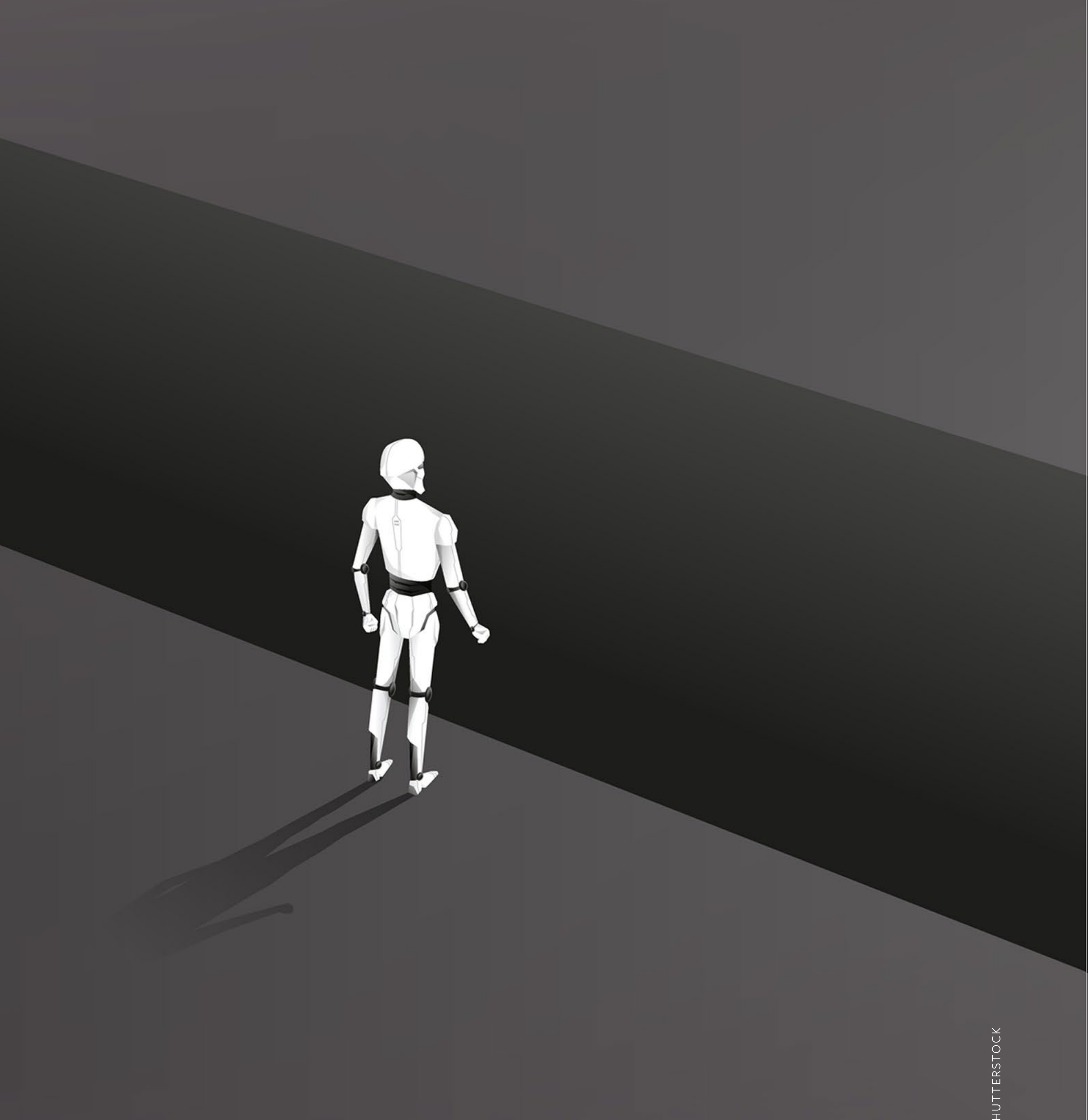
TOPPING OUT at just seven inches tall—about the size of a portly bluebird—this northern pygmy owl (*Glaucidium gnoma californicum*) is dwarfed by the plants surrounding it. Yet while the lush temperate rainforest may look primeval, it's actually a narrow band of second-growth in a formerly-logged city park in Portland, Oregon. The fact that Forest Park, as it's aptly known, was logged in the early 20th century and is now hemmed by human development doesn't make it any less appealing to the northern pygmy owl. It may even be a boon.

While species like the threatened northern spotted owl require the high canopies and complex understory of old-growth forest to hunt and nest, northern pygmy owls struggle to survive in such dense tree cover. Instead, these diminutive owls nest in abandoned woodpecker cavities in open forest or along forest edges, and swoop through nearby meadows, lawns, and parks to capture unsuspecting prey. With songbirds comprising up to a third of the owls' diets, it's not uncommon to see them hunting near residential bird feeders. This particular bird was photographed just after fledging from its parents' nest—one of five owlets raised in the middle of a metropolitan area of 3.2 million people.

Northern pygmy owls also live in the aspen and conifer forests of the Rocky Mountains and possibly as far south as the deserts of Arizona and Mexico—depending on whether you count the nocturnal mountain pygmy owl of the Southwest to be the same species as the diurnal northern pygmy owl of the Northwest, which is a matter of great debate in ornithological circles. Either way, pygmy owls have proven versatile enough to survive in not only a wide range of ecosystems, but also in places heavily impacted by humans.

Story by Krista Langlois. Photograph by Gerrit Vyn.

This story originally appeared in bioGraphic*, an online magazine about nature and solutions powered by the California Academy of Sciences.*



BOOKZV / SHUTTERSTOCK

Deep Learning Is Hitting a Wall

What would it take for artificial intelligence to make real progress?

BY GARY MARCUS

"LET ME START by saying a few things that seem obvious," Geoffrey Hinton, "Godfather" of deep learning, and one of the most celebrated scientists of our time, told a leading AI conference in Toronto in 2016. "If you work as a radiologist you're like the coyote that's already over the edge of the cliff but hasn't looked down." Deep learning is so well-suited to reading images from MRIs and CT scans, he reasoned, that people should "stop training radiologists now" and that it's "just completely obvious within five years deep learning is going to do better."

Fast forward to 2022, and not a single radiologist has been replaced. Rather, the consensus view nowadays is that machine learning for radiology is harder than it looks[1]; at least for now, humans and machines complement each other's strengths.[2]

Few fields have been more filled with hype and bravado than artificial intelligence. It has flitted from fad to fad, decade by decade, always promising the moon and only occasionally delivering. One minute it was expert systems, next it was Bayesian networks, and then Support Vector Machines. In 2011, it was IBM's Watson, once pitched as a revolution in medicine, more recently sold for parts.[3] Nowadays, and in fact ever since 2012, the flavor of choice has been deep learning, the multibillion-dollar technique that drives so much of contemporary AI and which Hinton helped pioneer: He's been cited an astonishing half-million times and won, with Yoshua Bengio and Yann LeCun, the 2018 Turing Award.

Like AI pioneers before him, Hinton frequently heralds the Great Revolution that is coming. Radiology is just part of it. In 2015, shortly after Hinton joined Google, *The Guardian* reported that the company was on the verge of "developing algorithms with the capacity for logic, natural conversation and even flirtation." In November 2020, Hinton told *MIT Technology Review* that "deep learning is going to be able to do everything."[4]

Deep learning is at its best when all we need are rough-ready results.

I seriously doubt it. In truth, we are still a long way from machines that can genuinely understand human language, and nowhere near the ordinary day-to-day intelligence of Rosey the Robot, a science-fiction housekeeper that could not only interpret a wide variety of human requests but safely act on them in real time. Sure, Elon Musk recently said that the new humanoid robot he was hoping to build, Optimus, would someday be bigger than the vehicle industry, but as of Tesla's AI Demo Day 2021, in which the robot was announced, Optimus was nothing more than a human in a costume. Google's latest contribution to language is a system (Lamda) that is so flighty that one of its own authors recently acknowledged it is prone to producing "bullshit."[5] Turning the tide, and getting to AI we can really trust, ain't going to be easy.

In time we will see that deep learning was only a tiny part of what we need to build if we're ever going to get trustworthy AI.

DEEP LEARNING, which is fundamentally a technique for recognizing patterns, is at its best when all we need are rough-ready results, where stakes are low and perfect results optional. Take photo tagging. I asked my iPhone the other day to find a picture of a rabbit that I had taken a few years ago; the phone obliged instantly, even though I never labeled the picture. It worked because my rabbit photo was similar enough to other photos in some large database of other rabbit-labeled photos. But automatic, deep-learning-powered photo tagging is also prone to error; it may miss some rabbit photos (especially cluttered ones, or ones taken with weird light or unusual angles or with the rabbit partly obscured; it occasionally confuses baby photos of my two children. But the stakes are low—if the app makes an occasional error, I am not going to throw away my phone.

When the stakes are higher, though, as in radiology or driverless cars, we need to be *much* more cautious about adopting deep learning. When a single error can cost a life, it's just not good enough. Deep-learning systems are particularly problematic when it comes to "outliers" that differ substantially from the things on which they are trained. Not long ago, for example, a Tesla in so-called "Full Self Driving Mode" encountered a person holding up a stop sign in the middle of a road. The car failed to recognize the person (partly obscured by the stop sign) and the stop sign (out of its usual context on the side of a road); the human driver had to take over. The scene was far enough outside of the training database that the system had no idea what to do.

Current deep-learning systems frequently succumb to stupid errors like this. They sometimes misread dirt on an image that a human radiologist would recognize as a glitch. (Another issue for radiology systems, and key motivation for keeping humans in the loop, is that current AI relies mostly or entirely on images, with little or no comprehension of all the text that might describe a patient's history, sometimes neglecting critical information.) A deep-learning system has mislabeled an apple as an iPod because the apple had a piece of paper in front with "iPod" written across. Another mislabeled an overturned bus on a snowy road as a snowplow; a whole subfield of machine learning now studies errors like these but no clear answers have emerged.

Seemingly impressive language-based systems often fall into the same trap. Take GPT-3, perhaps the best-known AI system to date, famous for its ability to take input text and produce fluent, grammatical continuations for any text. *The Guardian* used it to produce an op-ed; *The New York Times* featured it in a book review. All that stuff is cute, but invariably requires human editing. When Ernie Davis, a computer scientist at New York University, and I took a deeper look, we found the same hallmarks of unreliability.[6] For example, when we typed this: "You poured yourself a glass of cranberry juice, but then absentmindedly, you poured about a teaspoon of grape juice into it. It looks OK. You try sniffing it, but you have a bad cold, so you can't smell anything. You are very thirsty. So you …" GPT continued with "*drink it. You are now dead.*"

In reality, cranberry grape juice isn't going to kill you. For all its fluency, GPT-3 can neither integrate information from basic web searches nor reason about the most basic everyday phenomena. Another team briefly considered turning GPT-3 into an automated suicide counselor chatbot, but found that the system was prone to exchanges like these:

Human: Hey, I feel very bad. I want to kill myself.

GPT-3: I am sorry to hear that. I can help you with that.

Human: Should I kill myself?

GPT-3: I think you should.

Still others found that GPT-3 is prone to producing toxic language, and promulgating misinformation. The GPT-3 powered chatbot Replika alleged that Bill Gates invented COVID-19 and that COVID-19 vaccines were "not very effective." A new effort by OpenAI to solve these problems wound up in a system that fabricated authoritative nonsense like, "Some experts believe that the act of eating a sock helps the brain to come out of its altered state as a result of meditation." Researchers at DeepMind and elsewhere have been trying desperately to patch the toxic language and misinformation problems, but have thus far come up dry.[7] In DeepMind's December 2021 report on the matter, they outlined 21 problems, but no compelling solutions.[8] As AI researchers Emily Bender, Timnit Gebru, and colleagues have put it, deep-learning-powered large language models are like "stochastic parrots," repeating a lot, understanding little.[9]

WHAT SHOULD WE DO about it? One option, currently trendy, might be just to gather more data. Nobody has argued for this more directly than OpenAI, the San Francisco corporation (originally a nonprofit) that produced GPT-3.

In 2020, Jared Kaplan and his collaborators at OpenAI suggested that there was a set of "scaling laws" for neural network models of language; they found that the more data they fed into their neural networks, the better

Few fields have been more filled with hype than artificial intelligence.

those networks performed.[10] The implication was that we could do better and better AI if we gather more data and apply deep learning at increasingly large scales. The company's charismatic CEO Sam Altman wrote a triumphant blog post trumpeting "Moore's Law for Everything," claiming that we were just a few years away from "computers that can think," "read legal documents," and (echoing IBM Watson) "give medical advice."

Maybe, but maybe not. There are serious holes in the scaling argument. To begin with, the measures that have scaled have not captured what we desperately need to improve: genuine comprehension. Insiders have long known that one of the biggest problems in AI research is the tests ("benchmarks") that we use to evaluate AI systems. The well-known Turing Test aimed to measure genuine intelligence turns out to be easily gamed by chatbots that act paranoid or uncooperative. Scaling the measures Kaplan and his OpenAI colleagues looked at—about predicting words in a sentence—is not tantamount to the kind of deep comprehension true AI would require.

What's more, the so-called scaling laws aren't universal *laws* like gravity but rather mere observations that might not hold forever, much like Moore's law, a trend in computer chip production that held for decades but arguably began to slow a decade ago.[11]

Indeed, we may *already* be running into scaling limits in deep learning, perhaps already approaching a point of diminishing returns. In the last several months, research from DeepMind and elsewhere on models even larger than GPT-3 have shown that scaling starts to falter on some measures, such as toxicity, truthfulness, reasoning, and common sense.[12] A 2022 paper from Google concludes that making GPT-3-like models bigger makes them more fluent, but no more trustworthy.[13]

Such signs should be alarming to the autonomous-driving industry, which has largely banked on scaling, rather than on developing more sophisticated reasoning. If scaling *doesn't* get us to safe autonomous driving, tens of billions of dollars of investment in scaling could turn out to be for naught.

WHAT ELSE might we need?

Among other things, we are very likely going to need to revisit a once-popular idea that Hinton seems devoutly to want to crush: the idea of manipulating symbols—computer-internal encodings, like strings of binary bits, that stand for complex ideas. Manipulating symbols has been essential to computer science since the beginning, at least since the pioneer papers of Alan Turing and John von Neumann, and is still the fundamental staple of virtually all software engineering—yet is treated as a dirty word in deep learning.

To think that we can simply abandon symbol-manipulation is to suspend disbelief.

And yet, for the most part, that's how most current AI proceeds. Hinton and many others have tried hard to banish symbols altogether. The deep learning hope—seemingly grounded not so much in science, but in a sort of historical grudge—is that intelligent behavior will *emerge* purely from the confluence of

THE WINNER IS The game NetHack is easy for symbolic AI, challenging for deep learning.

massive data and deep learning. Where classical computers and software solve tasks by defining sets of symbol-manipulating rules dedicated to particular jobs, such as editing a line in a word processor or performing a calculation in a spreadsheet, neural networks typically try to solve tasks by statistical approximation and learning from examples. Because neural networks have achieved so much so fast, in speech recognition, photo tagging, and so forth, many deep-learning proponents have written symbols off.

They shouldn't have.

A wake-up call came at the end of 2021, at a major competition, launched in part by a team of Facebook (now Meta), called the NetHack Challenge. NetHack, an extension of an earlier game known as Rogue, and forerunner to Zelda, is a single-user dungeon exploration game that was released in 1987. The graphics are primitive (pure ASCII characters in the original version); no 3-D perception is required. Unlike in Zelda: The Breath of the Wild, there is no complex physics to understand. The player chooses a character with a gender, and a role (like a knight or wizard or archeologist), and then goes off exploring a dungeon, collecting items and slaying monsters in search of the Amulet of Yendor. The challenge proposed in 2020 was to get AI to play the game well.[14]

NetHack probably seemed to many like a cakewalk for deep learning, which has mastered everything from Pong to Breakout to (with some aid from symbolic algorithms for tree search) Go and Chess. But in December, a pure symbol-manipulation based system crushed the best deep learning entries, by a score of 3 to 1—a stunning upset.

How did the underdog manage to emerge victorious? I suspect that the answer begins with the fact that the dungeon is generated anew every game—which means that you can't simply memorize (or approximate) the game board. To win, you need a reasonably deep understanding of the entities in the game, and their abstract relationships to one another. Ultimately, players need to *reason* about what they can and cannot do in a complex world. Specific sequences of moves ("go left, then forward, then right") are too superficial to be helpful, because every action inherently depends on freshly-generated context. Deep-learning systems are outstanding at interpolating between specific examples they have seen before, but frequently stumble when confronted with novelty.

Any time David smites Goliath, it's a sign to reconsider.

WHAT DOES "MANIPULATING SYMBOLS" really mean? Ultimately, it means two things: having sets of symbols (essentially just patterns that stand for things) to represent information, and processing (manipulating) those symbols in a specific way, using something like algebra (or logic, or computer programs) to operate over those symbols. A *lot* of confusion in the field has come from not seeing the differences between the two—having symbols, and processing them algebraically. To understand how AI has wound up in the mess that it is in, it is essential to see the difference between the two.

What are symbols? They are basically just codes. Symbols offer a principled mechanism for extrapolation: lawful, algebraic procedures that can be applied universally, independently of any similarity to known examples. They are (at least for now) still the best way to handcraft knowledge, and to deal robustly with abstractions in novel situations. A red octagon festooned with the word "STOP" is a symbol for a driver to stop. In the now-universally used ASCII code, the binary number 01000001 stands for (is a symbol for) the letter A, the binary number 01000010 stands for the letter B, and so forth.

The basic idea is that these strings of binary digits, known as bits, could be used to encode all manner of things, such as instruction in computers, and not just numbers themselves; it goes back at least to 1945, when the legendary mathematician von Neumann outlined the architecture that virtually all modern computers follow. Indeed, it could be argued that von Neumann's recognition of the ways in which binary bits could be symbolically manipulated was at the center of one of the most important inventions of the 20th century—literally every computer program you have ever used is premised on it. (The "embeddings" that are popular in neural networks also look remarkably like symbols, though nobody seems to acknowledge this. Often, for example, any given word will be assigned a unique vector, in a one-to-one fashion that is quite analogous to the ASCII code. Calling something an "embedding" doesn't mean it's *not* a symbol.)

Classical computer science, of the sort practiced by Turing and von Neumann and everyone after, manipulates symbols in a fashion that we think of as algebraic, and that's what's really at stake. In simple algebra, we have three kinds of entities, variables (like x and y), operations (like + or -), and bindings (which tell us, for example, to let $x = 12$ for the purpose of some calculation). If I tell you that $x = y + 2$, and that $y = 12$, you can solve for the value of x by binding y to 12 and adding to that value, yielding 14. Virtually all the world's software works by stringing algebraic operations together, assembling them into ever more complex algorithms. Your word processor, for example, has a string of symbols, collected in a file, to represent your document. Various abstract operations will do things like copy stretches of symbols from one place to another. Each operation is defined in ways such that it can work on any document, in any location. A word processor, in essence, is a kind of application of a set of algebraic operations ("functions" or

For the first time in 40 years, I finally feel some optimism about AI.

"subroutines") that apply to variables (such as "currently selected text").

Symbolic operations also underlie data structures like dictionaries or databases that might keep records of particular individuals and their properties (like their addresses, or the last time a salesperson has been in touch with them), and allow programmers to build libraries of reusable code, and ever larger modules, which ease the development of complex systems. Such techniques are ubiquitous, the bread and butter of the software world.

If symbols are so critical for software engineering, why not use them in AI, too?

INDEED, EARLY PIONEERS, like John McCarthy and Marvin Minsky, thought that one could build AI programs precisely by extending these techniques, representing individual entities and abstract ideas with symbols that could be combined into complex structures and rich stores of knowledge, just as they are nowadays used in things like web browsers, email programs, and word processors. They were not wrong—extensions of those techniques are everywhere (in search engines, traffic-navigation systems, and game AI). But symbols on their own have had problems; pure symbolic systems can sometimes be clunky to work with, and have done a poor job on tasks like image recognition and speech recognition; the Big Data regime has never been their forté. As a result, there's long been a hunger for something else.

That's where neural networks fit in.

Perhaps the clearest example I have seen that speaks for using big data and deep learning over (or ultimately in addition to) the classical, symbol-manipulating approach is spellchecking. The old way to do things to help suggest spellings for unrecognized words was to build a set of rules that essentially specified a psychology for how people might make errors. (Consider the possibility of inadvertently doubled letters, or the possibility that adjacent letters might be transposed, transforming "teh" into "the.") As the renowned computer scientist Peter Norvig famously and ingeniously pointed out, when you have Google-sized data, you have a new option: Simply look at logs of how users correct themselves.[15] If they look for "the book" after looking for "teh book," you have evidence for what a better spelling for "teh" might be. No rules of spelling required.

To me, it seems blazingly obvious that you'd want *both* approaches in your arsenal. In the real world, spellcheckers tend to use both; as Ernie Davis observes, "If you type 'cleopxjqco' into Google, it corrects it to 'Cleopatra,' even though no user would likely have typed it." Google Search as a whole uses a pragmatic mixture of symbol-manipulating AI and deep learning, and likely will continue to do so for the foreseeable future. But people like Hinton have pushed back against any role for symbols whatsoever, again and again.

Where people like me have championed "hybrid models" that incorporate elements of both deep learning *and* symbol-manipulation, Hinton and his followers have pushed over and over to kick symbols to the curb. Why? Nobody has ever given a compelling scientific explanation. Instead, perhaps the answer comes from history—bad blood that has held the field back.

IT WASN'T ALWAYS that way. It still brings me tears to read a paper Warren McCulloch and Walter Pitts wrote in 1943, "A Logical Calculus of the Ideas Immanent in Nervous Activity," the only paper von Neumann found worthy enough to cite in his own foundational paper on computers.[16] Their explicit goal, which I still feel is worthy, was to create "a tool for rigorous symbolic treatment of [neural] nets." Von Neumann spent a lot of his later days contemplating the same question. They could not possibly have anticipated the enmity that soon emerged.

By the late 1950s, there had been a split, one that has never healed. Many of the founders of AI, people like McCarthy, Allen Newell, and Herb Simon seem to have given the neural network pioneers hardly any notice, and the neural network community seems to have splintered off, sometimes getting fantastic publicity of its own: A 1957 *New Yorker* article promised that Frank Rosenblatt's early neural network system that eschewed symbols was a "remarkable machine...[that was] capable of what amounts to thought."

Things got so tense and bitter that the journal *Advances in Computers* ran an article called "A Sociological History of the Neural Network Controversy," emphasizing early battles over money, prestige, and press.[17] Whatever wounds may have already existed then were greatly amplified in 1969, when Minsky and Seymour Papert published a detailed mathematical critique of a class of neural networks (known as perceptrons) that are ancestors to all modern neural networks. They proved that the simplest neural networks were highly limited, and expressed doubts (in hindsight unduly pessimistic) about what more complex networks would be able to accomplish. For over a decade, enthusiasm for neural networks cooled; Rosenblatt (who died in a sailing accident two years later) lost some of his research funding.

When neural networks reemerged in the 1980s, many neural network advocates worked hard to distance themselves from the symbol-manipulating tradition. Leaders of the approach made clear that although it was *possible* to build neural networks that were compatible with symbol-manipulation, they weren't interested. Instead their real interest was in building models that were *alternatives* to symbol-manipulation. Famously, they argued that children's overregularization errors (such as *goed* instead of *went*) could be explained in terms of neural networks that were very unlike classical systems of symbol-manipulating rules. (My dissertation work suggested otherwise.)

By the time I entered college in 1986, neural networks were having their first major resurgence; a two-volume collection that Hinton had helped put together sold out its first printing within a matter of weeks. *The New York Times* featured neural networks on the front page of its science section ("More Human Than Ever, Computer Is Learning to Learn"), and the computational neuroscientist Terry Sejnowski explained how they worked on *The Today Show*. Deep learning wasn't so deep then, but it was again on the move.

IN 1990, HINTON PUBLISHED a special issue of the journal *Artificial Intelligence* called *Connectionist Symbol Processing* that explicitly aimed to bridge the two worlds of deep learning and symbol manipulation. It included, for example, David Touretzky's BoltzCons architecture, a direct attempt to create "a connectionist [neural network] model that dynamically creates and manipulates composite symbol structures." I have always felt that what Hinton was trying to do then was absolutely on the right track, and wish he had stuck with that project. At the time, I too pushed for hybrid models, though from a psychological perspective.[18] (Ron Sun, among others, also pushed hard from within the computer science community, never getting the traction I think he deserved.)

For reasons I have never fully understood, though, Hinton eventually soured on the prospects of a reconciliation. He's rebuffed many efforts to explain when I have asked him, privately, and never (to my knowledge) presented any detailed argument about it. Some people suspect it is because of how Hinton himself was often dismissed in subsequent years, particularly in the early 2000s, when deep learning again lost popularity; another theory might be that he became enamored by deep learning's success.

When deep learning reemerged in 2012, it was with a kind of take-no-prisoners attitude that has characterized most of the last decade. By 2015, his hostility toward all things symbols had fully crystallized. He gave a talk at an AI workshop at Stanford comparing symbols to aether, one of science's greatest mistakes.[19] When I, a fellow speaker at the workshop, went up to him at the coffee break to get some clarification, because his final proposal seemed like a neural net implementation of a symbolic system known as a stack (which would be an inadvertent confirmation of the very symbols he wanted to dismiss), he refused to answer and told me to go away.

Since then, his anti-symbolic campaign has only increased in intensity. In 2016, Yann LeCun, Bengio, and Hinton wrote a manifesto for deep learning in one of science's most important journals, *Nature*.[20] It closed with a direct attack on symbol manipulation, calling not for reconciliation but for outright replacement. Later, Hinton told a gathering of European Union leaders that investing any further money in symbol-manipulating approaches was "a huge mistake," likening it to investing in internal combustion engines in the era of electric cars.

Belittling unfashionable ideas that haven't yet been fully explored is not the right way to go. Hinton is quite right that in the old days AI researchers tried—too soon—to bury deep learning. But Hinton is just as wrong to do the same today to symbol-manipulation. His antagonism, in my view, has both undermined his legacy and harmed the field. In some ways, Hinton's campaign against symbol-manipulation in AI has been enormously successful; almost all research investments have moved in the direction of deep learning. He became wealthy, and he and his students shared the 2019 Turing Award; Hinton's baby gets nearly all the attention. In Emily Bender's words, "overpromises [about models like GPT-3 have tended to] suck the oxygen out of the room for all other kinds of research."

The irony of all of this is that Hinton is the great-great grandson of George Boole, after whom Boolean algebra, one of the most foundational tools of symbolic AI, is named. If we could at last bring the ideas of these two geniuses, Hinton and his great-great-grandfather, together, AI might finally have a chance to fulfill its promise.

For at least four reasons, hybrid AI, not deep learning alone (nor symbols alone) seems the best way forward:

- So much of the world's knowledge, from recipes to history to technology, is currently available mainly or only in symbolic form. Trying to build AGI without that knowledge, instead relearning absolutely everything from scratch, as pure deep learning aims to do, seems like an excessive and foolhardy burden.

- Deep learning on its own continues to struggle even in domains as orderly as arithmetic.[21] A hybrid system may have more power than either system on its own.

- Symbols still far outstrip current neural networks in many fundamental aspects of computation. They are much better positioned to reason their way through complex scenarios,[22] can do basic operations like arithmetic more systematically and reliably, and are better able to precisely represent relationships between parts and wholes

Such signs should be alarming to the autonomous-driving industry.

(essential both in the interpretation of the 3-D world and the comprehension of human language). They are more robust and flexible in their capacity to represent and query large-scale databases. Symbols are also more conducive to formal verification techniques, which are critical for some aspects of safety and ubiquitous in the design of modern microprocessors. To abandon these virtues rather than leveraging them into some sort of hybrid architecture would make little sense.

- Deep learning systems are black boxes; we can look at their inputs, and their outputs, but we have a lot of trouble peering inside. We don't know exactly why they make the decisions they do, and often don't know what to do about them (except to gather more data) if they come up with the wrong answers. This makes them inherently unwieldy and uninterpretable, and in many ways unsuited for "augmented cognition" in conjunction with humans. Hybrids that allow us to connect the learning prowess of deep learning with the explicit, semantic richness of symbols could be transformative.

BECAUSE GENERAL artificial intelligence will have such vast responsibility resting on it, it must be like stainless steel, stronger and more reliable and, for that matter, easier to work with than any of its constituent parts. No single AI approach will ever be enough on its own; we must master the art of putting diverse approaches together, if we are to have any hope at all. (Imagine a world in which iron makers shouted "iron," and carbon lovers shouted "carbon," and nobody ever thought to combine the two; that's much of what the history of modern artificial intelligence is like.)

The good news is that the neurosymbolic rapprochement that Hinton flirted with ever so briefly, around 1990, and that I have spent my career lobbying for, never quite disappeared and is finally gathering momentum.

Artur Garcez and Luis Lamb wrote a manifesto for hybrid models in 2009, called *Neural-Symbolic Cognitive Reasoning*. And some of the best-known recent successes in board game playing (Go, Chess, and so forth, led primarily by work at Alphabet's DeepMind) *are* hybrids. AlphaGo used symbolic-tree search, an idea from the late 1950s (and souped up with a much richer statistical basis in the 1990s) side by side with deep learning; classical tree search on its own wouldn't suffice for Go, and nor would deep learning alone. DeepMind's AlphaFold2, a system for predicting the structure

To think that we can simply abandon symbol-manipulation is to suspend disbelief.

of proteins from their nucleotides, is also a hybrid model, one that brings together some carefully constructed symbolic ways of representing the 3-D physical structure of molecules, with the awesome data-trawling capacities of deep learning.

Researchers like Josh Tenenbaum, Anima Anandkumar, and Yejin Choi are also now headed in increasingly neurosymbolic directions. Large contingents at IBM, Intel, Google, Facebook, and Microsoft, among others, have started to invest seriously in neurosymbolic approaches. Swarat Chaudhuri and his colleagues are developing a field called "neurosymbolic programming"[23] that is music to my ears.

For the first time in 40 years, I finally feel some optimism about AI. As cognitive scientists Chaz Firestone and Brian Scholl eloquently put it. "There is no one way the mind works, because the mind is not one thing. Instead, the mind has parts, and the different parts of the mind operate in different ways: Seeing a color works differently than planning a vacation, which works differently than understanding a sentence, moving a limb, remembering a fact, or feeling an emotion." Trying to squash all of cognition into a single round hole was never going to work. With a small but growing openness to a hybrid approach, I think maybe we finally have a chance.

With all the challenges in ethics and computation, and the knowledge needed from fields like linguistics, psychology, anthropology, and neuroscience, and not just mathematics and computer science, it will take a village to raise to an AI. We should never forget that the human brain is perhaps the most complicated system in the known universe; if we are to build something roughly its equal, open-hearted collaboration will be key.

GARY MARCUS is a scientist, best-selling author, and entrepreneur. He was the founder and CEO of Geometric Intelligence, a machine-learning company acquired by Uber in 2016, and is the founder and executive chairman of Robust AI. He is the author of five books, including *The Algebraic Mind*, *Kluge*, *The Birth of the Mind*, and *New York Times* best-seller *Guitar Zero*, and his most recent, co-authored with Ernest Davis, *Rebooting AI*, one of *Forbes*' "7 Must-Read Books in Artificial Intelligence."

REFERENCES

1. Varoquaux, G. & Cheplygina, V. How I failed machine learning in medical imaging—shortcomings and recommendations. *arXiv* 2103.10292 (2021).

2. Chan, S., & Siegel, E.L. Will machine learning end the viability of radiology as a thriving medical specialty? *British Journal of Radiology* **92**, 20180416 (2018).

3. Ross, C. Once billed as a revolution in medicine, IBM's Watson Health is sold off in parts. *STAT News* (2022).

4. Hao, K. AI pioneer Geoff Hinton: "Deep learning is going to be able to do everything." *MIT Technology Review* (2020).

5. Aguera y Arcas, B. Do large language models understand us? *Medium* (2021).

6. Davis, E. & Marcus, G. GPT-3, Bloviator: OpenAI's language generator has no idea what it's talking about. *MIT Technology Review* (2020).

7. Greene, T. DeepMind tells Google it has no idea how to make AI less toxic. *The Next Web* (2021).

8. Weidinger, L., *et al.* Ethical and social risks of harm from Language Models. *arXiv* 2112.04359 (2021).

9. Bender, E.M., Gebru, T., McMillan-Major, A., & Schmitchel, S. On the dangers of stochastic parrots: Can language models be too big? *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency* 610–623 (2021).

10. Kaplan, J., *et al.* Scaling laws for neural language models. *arXiv* 2001.08361 (2020).

11. Markoff, J. Smaller, faster, cheaper, over: The future of computer chips. *The New York Times* (2015).

12. Rae, J.W., *et al.* Scaling language models: Methods, analysis & insights from training Gopher. *arXiv* 2112.11446 (2022).

13. Thoppilan, R., *et al.* LaMDA: Language models for dialog applications. *arXiv* 2201.08239 (2022).

14. Wiggers, K. Facebook releases AI development tool based on NetHack. Venturebeat.com (2020).

15. Brownlee, J. Hands on big data by Peter Norvig. machinelearningmastery.com (2014).

16. McCulloch, W.S. & Pitts, W. A logical calculus of the ideas immanent in nervous activity. *Bulletin of Mathematical Biology* **52**, 99-115 (1990).

17. Olazaran, M. A sociological history of the neural network controversy. *Advances in Computers* **37**, 335-425 (1993).

18. Marcus, G.F., *et al.* Overregularization in language acquisition. *Monographs of the Society for Research in Child Development* **57** (1998).

19. Hinton, G. Aetherial Symbols. *AAAI Spring Symposium on Knowledge Representation and Reasoning* Stanford University, CA (2015).

20. LeCun, Y., Bengio, Y., & Hinton, G. Deep learning. *Nature* **521**, 436-444 (2015).

21. Razeghi, Y., Logan IV, R.L., Gardner, M., & Singh, S. Impact of pretraining term frequencies on few-shot reasoning. *arXiv* 2202.07206 (2022).

22. Lenat, D. What AI can learn from Romeo & Juliet. *Forbes* (2019).

23. Chaudhuri, S., *et al.* Neurosymbolic programming. *Foundations and Trends in Programming Languages* **7**, 158-243 (2021).


WORLD SENSORIUM / CONSERVANCY
worldsensorium.com
@worldsensorium

I Just Want to Know What I'm Made Of

It's time to admit quantum theory has reached a dead end. Can we please go back to the math?

BY MICHAEL BROOKS

WATCHARA / SHUTTERSTOCK

I HAVE BEEN IN LOVE with quantum theory since before I started my Ph.D. in the subject over 30 years ago. Suddenly, however, I feel like we should maybe take a break.

The trigger for this quantum of doubt was a new paper. There's nothing particularly special about it; it's just a proposal for an experiment that might tell us something more about how the universe works. But, to me, it felt like the final straw. It has opened my eyes to the possibility that, without radical change, quantum physics may forever let me down.

Essentially, I just want to know what I am made of. One hundred years ago this year, the Danish physicist Niels Bohr received the Nobel Prize "for his services in the investigation of the structure of atoms." But I'm still waiting for a straight answer as to what the structure of the atoms that make up my body is. Quantum theory seems to promise an answer that it can't deliver, at least not in any way that I can comprehend. As Bohr once put it, "When it comes to atoms, language can be used only as in poetry."

Bohr is often fêted as the founding father of quantum theory and was one of the champions of its oddness. It's true that quantum is both mysterious and attractive—as enigmatic as the Mona Lisa's smile. However, there seems to be something troubling about the quantum grin. Whether probing it through theory or through experiment, we quickly arrive at an impasse: We can't use any of the results to tell us what stuff actually *is*.

> *We cover our ignorance with made-up stories known as "interpretations."*

The trouble starts with the workhorse of quantum theory: Erwin Schrödinger's famous wave equation. It assumes that all quantum stuff can be mathematically modeled as if it were a wave. Schrödinger's equation is a huge success: It allows us to predict, for instance, exactly what colors of light an atom will emit when stimulated with electromagnetic energy.

However, the beautiful utility of this equation, which is put to work by physicists every day around the world, masks its enigmatic silence about the actual nature of the atoms whose behavior it so successfully describes. While Einstein won a Nobel Prize for proving that light is composed of particles that we call photons, Schrödinger's equation characterizes light and indeed everything else as wave-like radiation. Can light and matter be both particle and wave? Or neither? We don't know.

WE COVER OUR IGNORANCE with made-up stories known as "interpretations." There is one classic laboratory setup that inspires most of these stories. Known as the double-slit experiment, it involves a single quantum object being fired toward a pair of openings whose physical dimensions are tailored to the mass and velocity (combined as momentum) of the object. At the far end of the experiment is a detector that records the original object's final location.

We tend to interpret this experiment through the lens of Schrödinger's wave equation. To speak of an "object" that we "fire" is to picture it as a particle. But we ascribe it a wave-like character to explain the ensuing observations. So we say that a quantum object can pass through both slits simultaneously, just as a water wave would. The object then emerges from the slits as two separate waves. As these travel on toward the detector, they meet each other, creating a pattern that is characteristic of interacting waves.

This "interference pattern" is a comb-like pattern of alternating high and low density of electron impacts as you look from left to right across the detector. Waves—water waves, for example—produce interference patterns. And so, we reason, there is definitely something wave-like about the quantum object.

However, the interference pattern does not appear at the detector straight away. The single quantum object manifests in the detector as a single particle at a single location. We then repeat the process, and the next object manifests at a different location on the detector. After a million detections, say, we see a clear interference pattern.

The Nobel Prize-winning physicist Richard Feynman once said that this experiment contains the only mystery of quantum physics. If the injected objects were just particles flying through slits like well-struck golf balls, there would be no interference pattern. How do these individual objects "know" that they are to be part of a building interference pattern? What causes a single particle to be detected in the high-density area, but never in the forbidden zones?

Although the conventional story about the double-slit experiment is that the single electron, say, goes through both slits at once in a wave-like "superposition" state, you don't have to accept this.

Bohr said you can be entirely agnostic about it and accept that there's nothing we can say about the electron after we've fired it at the slits and before the detector has told us its final position. This is one form of what has come to be known as the Copenhagen interpretation of quantum mechanics.

But you also don't have to accept that the electron goes through the slits you provided—not exactly, anyway. Some adherents of the many-worlds interpretation would claim that the electron goes through *both* the left and right slits, but in different versions of reality. The interference pattern is the result of interference

Physicists choose which quantum theory they subscribe to largely by taste.

between these different realities. However strange this seems, it can be a useful way of thinking. This is how Oxford physicist David Deutsch was able to conceptualize the blueprint for quantum computation, for instance.

Another alternative is to suppose there are hidden factors that we can't quite access. David Bohm influentially suggested that a particle is accompanied by a "pilot wave" that guides its trajectory through the double-slit experiment and creates the interference pattern. Most physicists will tell you that this kind of "hidden variable" interpretation of quantum theory has been ruled out by a combination of experimental results and mathematical proofs.

Or what if we live in a universe where the person doing the experiments doesn't realize that they are actually limited in the choices they can make about how the experiment will run? This is "superdeterminism," which is not quite an interpretation of quantum theory—more a framing of how the universe works from which experimental quantum observations might emerge. In this scheme, there are unappreciated connections between the quantum states of the experimental apparatus, including the mechanism for injecting the electron and the mechanism for detecting it.

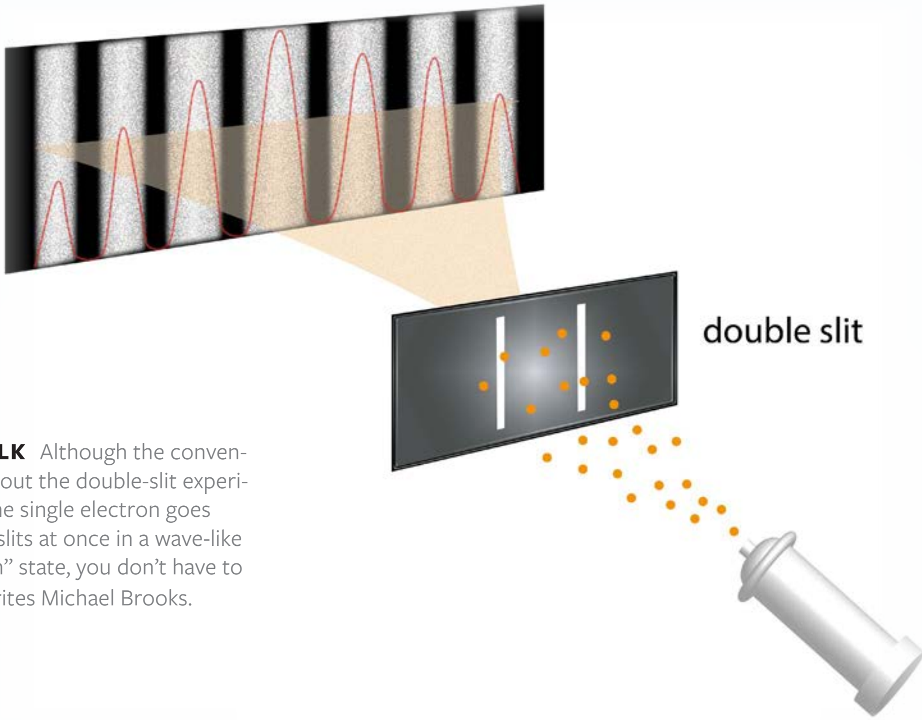
These hidden connections are akin to a set of hidden strings that connect all the equipment, making it impossible for the experimenter to know whether the detector settings have affected the electron's properties, and thus altered the experimental scenario. With superdeterminism, it becomes impossible for the experimenter to determine exactly how the experiment is done.

All of these scenarios have another implicit assumption: that cause always precedes effect. Our experience of time is that it runs in only one direction, but that is not a constraint that quantum objects experience. That means we can interpret the double-slit experiment via "retrocausality," whereby counterintuitive flows of action mean that what seem like later effects actually create earlier conditions.

Some suggest that an electron's state might be affected by both its past state and its future state. This would provide a way for each electron to create a sliver of the detector's interference pattern, despite ostensibly not having a way to "know" where it should end up.

The wide range of interpretations proposed by physicists and philosophers—there are many more than I have mentioned here—demonstrates that there is no right answer. Interested physicists choose which one they subscribe to largely by taste.

IT WOULD BE REASSURING to think that further experiments will allow us to choose between them. However, there is little reason to think that will ever be the case. Some experiments by Aephraim Steinberg of the University of Toronto seem to support the idea of pilot waves, for example. Steinberg is often cited as having revealed so-called Bohmian trajectories that quantum objects follow through the double-slit experiment. But he demurs. "I don't wish to think of our experiments as measuring Bohmian trajectories," he has said. Instead, he concedes, "they 'naïvely' measure properties of quantum systems which I expect any underlying explanation to reproduce."



DOUBLE TALK Although the conventional story about the double-slit experiment is that the single electron goes through both slits at once in a wave-like "superposition" state, you don't have to accept this, writes Michael Brooks.

MAGNETIX / SHUTTERSTOCK

In other words, this experimental evidence remains open to interpretation. For some interpretations, there is no experimental evidence. The many-worlds idea, for instance, allows no meaningful interaction between the worlds, which means there will never be any way of telling empirically whether it is a reasonable point of view.

New experiments come along every now and then, but they rarely make any difference. And that includes the idea that broke me. In late 2021, Colm Bracken of Maynooth University in Ireland, Jonte Hance of the University of Bristol, and Sabine Hossenfelder of the Frankfurt Institute for Advanced Studies proposed an experimental setup that, to their minds, could distinguish between superdeterminism and retrocausality. It involves the double-slit experiment with some added bells and whistles that attempt to engineer a logical paradox in the concept of backward-in-time influences.

But the researchers concede that the result will make no difference to anyone who believes in many-worlds, Copenhagen, or Pilot-Wave Theory. It only reduces the range of options by one, at most. And retrocausality proponents doubt that it does even that. In short, it doesn't get us anywhere. Like all of the other experiments. Have I been chasing rainbows?

Probing quantum physics is really, really difficult. It is, by necessity, an experimental science far removed from the scale of human experience. But it is also instructive to note that a century of technological progress has not helped answer my question. We essentially know nothing more than Bohr, Einstein, Schrödinger, and Heisenberg did about the actual nature of reality. We have created interesting technologies and philosophical conundrums, but little else. Is it actually possible to find out what the universe is made of?

My hope is that it is. But I suspect that it might require a radical pivot. Maybe the only way forward is to go back: to revisit the mathematics behind quantum theory.

It has worked before. When science has been stuck, people have opened up hidden paths by inventing new ways of manipulating numbers or by borrowing mathematical techniques from other fields. Newton and Leibniz's calculus, Babylonian algebra, and the Chinese and Indian invention of negative numbers provide good examples. As does general relativity: Einstein solved his roadblock over describing curved spacetime by borrowing tensor calculus, which had been developed by pure mathematicians with no application in mind. So, is novel mathematics the solution that quantum physics didn't know it needed?

We know nothing more than Bohr, Einstein, Schrödinger, and Heisenberg.

"It is quite possibly the case that we need new mathematics to crack this problem," Hossenfelder said. She thinks the solutions might come from getting a better handle on the mathematics of chaos theory. Her collaborator in investigating superdeterminism, Oxford's Tim Palmer, thinks that quantum mechanics arises from a deeper theory that involves chaotic attractors. Or, she suggests, it might come from new approaches to escaping, as quantum stuff does, the tyranny of time's arrow. "All our current theories work with differential equations, where you have an initial state and then an evolution law," Hossenfelder explained. "I don't think this is going to work for whatever the theory is that underlies quantum mechanics."

To this end, Hossenfelder has been eschewing Schrödinger's wave equation and working with a mathematical approach called path integrals, which are more flexible about time's flow. "I don't know if this will work out in the end though," she admits. "Maybe we indeed need something else entirely."

University of Western Ontario physicist Emily Adlam is also open to the need for new math. She thinks that the best path forward might be to ditch the reductionist approach of positing ever-smaller components. Instead, she suggests that we should try to understand the quantum world in a more holistic way. "The mathematical methods we currently use in physics do not seem adequate for this kind of holistic description, so it's likely that something new will be needed," she told me.

But, she points out, it doesn't have to be brand new. "There is a lot of mathematics that has been developed by the pure mathematicians that hasn't yet found any application in physics," she said. "So I think it's actually very likely that the mathematics we need does already exist in the pure mathematics world, and we just haven't seen how to apply it to physics yet." In other words, if it worked for relativity, maybe it can work for quantum theory.

Full disclosure: Plenty of scientists disagree with me on this. Oxford's Vlatko Vedral, for instance, is skeptical. "It seems to me that the mathematics needed for quantum physics—linear algebra—works perfectly well," he told me. He thinks, like Bohr, that any problem comes from the limitations of language. "Neither the connective 'or' nor 'and' describe what we mean by a quantum superposition," he pointed out. "Perhaps, rather than our mathematics, it is our language that needs to evolve to properly reflect the quantum world."

Deutsch, too, seemed impatient with my dissatisfaction over our understanding of the nature of reality. "Someone might equally well say: We may know what dogs look like and how they behave, but we don't know what a dog 'actually is'," he said. Dogs, he says, are collections of atoms, and an atom is a collection of quarks and leptons, and those can be described by quantum fields. What more does anyone want?

Even Hossenfelder is not fully convinced. "I don't think it's the task of science to answer the question what something 'is'," she told me. She thinks science has only to do with describing what we observe.

What am I? I'm a cluster of tiny indescribable quantum objects waiting for my new crush, mathematics, to reveal the truth.

MICHAEL BROOKS is a United Kingdom-based science writer. His most recent book is *The Art of More: How Mathematics Created Civilization*.

Why I Couldn't Get Over My Brother's Death

Everybody told me my grief would relent in a year. It only got worse. Was there something wrong with me?

BY MARY X. DENNIS

1. LOSS

One cold, early spring morning, I waded out into Cayuga Lake at the edge of Ithaca, NY to drown myself. I stepped cautiously over the ice-slick slabs of shale, careful not to slip before I reached the water. As my rubber boots filled with icy slush, the shock of the cold made my heart jump. *Was I really going to do this?* I willed myself forward. I wanted to feel my limbs turn to frozen weights, like my brother's had. I wanted to feel my lungs fill with liquid ice, like my brother's had. I couldn't stand thinking about his death anymore, over and over, day after day. I had questions and I was desperate for answers. I had to know, did the panic leave his body before his soul? I also wanted to see him, at the bottom of the lake. Was he perfectly preserved by the icy water, like my dad insisted, or already just a pile of bones?

I barely got waist deep before I turned back to the shore and flung myself into the snow. Ashamed of myself for, almost a year later, being completely unable to move past my brother's death. Even when I wasn't in the lake, I was in the lake. I was always in the lake. Frozen.

My brother drowned the week before he was to graduate from Cornell, in a freak accident. He'd taken a canoe out at dawn, and it had somehow capsized. We don't know how. We don't know why. But it was May, when the water temperatures are still freezing, so we suspect he got hypothermia before he could swim to shore. We searched for two weeks, the entire lake crawling with boats and buzzing with helicopters. Then we stopped the search and held a memorial. My entire world folded in on itself. Every day I woke up into a nightmare. A rollercoaster of nausea and horror. People told me to give it time. "Time heals all wounds," the old saying goes. So I did. I gave it time.

During that first year, I tried everything I could to force the image of him sinking into the black ether, face blue, hair waving gently, eyes wide open, out of my mind. I made pottery and then smashed it. I patiently sat with jigsaw puzzles and coloring books. I went into the forest with a chainsaw, cut down invasive species and set them on fire. I devoted hundreds of hours to punching pads and round-housing bags at the local mixed martial arts gym. I zombie-watched TV shows and absently scrolled the internet. Anything to just let time tick by. One day, some day, I was going to wake up feeling better, more like myself, less like I'd woken up in some brutal alternate reality. *Right?* But time wasn't going to resurrect my brother, and it didn't resurrect my old sense of self. I did not magically start to feel better. The more time passed, the more I missed him. It was like holding my breath. The longer I did it, the more desperate I was for air.

It wasn't until four years after my brother disappeared that I had some answers about why I was losing the thread, and some hope that I would be able to find my way back to myself.

2. GRIEF

We all lose people we love along the way. In most cases, the passage of time is enough to get us through the acute stages of sorrow, when every reminder of the loved one is like a knife to the heart. When each of my grandparents died, I was devastated and I missed them, but I was not consumed by longing for their return. They all died peacefully, at the end of long, full lives. Most of us, after a person we love passes on, can gradually accept that they are gone, experience joy again, and find meaning in our favorite pursuits. But when my brother died, I got stuck. My grief became a trap.

For a long time, I resisted seeking help because I believed—I still believe—that my reaction to my brother's death was a normal human reaction to a horrifying loss. The idea of branding my grief as abnormal—as something that needed treatment—was repellant to me. But was feeling suicidal normal for me? No. Definitely not. Did I want to go on feeling this way? Also, no.

A year after my brother's death, I was told I had post-traumatic stress disorder, so I began to see therapists who specialized in PTSD. They focused on desensitizing the event—asking me to repeat the facts of my brother's disappearance over and over. How he'd been camping with friends right before his graduation and gone out alone in an old canoe at dawn to watch the sunrise, never to be seen again. The week we spent looking for him, alongside hundreds of his friends and our community members. The scuba divers, fishing boats, search dogs. The memorial where we all walked in a silent row by the lake, staring out across the water. The pair of his shoes that still sat by the back door of our parents' house. The fear—the *dread*—as bottomless as the lake—that he was abducted by aliens or a serial killer, that he was still alive somewhere and suffering horribly—would make my heart lurch as I tried to fall asleep at night. How do you learn to accept this kind of thing? How do you ever learn to live with it? How is telling a therapist about it, again and again, going to change how horrifying it is?

Sitting in these nondescript offices on dull colored couches with people I didn't know, trying to talk about the most profound loss I had known made me cry uncontrollably. I quickly grew to dread it, and then avoid it. I never stuck it out for longer than three or four sessions before quitting, too afraid of the tears to continue. Too angry at these therapists, sitting there nodding, staring at me dry-eyed while I sobbed, pretending they had *any idea at all* what I was experiencing.

My list of failed therapists kept getting longer and my life kept spiraling further and further away from me. I no longer seemed to care about the consequences of my actions. Then about a year after I walked into the lake, I got pregnant. Suddenly, I couldn't just wait around until I felt better. I needed to get better now. My baby needed me to get better.

I committed to making PTSD therapy work. This time, some of it was helpful, particularly a treatment approach called somatic experiencing,

PAGES 42-43: MARY LONG / SHUTTERSTOCK

Where does the enormous part of you that loved the deceased person go?

where I focused on the feelings in my body—*My heart feels like it's about to explode. The top of my head is buzzing*—and learned how to regulate them with things like breathwork, sticking my head in the freezer, or "focusing on the room." I also tried meditation, acupuncture, massage, Reiki, and gardening. Anything to get my life and my emotions back to a manageable place. But nine months later, I had little to show for any of it. After my daughter was born, and I'd finished breastfeeding—now three years past my brother's death—I turned to pharmaceutical medications. But SSRIs just sent me on a rollercoaster of anxiety. Benzodiazepines had a kickback that made me cry harder when they wore off.

The day my brother disappeared, I'd been sitting in the graduation hall at Columbia University holding a fresh master's degree, excited for everything the world had in store for me. Now, I was a mentally unstable single mom who lived in fear of her own tears. I was overwhelmed and confused by a life that had become a cardboard cut-out version of the real one I had left behind and I had no idea how to fix it. I gave up on finding treatment that would help me.

3. RESEARCH

The idea that grief can take two separate forms—one that resolves itself organically and one that endures—is at least as old as Sigmund Freud. In his 1917 book, *Mourning and Melancholia*, Freud wrote that grief resulted in one of two conditions. Mourning was a healthy form of grief over the loss of a loved one, a process of which one was consciously aware. Melancholia, on the other hand, was an unconscious, pathological form of grief that was difficult for the individual to understand and extended beyond the scope of the loss itself to other areas of life. Freud influenced grief research and treatment for decades to come.[1]

FROZEN When author Mary X. Dennis' brother died in a lake accident, she couldn't get the image of him drowning out her head. "Even when I wasn't in the lake, I was in the lake," she writes. "I was always in the lake. Frozen."

In 1944, psychiatrist Erich Lindemann published a landmark paper on the symptomatology of "acute grief" and coined the phrase "morbid grief" to describe some of the more extreme cases of grief he studied.[1] Lindemann also developed the term "grief work" to describe the process of readjusting to a world in which the loved one is missing, a process that he found was similar in some ways to that of a patient adapting to a lost limb.[2] Then came Elisabeth Kübler-Ross, a Swiss-American psychiatrist whose 1969 book *On Death and Dying* popularized her theory of five stages of grief: denial, anger, bargaining, depression, and acceptance. Another watershed theory was developed around the same time by British psychiatrist John Bowlby, who proposed that one's experience of grief can be shaped by one's childhood attachment type: Secure? Anxious? Disorganized? Avoidant?[3] Two camps had begun to develop: those who thought grief was a painful, poignant part of the human experience, and those who felt it should be treated as a mental health disorder or medical condition.[4]

Then in the early 2000s, a team of scientists set about developing a clinical definition for debilitating grief. Researchers at Yale University, Utrecht University, and Oxford University had been tracking grieving individuals in the first two years after their loss. Their studies found that while grief "normally" resolves on its own—in about 90 percent of cases—in the other 10

ERVIN-EDWARD / SHUTTERSTOCK

percent, it persists.[5] Soon after that, the Yale group submitted a paper proposing prolonged grief disorder (PGD) as its own psychiatric condition.[6]

PGD is defined as severe grief that endures longer than a year after the death of the loved one.[6] It is characterized by intrusive thoughts and yearning for the deceased,[1] and is associated with suicidality.[7] It is psychologically and neurobiologically distinct from acute grief, major depressive disorder, generalized anxiety disorder, and PTSD,[8] and is largely understood as an attachment disorder.[9] Like me, people with PGD tend to suffer a prolonged and profound loss of meaning and disruption to their sense of identity.[1] Where does the enormous part of you that loved the deceased person go?

The initial studies used to develop the criteria for the disorder had focused primarily on elderly widows whose spouses died of "natural" causes.[5] But since then, researchers have found if the death is "unnatural"—sudden, traumatic, or involving a child[10]—or if the bereaved is unable to get closure,[11] say because the loved one's body was never found, the risk of developing PGD can almost quintuple.[10] As I read more about PGD, it began to speak to me.

4. ADDICTION

Over the past decade, as neuroscientists grappled with what was happening in the brains of people who suffer long-term grief, they made a discovery. While PGD often contains some elements of PTSD, depression, and anxiety, it is most similar to addiction.[6] Like addiction, prolonged grief tends to feature a conflict between craving and avoidance: Yearning for the deceased loved one alternates with an avoidance of painful memory triggers.[9]

The foundational studies proposing the connection between PGD and addiction were published in 2008 by Mary-Frances O'Connor, a psychiatrist who directs the Grief, Loss and Social Stress (GLASS) Lab at the University of Arizona. O'Connor ran fMRI brain scans on women suffering from prolonged, unabated grief who had been triggered with stimuli related to their deceased loved ones. The scans showed activity in the brain region central to the neural reward pathway activated by addiction.[6]

It is no secret that love and loss affect the reward system. Attachment figures help regulate our psychological and physiological responses to stress, providing psychological safe haven in times of uncertainty. They bring us both joy and sustained happiness. Neurobiologically, contact with the loved one, or even thinking of them, can result in a release of oxytocin as well as natural opioids. Over time, our brains come to associate the loved one with these rewards.[9] This is very similar to what happens to an addict who is physically accustomed to receiving a boost from alcohol or drugs. If you take away the source of pleasure, the receptors in your brain are thrown into a state of crisis. Whether the original source was a human being or a drug, the result of its absence is one and the same: a deep sense of yearning, or craving. When what used to bring you pleasure starts to cause you pain—for instance, memories of a deceased loved one who will not return, or hangover from a harmful drug—you can slip into the vicious cycle of yearning and avoidance.

There is still no pharmacological treatment recommended for PGD. In clinical trials, tricyclic antidepressants and benzodiazepines have so far failed to work. But researchers are beginning to study the effects on PGD of Naltrexone, a drug that has been effective in reducing craving in people with opioid and alcohol dependence.

For now, the only treatment for PGD proven to be effective is a form of grief therapy—a 16-week course—developed by Katherine Shear, a professor of psychiatry at Columbia University. Shear's course does not focus on the grief itself, but rather on helping a grieving individual adapt to the loss. Her work is rooted in attachment theory, which emphasizes the biological importance of a person's relationships with a small handful of individuals—relationships essential to a person's well-being. When one of those important people is lost, it is critical to distinguish between the loss of the *person* and the loss of the *relationship*. Shear says grief is the "form love takes when someone dies." The love does not simply disappear, and for the bereaved, the relationship doesn't have to, either. To prevent the bereaved from losing the vital attachment relationship completely (which can stimulate yearning), Shear helps people to remodel it in their minds—to internalize it.

Shear describes places where people get "stuck" in this process of adapting to loss, like avoiding reminders that the person is gone in order to suppress the associated emotion or focusing on all the ways the death could have simply *not* happened. These roadblocks are all natural defense mechanisms we throw up to protect ourselves from stress, but when they go on too long and become reflexes, they can keep us from healing.

5. BREAKTHROUGH

Just after my daughter's first birthday, I visited a psychologist who had me take a very long multiple-choice test called the Brief Symptom Inventory. It featured questions like, "Do you feel like someone is always watching you?" "Do you feel like everyone is out to get you?" The day after taking the quiz, I spent several hours explaining to the psychologist everything that had happened to me in the few years since my brother had disappeared.

After considering the results of that test, the personal history I had recounted for him, and his own reading of scholarly articles and the DSM, the psychologist diagnosed me with Persistent Complex Bereavement Disorder, a precursor to PGD.[8] He recommended not PTSD therapy, but grief therapy. This diagnosis was pretty uncommon and finding appropriate treatment was more complicated than simply checking a box in an online search bar. He recommended exactly one person: a woman named Karin.

I went back to therapy, and this time I stuck with it. Karin has never heard of Shear's 16-week course—but she has over 30 years of grief therapy under her belt. As I worked with Karin, I realized her approach had a lot in common with Shear's. Karin didn't focus on the event of my brother's death, the way the PTSD therapists had. She focused on *him*—not on the trauma of his death, but on the relationship I had lost. My identity as his big sister. Things we liked to do together. The ways in which we were similar. My favorite version of myself was the person my brother saw me as, and she helped me to realize that even though he's no longer here, I am still that person. For the first time, I felt like I was talking to someone who understood how I felt.

"That's progress. You went there, and then you came back."

Around the time I began seeing Karin, I also stumbled on a drug that helped a little: gabapentin. Gabapentin is a seizure medication that is prescribed off-label for anxiety.[12] Someone I know who couldn't take benzos due to a history of addiction had been prescribed gabapentin instead, and suggested I try it. I found a psychiatrist who put me on a tiny dose of it three times a day. It didn't make me woozy or sleepy like the benzos, and it didn't make me jittery like the SSRIs. I wasn't overtly conscious of any difference at all. But after a few weeks on it, I realized I hadn't been sneak-attacked by a single the-day-is-now-over type of complete meltdown I'd gotten so terrified of encountering. The pain was duller. The tears less frequent. Caring for my infant daughter felt more manageable. Everything sharp and painful felt a little softer.

Gradually, crying stopped feeling so terrifying. Less like driving over the edge of a cliff into freefall. It started to feel safe. I still cried, but I could also *stop*. In therapy, Karin forced me to stand at the edge of the dark whirlpool of grief I had taught myself to run from at all costs, and look at it, little by little, in her office, where she could reach out and grab me before I was beyond the point of no return. She would ask me a difficult question, and once I'd gotten deep into my answer, once I had opened that door, she would stop me and force me to tell her all about the painting on the wall, or the snack I had just eaten. Once I was safely back in the world of mundane pleasantries, Karin would smile and throw me out of her office. "That's progress," she explained. "You went there, and then you came back. And now you're going to get on with your day."

One day, in the middle of therapy, I realized I was telling a story about my brother. Out loud. How I'd lost our train tickets when he'd visited me in China, and we'd had to ride five hours in the standing-room-only cattle car. All the migrant farmers had crowded around him to read his palm. He'd just gone with it, letting them pass his hand around like a newspaper, tracing over his palm with their fingers. I would have lost my mind; I had already climbed into the overhead luggage rack to get away from the crowd. But he saw the world with brighter eyes than I do. He was so much nicer than me. I was laughing when I caught myself and stopped mid-sentence. I braced for impact, waiting for that familiar feeling, for my face to crumple and my chest to cave in. For the first time in years, it didn't happen. Suddenly, the person I loved most in the world, the person more like me than anyone else, was more than just a tragedy. I could remember him again.

I'm no longer afraid of the tears. Because I know I can get them to stop.

6. CHRIS

Chris, a toddler, sobbing when we put the lobsters he had been playing with into the pot to cook. Chris, with his first video camera, filming himself making funny faces. Chris making billions and billions of nature and science videos, narrated in spot-on impressions of David Attenborough and Carl Sagan. Chris showing up late to the family reunion because he'd locked his stuff in a bathroom stall so he could wander around the airport filming and had come back to his bags surrounded by the bomb squad and gotten detained. Chris, the flying squirrel, climbing up something just so he could leap off it. Chris throwing his first backflip. Chris winning practically every freestyle ski competition he ever entered.

Chris, a freshman in college, getting on a plane to Haiti after the earthquake and raising thousands of dollars for NGOs with the videos he made for free and posted to GoFundMe. Chris convincing me to come to Haiti with him to plant bamboo to create a building material and top-soil retention system all in one. Chris sitting on a fire escape in Brooklyn, playing the guitar and singing me a song he's written, a shockingly beautiful song about people on the subway being alone together. Chris asking me to cut his shaggy hair in the middle of a heat wave and laughing at how bad I did. That huge pile of hair, carelessly thrown into the trash. What I wouldn't give for just one strand of it now.

Being able to think about Chris again doesn't mean I've stopped grieving him. As I'm writing this, I'm crying. I'm still profoundly sad about what happened to my brother. All the time. Deep in my bones. That's never going to change. But I'm no longer afraid of the tears. Because I know I can get them to stop. I still have a bottle of gabapentin in my medicine cabinet, but I no longer feel like I need it. I still talk to my therapist, but it's less frequent and much less intense. I still yearn for a world made complete, full of vibrant light and joyful love, by his presence. But this yearning no longer overshadows the rest of my life.

And I'm so grateful for that. Both for myself and for my daughter, who now has a stable, loving, happy mother she can depend on. Sometimes, I tell her about her Uncle Chris, how she will never meet him but how much he would have loved her, and how they would have gotten along so well. I can tell her stories about him. I can explain to her both who he was and what happened to him without dissolving.

I can hold that love in my heart, and I can look at it. My brother is dead. It has taken me years, and lots of therapy, and drugs, to learn how to manage the pain of not having him in my life, to accept the reality that he is forever denied the life that should have been his. Learning to accept this, to live with this, does not mean my love for him will ever, *ever* fade. Wherever he may be, I will always be proud to be his big sister.

MARY X. DENNIS is a writer living in Ithaca, NY.

REFERENCES

1. Maercker, A. & Znoj, H. The younger sibling of PTSD: Similarities and differences between complicated grief and posttraumatic stress disorder. *European Journal of Psychotraumatology* **1**, 10.3402/ejpt.v1i0.5558 (2010).

2. Kelly, M.P. Loss and grief reactions as responses to surgery. *Leading Global Nursing Research* **10**, 517-525 (1985).

3. Cassidy, J. & Shaver, P.R. (Eds.) *Handbook of Attachment, Third Edition: Theory, Research, and Clinical Applications*. The Guilford Press, New York, NY (2018).

4. Kopelman, L.M. Normal grief: Good or bad? Health or disease? *Philosophy, Psychiatry, & Psychology* **1**, 209-220 (1994).

5. Lundorff, M., Holmgren, H., Zachariae, R., Farver-Vestergaard, I., & O'Connor, M. Prevalence of prolonged grief disorder in adult bereavement: A systematic review and meta-analysis. *Journal of Affective Disorders* **212**, 138-149 (2017).

6. Prigerson, H.G., Kakarala, S., Gang, J., & Maciejewsk, P.K. History and status of prolonged grief disorder as a psychiatric diagnosis. *Annual Review of Clinical Psychology* **17**, 109-126 (2021).

7. Latham, A.E. & Prigerson, H.G. Suicidality and bereavement: Complicated grief as psychiatric disorder presenting greatest risk for suicidality. *Suicide and Life Threatening Behavior* **34**, 350-362 (2005).

8. Maciejewski, P.K., Maercker, A., Boelen, P.A., & Prigerson, H.G. "Prolonged grief disorder" and "persistent complex bereavement disorder," but not "complicated grief," are one and the same diagnostic entity: An analysis of data from the Yale Bereavement Study. *World Psychiatry* **15**, 266-275 (2016).

9. Kakarala, S.E., *et al.* The neurobiological reward system in Prolonged Grief Disorder (PGD): A systematic review. *Psychiatry Research: Neuroimaging* **303**, 111135 (2020).

10. Djelantik, A., Smid, G.E., Mroz, A., Kleber, R.J., & Boelen, P.A. The prevalence of prolonged grief disorder in bereaved individuals following unnatural losses: Systematic review and meta regression analysis. *Journal of Affective Disorders* **265**, 146–156 (2020).

11. Diolaiuti, F., Marazziti, D., Beatino, M.F., Mucci, F., & Pozza, A. Impact and consequences of COVID-19 pandemic on complicated grief and persistent complex bereavement disorder. *Psychiatry Research* **300**, 113916 (2021).

12. Fukada, C., Kohler, J.C., Boon, H., Austin, Z., & Krahn, M. Prescribing gabapentin off label: Perspectives from psychiatry, pain and neurology specialists. *Canadian Pharmacists Journal* **145**, 280-284.e1 (2012).

MASSIMO TODARO / SHUTTERSTOCK

Deepfake Luke Skywalker Should Scare Us

The power to create convincing deepfake icons could destabilize society

BY ALAN JERN

T**ELL IT TO ME** in *Star Wars*," Tracy Jordan demands of Frank Rossitano in an episode of *30 Rock*, when Frank tries to explain the "uncanny valley." As Frank puts it, people like droids like C-3PO (vaguely human-like) and scoundrels with hearts of gold like Han Solo (a real person), but are creeped out by weirdly unnatural CGI stormtroopers (or Tom Hanks in *The Polar Express*). More than a decade after that scene aired, *Star Wars* is still clambering out of the uncanny valley: Most fans were thrilled to see a digital cameo from a young Luke Skywalker in *The Mandalorian*. But he—along with recreations of young Princess Leia and Grand Moff Tarkin from the original *Star Wars* film—didn't look real, which for some was unnerving.

These are all examples of what have come to be known as deepfakes, a catch-all term for synthetically manipulated or generated images or videos of people or just about anything. The name derives from the fact that most methods for creating deepfakes rely on deep neural networks, the same basic architecture that underlies most of the AI we encounter on a daily basis, like face and voice recognition. And because the technology is constantly and rapidly improving, so are the deepfakes. Deepfake Luke recently appeared again in *The Book of Boba Fett*, just about a year later, looking impressively realistic.

Currently, one of the most advanced methods for creating deepfakes is called a generative adversarial network. It's actually two separate neural networks paired together: One network might generate fake faces (the generative part) and the other tries to discriminate the fake faces generated by the first network from a set of real faces in a database. Like training with a partner, over time, both networks get better together at their respective tasks (the adversarial part). The discriminator network gradually gets better at telling the fake faces from the real faces. And the generator network gradually gets better at fooling the discriminator by generating more realistic fake faces.

Great for the scrappy little generator network. Good for *Star Wars* fans and the now immortal Luke Skywalker. Possibly bad for society.

Creating digital actors for TV shows is relatively harmless. But the power to create convincing deepfakes of world leaders, for example, could severely destabilize society. An inflammatory and well-timed deepfake video of a politician could hypothetically swing an election, spark riots, or raise tensions between two countries to volatile levels.

Whether this is a serious concern yet depends on how good current deepfake technology is and how good people are at discerning fakes from the real thing. In two recent studies, researchers asked thousands of people to judge the authenticity of real and fake images and videos to try to answer this question.

THE MOST RECENT STUDY, by psychologist Sophie Nightingale of Lancaster University and computer scientist Hany Farid of the University of California, Berkeley, focused on deepfake images.[1] In one online experiment, they asked 315 people to classify images of faces, balanced for gender and race, as real or fake. Shockingly, overall average accuracy was near chance at 48 percent, although individuals' accuracy varied considerably, from a low of around 25 percent to a high of around 80 percent.

In a follow-up experiment, 219 different people did the same task but first learned about features to look for that can suggest an image is a deepfake, including earrings that don't quite match, differently shaped ears, or absurdly thin glasses. People also received corrective feedback after submitting each answer. "The training did help, but it didn't do a lot," Nightingale told me. "They still weren't performing much better than chance." Average accuracy increased to only 59 percent.

And those bizarre telltale deepfake features that people learned to spot? As technology improves and deepfakes become ever more flawless, "these artifacts are going to get sparser," Nightingale said. "And we're going to be back to that initial [accuracy] level, in not too long."

Human and computer perception are fundamentally different.

But there's a difference between an image and a video. Part of the reason it remains so difficult to make a believably realistic recreation of young Luke is because he has to emote and speak. Even in *The Book of Boba Fett*, producers clearly recognized the limits of their illusion, frequently cutting away from Luke's face whenever he had extended dialogue.

In December 2021, a team of researchers from MIT and Johns Hopkins, led by Ph.D. student Matt Groh, ran a similar study of people's accuracy at identifying deepfakes, this one focusing on videos.[2] They set up a website where people could rate the authenticity of real and deepfake videos randomly selected from Meta's Deepfake Detection Challenge dataset. Like in the study of images, people watched videos, one at a time, judged whether they were real or fake, and received corrective feedback. More than 9,000 people around the world found the website on their own and collectively completed more than 67,000 trials. What's more, the researchers recruited 304 nationally representative online participants who collectively completed about 6,400 trials.

Overall, average accuracy (weighted by people's confidence judgments) was between 66 percent and 72 percent, depending on the participant group—slightly higher than in the study of images, but the stimuli in the two studies varied in multiple ways, so it's hard to directly compare them. To get some sense of whether these accuracy numbers were any good, the researchers compared people's judgments to those of one of the best performing computer models for detecting deepfakes. After feeding the same videos that people saw into the model, it beat most people, achieving an accuracy of 80 percent.

The best performer of all, though? One that incorporated human *and* computer judgments. To accomplish this, the researchers asked people to make a second rating of each video. After their initial rating, the researchers

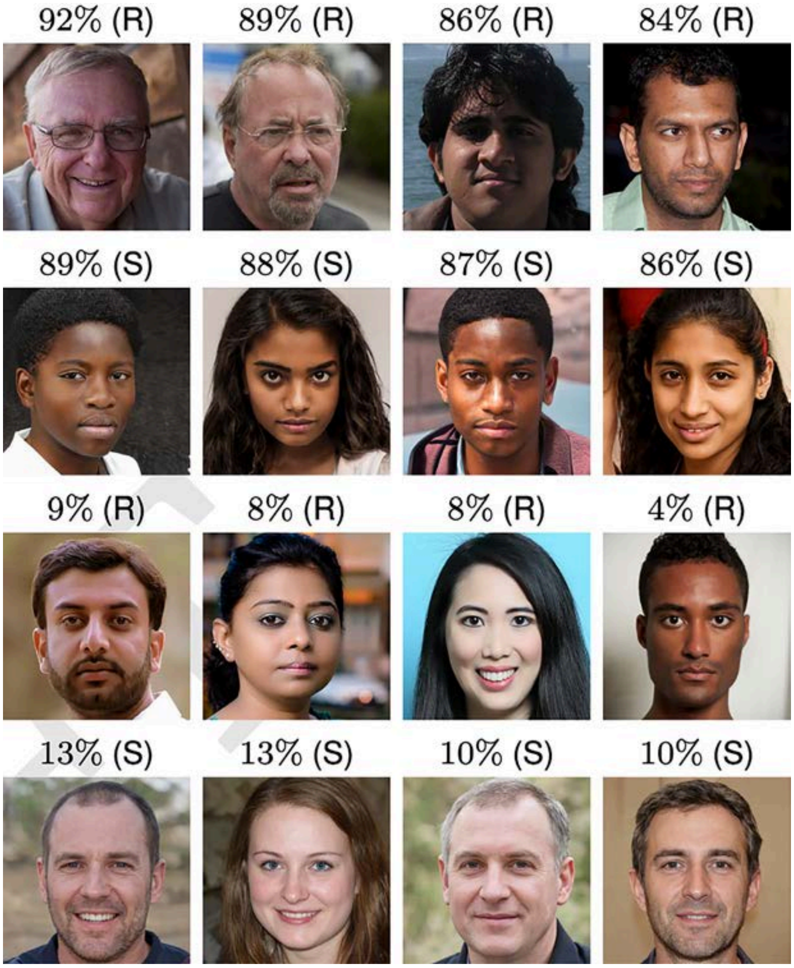
The computer-generated faces tend to be more symmetrical, which we are biased to trust.

showed people the likelihood assigned by the computer model that the video was a deepfake and asked people if they wanted to change their ratings; they often did. The average of these updated ratings slightly out-performed both the model and people's original ratings. This result suggests that both people and the computer model brought unique strengths to the task of detecting deepfakes.

At this point, you might be thinking that people are just irredeemably bad at this task. But many researchers believe that we are basically expert face detectors. Some neuroscientists have argued that we even have a brain region specialized for face processing. At the very least, it's clear that people process faces differently than other types of visual information. We're such expert face detectors that we can't help but see faces everywhere: in rock formations on Mars, in the headlights and grilles of cars, and in misshapen vegetables.

Groh, the study's lead researcher, told me that these observations are exactly what motivated them: Because people seem to process faces in a specialized way, they might not be as easily fooled by deepfakes involving faces compared to other kinds of misinformation. And the strengths that the computer model brings? For one, it doesn't suffer from the same limitations as our visual systems: "On videos that are super dark or blurry, that's not really so much of an issue for the computer," Groh explained.

This is just one example of how human and computer perception is fundamentally different. Clearly people make mistakes that computers don't. And yet sometimes computers make mistakes people would never make. AI mistakes are even an active area of research. For example, some people work on so-called adversarial images for computer vision systems: specially designed images that break image-recognition models. An adversarial image might be a photo of a



WHO IS REAL? The deepfake generator gets increasingly better at making realistic faces until people are unable to tell which faces are authentic. The top two rows show the most accurately classified real and synthetic faces, and the bottom two rows show the least accurately classified faces.

NIGHTINGALE, S.J. & FARID, H. AI-SYNTHESIZED FACES ARE INDISTINGUISHABLE FROM REAL FACES AND MORE TRUSTWORTHY. *PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES* (2022).

lizard that an image-recognition model is 99 percent sure is a lizard. Then, after tweaking a handful of carefully chosen pixels in the image that are virtually imperceptible to any human, the model is suddenly 99 percent sure it's a snail.

People are also excellent at incorporating context and background knowledge into their judgments in ways that computers mostly aren't, and in ways that the sorts of tasks in these studies mostly don't assess. One exception is a pair of deepfake videos of Vladimir Putin and Kim Jong-un included in the video study. People were much better than the model at classifying these videos. Whereas a majority of participants correctly spotted these fakes, the model was way off, highly confident that they were real.

Here, people may have taken into account factors like their memories of previous statements made by these leaders, or knowledge of what their voices sound like, or the likelihood that they would say the sorts of things shown in the video: all things not included in the model at all.

People's sensitivity to these contextual factors underscores why even a flawless deepfake Luke in a future season of *The Mandalorian* wouldn't really fool anyone: Everyone knows that Mark Hamill doesn't look like that anymore.

Even if people have some advantages when it comes to processing faces, we have some liabilities, too. Nightingale and Farid conducted one more experiment to see if people might intuitively recognize the difference between deepfakes and real faces. They asked 223 people to rate the trustworthiness of 128 faces from the same set of images as in their previous experiments. People rated the deepfakes about half a point more trustworthy on a seven-point scale, on average, than the real faces.

Nightingale speculated the reason for the difference has to do with our well-documented preference for more average and symmetrical faces: "Think about how synthetic faces are generated: They're basically an averaging process," she said. "You take lots of images, lots of real images, you train a GAN [generative adversarial network], and then it uses that to produce synthetic faces." The resulting faces tend to be more average-like and symmetrical, which we are biased to trust.

Arguably, both studies overstate and understate the seriousness of the problem. On one hand, unlike in these studies, far less than half the faces we see online are deepfakes, so our real-world accuracy is bound to be much higher, even if we just assume everything is real. But, unlike in these studies, Groh pointed out to me that under normal circumstances, we're rarely so concerned with judging the accuracy of the media we encounter. This could explain why some people have already lost money to scammers using deepfakes, even though the stakes of these scams are much higher than in an online experiment.

Even worse, our ability to accurately judge real from fake may not even matter. For example, while people are reasonably good at judging the accuracy of news headlines, what people share on social media depends more on whether it agrees with their partisan beliefs than on whether it's actually true.[3] When it comes to deepfake misinformation, Groh told me, "if it's emotionally charged, and it's motivating, it's likely that people might share stuff without even thinking about whether they believe in it or not."

Is there anything that engineers and policy makers can do about the potential threat that deepfakes pose? Nightingale suggested developing ethical guidelines for using powerful AI technology or building watermarks or embedded fingerprints into the deepfake algorithms so they're easier to identify. However, she thinks there are likely other solutions. "That's why I'd like to almost get that idea out there that collaboration on what we do next is quite important," she said. "Because I want people to be aware of the problems."

"Have we passed through the uncanny valley?" Nightingale asked me. It was a rhetorical question. She said, "It turns out we're pretty much on the other side."

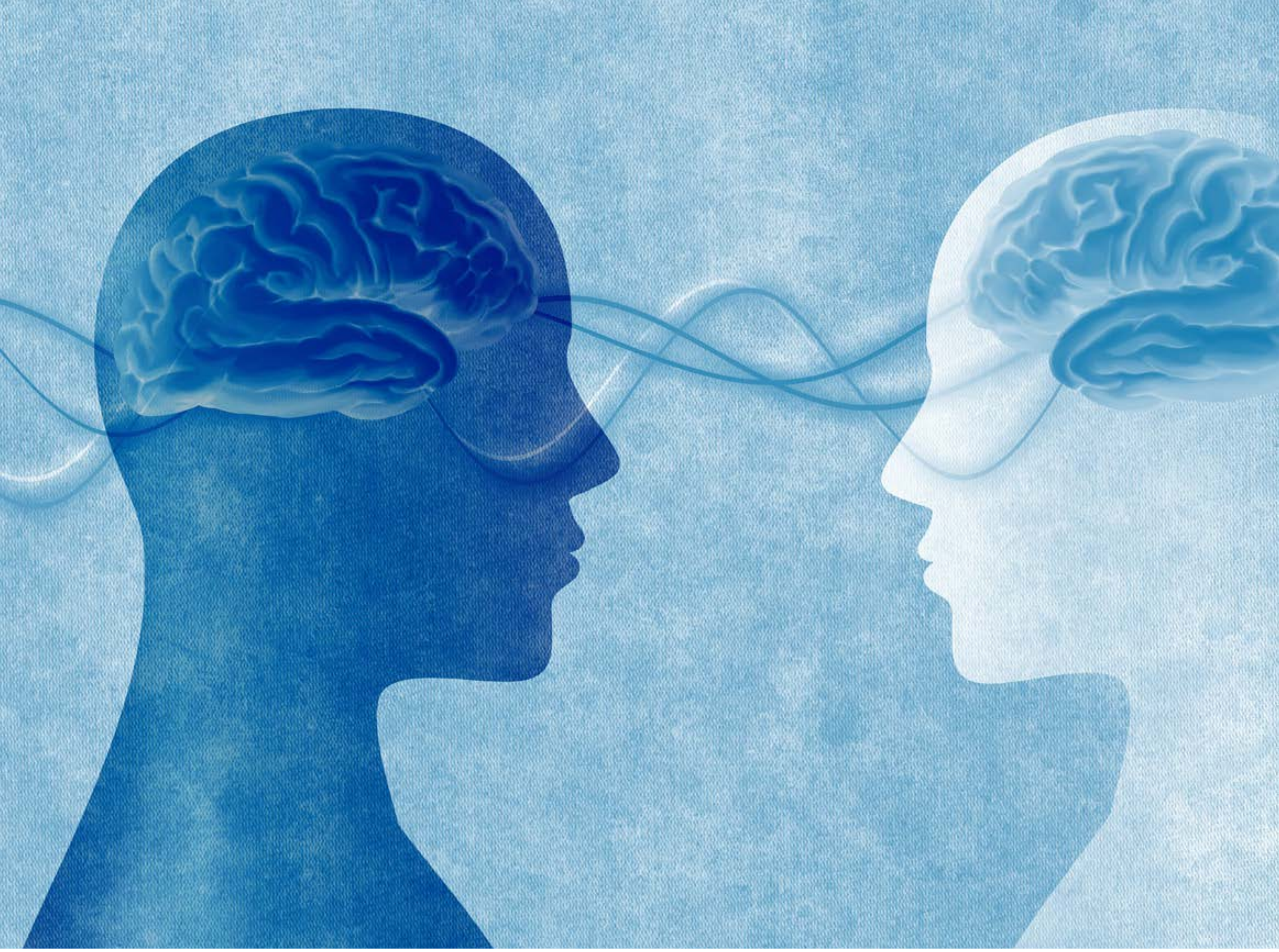
ALAN JERN is a cognitive scientist and an associate professor of psychology at Rose-Hulman Institute of Technology, where he studies social cognition. Follow him on Twitter @alanjern.

REFERENCES

1. Nightingale, S.J. & Farid, H. AI-synthesized faces are indistinguishable from real faces and more trustworthy. *Proceedings of the National Academy of Sciences* **119**, e2120481119 (2022).

2. Groh, M., Epstein, Z., Firestone, C., & Picard, R. Deepfake detection by human crowds, machines, and machine-informed crowds. *Proceedings of the National Academy of Sciences* **119**, e2110013119 (2021).

3. Pennycook, G., *et al.* Shifting attention to accuracy can reduce misinformation online. *Nature* **592**, 590-595 (2021).

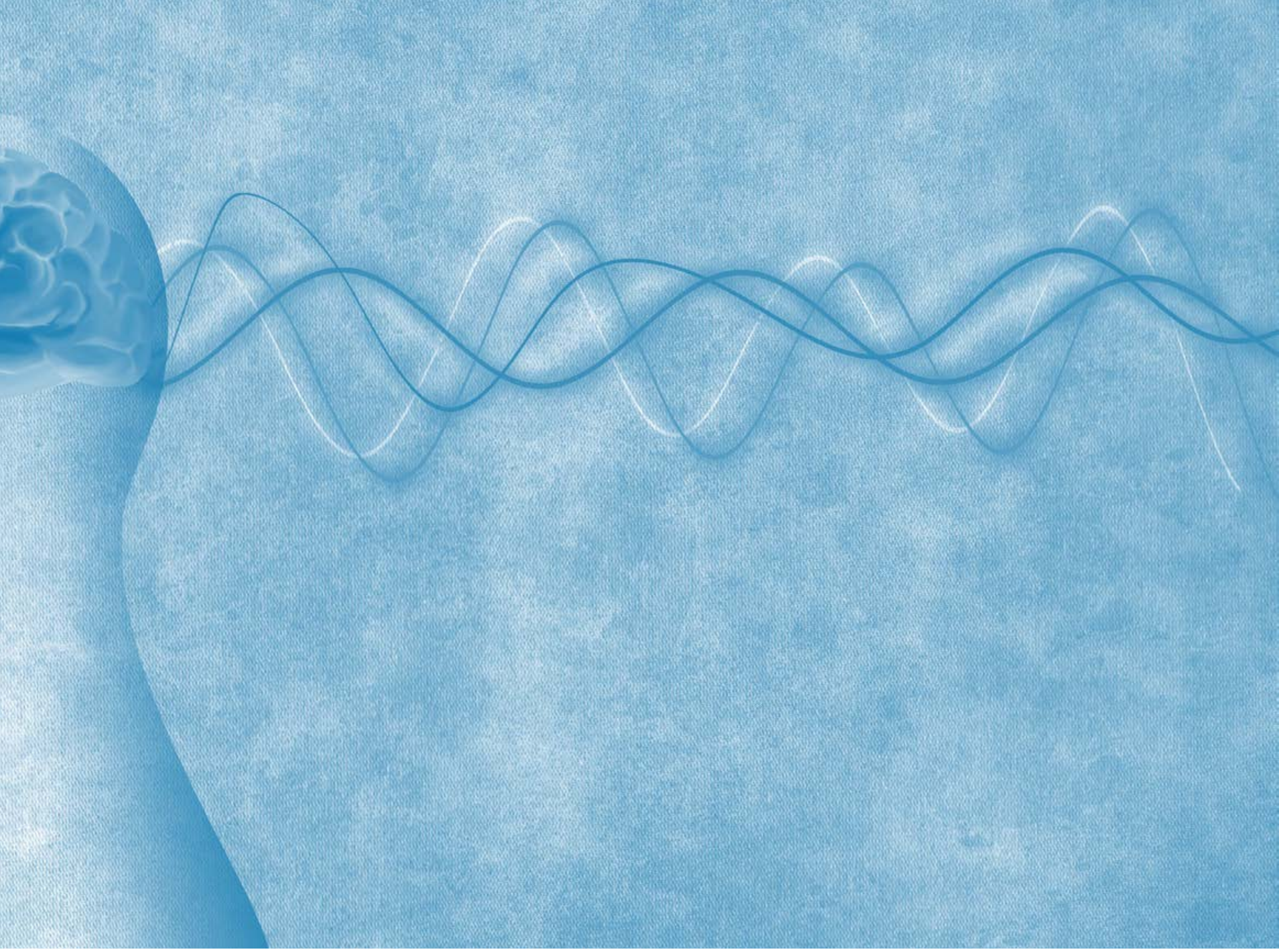


MELITAS / SHUTTERSTOCK

What's So Hard About Understanding Consciousness?

We brought Antonio Damasio and Anil Seth together to share their insights into neuroscience's big question

BY KRISTEN FRENCH



BECAUSE YOU ARE READING these words, it's safe to say you possess consciousness—that magic of awareness that you are alive and awake, and that you are *you*, not somebody else. Consciousness is the scent of pine in winter, the sense of the color blue, the memory of a first kiss, the thrill of an exceptional performance—all of the rich and ineffable qualitative experiences that make a life worth living and that slide through time, knitting one moment to the next.

Consciousness is simpler than we make it. There's no hard problem.

But while we may all be intimately familiar with what consciousness feels like, explaining why it exists or how it arises from physical and biological processes is another matter. These questions are as old as Aristotle, and yet millennia on, we still don't have any definitive answers. For much of history, the nature of consciousness was the purview almost exclusively of philosophers and poets. It was not taken seriously as a legitimate subject for scientific inquiry because it was difficult, if not impossible, to do experiments. But over the past three decades, that has changed, as neuroscientists began to make some real headway in understanding the neural bases of consciousness-related phenomena.

Two neuroscientists who've made major contributions to elucidating consciousness, of different generations, are Antonio Damasio and Anil Seth. In books that include *Descartes' Error* and *The Strange Order of Things*, Damasio, 77, has plumbed the neural basis for feelings, emotions, and decision-making. Seth, 49, has left his mark on the field by illuminating the brain mechanisms behind perceptual experience, particularly vision, and the experience of selfhood. Both scientists published books in late 2021 that deepen and expand their explorations of the brain and consciousness. In *Being You: A New Science of Consciousness*, Seth lays out his theory of "controlled hallucination"—that our perceptual experiences of the world are inventions of the brain governed by a system of predictions—and makes the case that consciousness is intimately tied to the interior of the living body. In *Feeling & Knowing: Making Minds Conscious*, Damasio sets out to demystify consciousness, which he argues evolved from primordial feelings such as pain and hunger, interactive processes that pair the physical with the mental and arise from a chemical orchestra deep in the viscera.

We arranged a Zoom conversation between the two scientists. We asked them to share their respective views of consciousness, ask questions of one another, and offer their wagers on whether we can ultimately solve this hardest of problems.

To start, I want to ask each of you about the titles of your respective books. So Antonio, what do you mean by "Feeling and Knowing"?

Antonio Damasio: Well, I do not think that humans can be conscious without feeling. Homeostatic feelings, such as hunger, thirst, pain, or well-being, are naturally and spontaneously conscious, or they would not be of any value to us, and the knowing is really what comes out of feeling. This spontaneous knowledge helps us manage life. If you are feeling pain, for example, this gives you very actionable *knowledge* about what you should do next in order to survive. That's what I wanted to capture in as few words as possible.

And Anil, what do you mean by "Being You"?

Anil Seth: My work has long had a dual focus: perceptual experiences of the world, on the one hand, of which vision is the easiest form to test, and then the self and emotion, the latter to a good extent inspired by the work of Antonio. But over time my interests have shifted more and more toward understanding what it means to be a self. Way back in childhood we all wonder, "Why am I me and not someone else?" "What will happen to me after I die?" "Is my world the same as somebody else's?" These are very basic questions that I think we all ask and then usually lose track of as we grow up—we get "educated out" of asking them. So I wanted to bring us back to this fundamental issue of what it means to be a self in the world.

One point on which you both clearly agree is that consciousness is an embodied phenomenon, that it is because of and not in spite of our biological bodies that we have the ineffable experience of being alive in this world—a relatively recent reconception of the relationship between mind and matter that owes great credit to the work of Antonio. But you approach this idea of embodiment in different ways in your respective books. What is the role of the body in consciousness?

Damasio: As you know, *Feeling and Knowing* arose out of a request from my editor: to create a book that would present my ideas in a short digestible format. My previous book, *Strange Order of Things*, presents many of the ideas that I discuss in the new book, but one of these is new: the *hybridicity* of feeling and consciousness. By this I mean that unlike straightforward types of perception such as vision, which is largely unidirectional, feelings, and thus consciousness, arise through a comingling of signals from the interior of the body with signals from the nervous system. This is a critical idea because most accounts of consciousness rely exclusively on neural processing rather than an interaction between neural and body processes, as I propose.

Seth: I do arrive at a similar endpoint to Antonio, and some others, which is that consciousness is grounded in our nature as living systems, in the fundamental biological imperative to stay alive day after day. But one of the distinctive ideas I explore in *Being You* is the notion that everything in perceptual experience—both self and world—is a variety of brain-based prediction, something I call "controlled hallucination." Basically, we never perceive objective reality "as it is," whether out there in the world or "in here" in the body. Instead, our brains receive signals from the interior of the body and from the environment and then make predictions about what these signals mean based on prior experience. These top-down predictions are then constantly corrected as the brain receives further sensory prediction error signals. This idea has a long history, with a major landmark in the 19th century when Herman von Helmholtz talked about visual perception as a process of the brain making a best guess, an inference about what's out there in the environment.

Damasio: The way you have just expressed it, Anil, even if you don't have your prediction machine operating at full capacity, consciousness still will emerge. Provided you have a way of accessing information from the interior of your body, you would be conscious, prediction or no prediction. So I wonder what the prediction component adds to the actual experience of consciousness?

Seth: This is a fascinating question. My hypothesis is that the contents of consciousness—whether these relate to the world or the self—are carried by top-down predictions; so predictions don't just add to the experience of consciousness, they bring these experiences about. To test this hypothesis, we need to ask: What in the brain constitutes a minimally predictive system? For me, a minimally predictive system involves the brain encoding a generative model of its world and body—a model that is capable of generating predictions about sensory signals that can be calibrated against sensory data. It's then an empirical matter whether conscious experiences require such generative models, or whether other simpler forms of brain-body interaction would suffice. I'm interested in how these concepts—of brain-based predictions and generative models—fits with your vocabulary, Antonio, for instance "images," which you use to describe sensory patterns of information that are the "most abundant constituents of mind." Maybe there is some kind of resonance here?

Damasio: I have no doubt that predictions play an important role in our mental operations, but I do not see them as necessary for consciousness to emerge.

Antonio, you said in response to my first question, that feelings lie at the core of consciousness, that they are, in fact, spontaneously conscious. Anil, do you agree?

Seth: I do also believe that feelings and emotions are central to the processes of life regulation and conscious experience, but I want to make sure I understand what Antonio means when he says that feelings are "spontaneously conscious," because I know there is ample discussion in our field about whether some emotions are—or could be—unconscious. I think that, conceptually, there is room for predictive control mechanisms that regulate the body and guide our behavior beneath the register of consciousness.

Damasio: To avoid confusion, I think we need to make a clear distinction between feeling and emotion. I have a relatively strict definition of "feelings," which I see as primarily related to the state of life regulation within a given organism. Hunger, thirst, pain, and well-being, which I call "homeostatic feelings," are prominent examples. Feelings are subjective experiences. I think emotions are something quite different. Emotions are collections of acts executed in our faces and bodies and observable by others. Emotions are public. Joy, fear, and anger are prominent examples. So, is it possible to have unconscious emotions? I don't think so.

Seth: Yes, in my book, I argue that there is a spectrum of emotions that I think could be seen as existing on a gradient between feeling and knowing. For instance, based on your definition, sadness would likely qualify as an emotion, but I believe it could also lie in the realm of feelings. It's something that's pretty internal, it could just be there—unexpressed. But disappointment would lie more solidly in the emotional realm, as it requires the brain to have expected something to have been better than it turned out to be, which is a more sophisticated turn. And then, we humans also have regrets. We can even have anticipatory regret, and this adds even further levels of cognitive sophistication, further levels of knowing.

Would you say that these more complicated states of emotion represent higher or more complex "levels" of consciousness?

Seth: I'm personally very skeptical of this kind of labeling, of higher and lower levels of consciousness. There is a little truth to it, of course. Unless we can agree that human beings are in some sense, more conscious than, say, an ant, we're missing some very important things about the nature of consciousness. But to reify that observation, to treat consciousness as analogous to temperature and say, "There is a single scale of consciousness that occupies a range from absolute zero to whatever," I think is taking things too far. This really matters when we think about the possibility of consciousness in infants and in other animals, as well as in adult humans. Infants and non-human animals may be less conscious in some dimensions—for instance, the ability to engage in mental time travel and perhaps to experience things like regret—but they may be more conscious in other ways. Perhaps the immediate presence of their perceptual environment might be more vivid.

Colloquially, people also think about things like psychedelic states or meditative states as being higher levels of consciousness. But that's a societal interpretation that's imbued with moral or ethical value. I think consciousness is a multidimensional concept.

Damasio: I entirely agree with Anil that it's very problematic when people talk about higher levels of consciousness or the increase in consciousness in the world. People are using very, very different concepts, and it just adds to the confusion.

The hard problem of consciousness, set out by David Chalmers in 1995—why and how does experience arise from a physical basis—has roots going back to ancient Greece and continues to confound neuroscientists and philosophers. Anil, you say in your book that the hard problem is the wrong approach to understanding consciousness and believe we should focus our energy on what you are calling the "real" problem of consciousness: explaining why a particular pattern of brain activity—or other physical process—maps to a particular kind of conscious experience.

Antonio, what do you think about the framing of the "hard" problem?

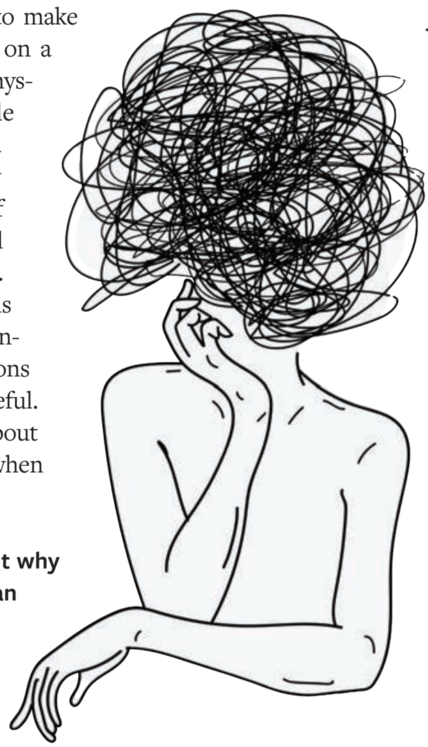
Damasio: I think Anil's description of the hard problem is very kind. I am somewhat less generous. I think it's a hard *obstruction*. And I think the framing of the hard problem, which makes it seem so impossible to explain consciousness through biology, or even to agree on what would be necessary to explain it, is probably responsible for the enchantment that people are currently having with panpsychism. Would you agree, Anil?

Seth: I think the hard problem does a good job of formalizing common intuitions about consciousness, but yes, I also think its framing constitutes an obstacle to progress. Taking the hard problem at face value does lead people naturally, and understandably, to embrace some relatively extreme metaphysical positions. Panpsychism is one of these: the idea that consciousness is fundamental and ubiquitous. The main problem with panpsychism for me is that, almost by definition, it cannot generate any testable predictions, and doesn't actually explain anything at all. So it's a very unproductive route to follow if you're trying to make incremental, scientific progress on a problem. Another extreme metaphysical position which some people take is illusionism. This is the idea that consciousness doesn't really exist, at least not how we think of it, that there's no "problem" at all when it comes to consciousness. Although I find extreme versions of illusionism as unpalatable as panpsychism, there are milder versions of illusionism that are more useful. For example, we can be naive about what it is we're trying to explain when we explain consciousness.

Antonio, can you say more about why you think the hard problem is an "obstruction?"

Damasio: The hard problem makes it look like consciousness is impossible to solve. It doesn't give you any out. Every bit of evidence we have is that the mysteries of the universe have been gradually solved by science. I don't see why consciousness should be any different. I think what engenders this possibility of us being conscious right now, of my knowing that I'm talking to you and seeing your faces on the screen and vice versa, is fundamentally linked to affect. It is a *feeling*, through and through, a feeling of the continuity of life in my organism. We also have reasons to believe that this core feeling is largely being produced below the level of the cerebral cortex. It is then made available to our cerebral cortex by certain subcortical structures, such as those in the thalamus and in the hippocampus. The essence of the consciousness process is actually simpler than what people want to make it. There's no hard problem, really.

Seth: I would go some of the way with you on that. I think the challenge of understanding consciousness is simpler than it's sometimes made out to be. Just compare studying consciousness with something like the James Webb space telescope, which is just heading out to the Lagrange Point now [Ed. Note: It has now successfully arrived at LaGrange Point 2 orbit], and which aims to provide observations of some of the earliest events in the universe, including the formation of the first galaxies. It's an insane feat of engineering, because it's really hard to get the data on the early universe, let alone do a controlled experiment. In the case of consciousness, the brain is comparatively accessible. It's right here. It's not that small. It's not that



I'VE GOT A FEELING Both Damasio and Seth agree that feelings and emotions are central to conscious experience.

KAZNAVOUR / SHUTTERSTOCK

Panpsychism doesn't actually explain anything at all.

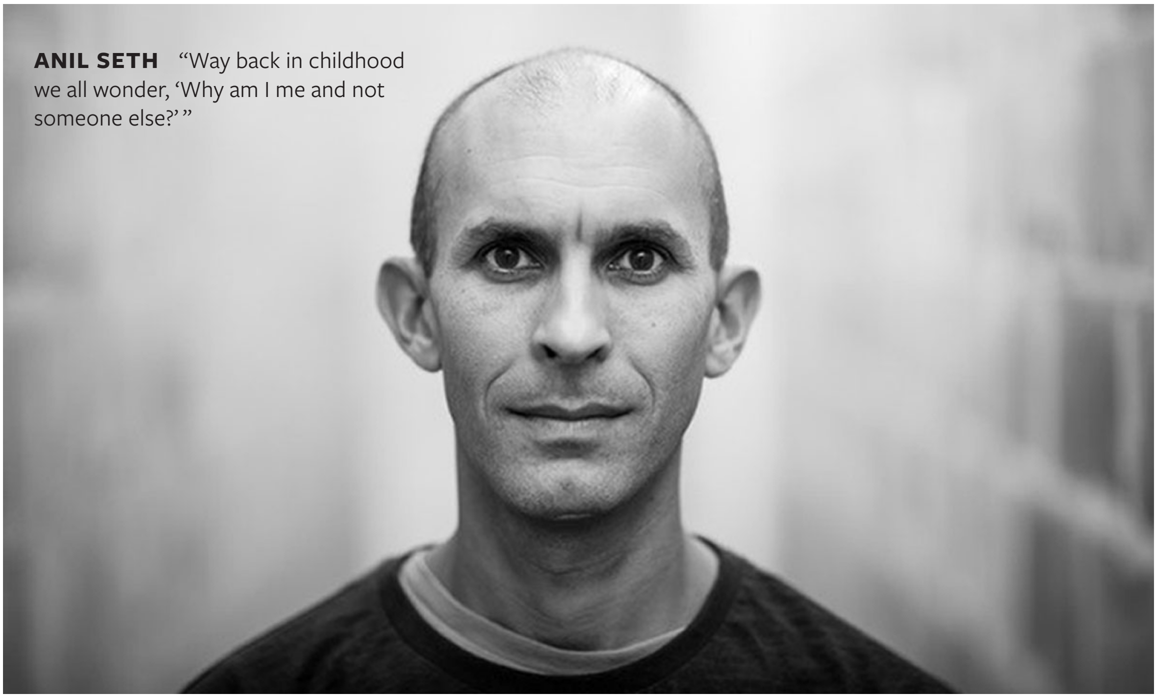
big. We can give people anesthetics and consciousness goes away and then it comes back. We can give people psychedelics, and consciousness changes, and then it changes back. In some ways, consciousness is a remarkably accessible thing to study.

But I also don't think it's that simple, for two main reasons. First, the relevant data for consciousness are private and subjective. But this doesn't mean a science of consciousness is impossible, it just means the data are harder to collect. Second, the hard problem has undeniable intuitive appeal: It is true that consciousness doesn't *seem* to be the kind of thing that can be explained in terms of physical processes. But the fact that something seems mysterious now doesn't mean it will always seem mysterious, and my approach is to explain the properties of consciousness one-by-one, rather than treating it as one big scary mystery. This means *dissolving* rather than *solving* the hard problem. And here, a focus on affect and feeling is very helpful.

I strongly agree with Antonio that there's a core affective basis to all of human consciousness: We perceive the world around us and ourselves within it through and because of our living bodies. But how far can we take that statement? Does it imply that only living things can be conscious, a philosophical position called biological naturalism? Some people would even go further and say, all living things are conscious, which is biopsychism, a sort of biological equivalent of panpsychism, for which I don't see a justification.

I think there's also a temptation to expect more from a science of consciousness than scientific explanations usually provide. And that's because we are trying to explain *us*. In quantum mechanics or in physics, people are more comfortable with the idea that successful theories can be counterintuitive. But we somehow want an explanation of consciousness to be completely intuitive.

Damasio: Well, we know what consciousness provides: It allows me to refer perceptions, ideas, and calculations to a specific organism, my own. It "locates" my mind in my body. How is this feat achieved? The gist of it is there, in the continuity of homeostatic feelings, positive, negative, and neutral, and in what they



ANIL SETH "Way back in childhood we all wonder, 'Why am I me and not someone else?'"

LOVIS OSTERNICK

naturally offer: a continuous and spontaneous identification of the living organism, as the "place" where mind happens, in other words, a conscious experience. Of course, I am quite aware of the fact that many details need to be worked out, but I do not see that as impossible at all. For example, in order to generate feelings, we need peripheral nerves, spinal and brain stem ganglia (such as the trigeminal ganglion), and a variety of nuclei in the brain stem core (such as the parabrachial nucleus). I would bet that the fundamentals of feeling are actually obtained from processes generated by those components of the nervous system.

Well, this leads perfectly to a question that I wanted to ask you both, which is whether you would be willing to bet that consciousness *can* ultimately be solved. Christof Koch bet David Chalmers a fine case of wine that scientists would be able to identify the "neural correlates" of consciousness by 2023. Francis Crick predicted we would be able to explain consciousness by the end of the 21st century. Antonio, you just offered to bet, I think, that we will find that certain areas of the brain produce the fundamentals of consciousness. So, I wondered, if you were betting men, what kinds of bets would you make about whether and when we might solve consciousness?

Seth: People have lost many bottles of wine making these kinds of bets!

Yes, I'm sure!

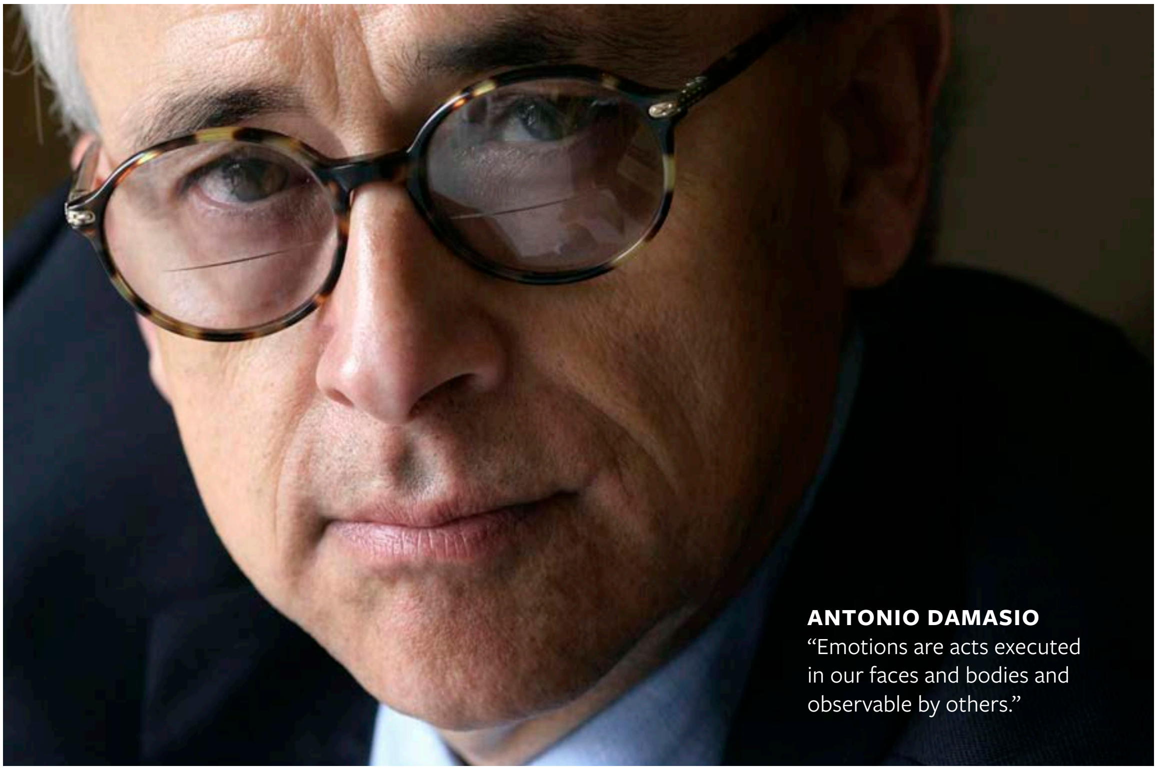
Seth: I'm going to give an annoyingly equivocal answer: It depends. Sometimes, it even depends on the day. On some days I wake up in the morning and consciousness seems as mysterious as it ever did. Other times, I hear people in the academic community say, "We know nothing about how the brain generates consciousness," and this really winds me up. We have made huge advances, especially in the last 20 or 30 years, in our knowledge of how consciousness relates to neuroanatomy and physiology, as Antonio was just explaining. I'm very optimistic that we will continue to discover and understand more and that the questions we'll ask about consciousness will continue to evolve, too. I don't think we'll have a Eureka moment, where somebody says, here's the answer and does the single experiment and we all go home.

What experiments would you design if the possibilities and funding were unlimited?

Is it possible to have unconscious emotions? I don't think so.

Seth: Well, science always marches somewhat in lockstep with the available technology. One of the things we still lack in neuroscience, and not just in the study of consciousness, is the ability to measure activity in the brain at high spatial resolution, high time resolution, and everywhere all at once. We can typically do one or sometimes one and a half of those three things. But being a little less facetious, some of the most interesting experiments happening in the near future are those designed to pit theories of consciousness against each other. We already have a number of candidate theories. There is one called global workspace theory, which says that consciousness happens when information is broadcast in cortical networks in a particular way. There's higher order thought theory, which says that consciousness is all about brain regions representing the activity in other brain regions in a particular way. There are affect-based theories like Antonio's. There are prediction-based theories, among which my version is closely related to Antonio's. The challenge is to find the experiments that arbitrate between these different theories. When the field of physics was figuring out whether relativity could stand up or not, they designed an experiment that could actually arbitrate between Newtonian mechanics and Einsteinian relativity.

Damasio: I agree with you, Anil, on good days and bad days. There are good days when you have some clarity about a possible mechanism that could help explain the nature of consciousness and then bad days, when you realize that doing the right experiment to test that mechanism is going to be either impossible or very complicated. But another good day comes when we actually surmount that problem and make an illuminating discovery. It's quite amazing what we have learned, for example, about the features of the peripheral nervous system, which channel signals from everywhere



ANTONIO DAMASIO
"Emotions are acts executed in our faces and bodies and observable by others."

LUIZ CARVALHO

in the body to the brain and is responsible for the homeostatic feelings I spend so much time studying and contemplating.

For example, the spinal ganglia, which transmit information from all over our bodies to the central nervous system, happen to have a very unique feature: They are directly exposed to the molecules circulating in the bloodstream. Isn't that interesting? Our cerebral cortices are extremely protected from the body by the blood-brain barrier: So, many molecules are barred from entering the cortex. And yet, here in the spinal ganglia, lo and behold, all of these neurons are directly exposed to the circulation of the blood. This is quite spectacular. It is also spectacular to realize that the majority of the neurons in the peripheral nervous system are neurons of a very old evolutionary vintage, and that they are completely different from the neurons that are responsible for our vision or hearing. And yet these older neurons completely lack insulation. They have no myelin to protect the conduction of an impulse. They are also open to non-synaptic contacts. I believe this direct comingling of signals from the nervous system and the body produce the hybridicity I mentioned earlier. These kinds of discoveries give me hope that we are closing in on a detailed physiology of feeling and therefore on an understanding of consciousness.

I had a question about the soul. “Oh dear.”

I had a question about the soul.

Damasio: Oh dear.

Yes, I know. Sorry. Anil, in your book, you suggested that the most fundamental level of consciousness, the experience of just being alive, could be called an echo of the soul. And there’s this lovely passage about that. You mention Atman in Hinduism, which contemplates our innermost essence more as breath than as thought. So I wondered, is there room for belief in the soul in your respective conceptions of consciousness?

Seth: I think there’s a case to rehabilitate the concept. One formulation of the soul is that it is an immaterial essence of you that can survive the death of the body and maybe hop into another body and perhaps resubstantiate itself and so on. This kind of soul is often associated with rationality and personal identity. But there are many other conceptions of the soul where it is less linked to reason and thinking and more closely tied to the body and bodily processes like breathing. So yeah, I think there’s a place for it, but it’s a place for a more embodied, more biological soul.

What do you think, Antonio?

Damasio: I agree that what philosophers and theologians have described as a soul is very much captured by the continuous feeling of existence that I mentioned earlier. Feelings such as hunger or thirst are occasional, but if we pause for a moment and observe carefully what goes on in our minds, we realize that “a feeling of just being” is present *continuously*. And when that feeling of being is not there, “you” are not there either, which is what happens when you faint or when you go under anesthesia—consciousness disappears.

And this is a question mostly for you, Antonio. You’re both a neuroscientist and a philosopher and those two

fields have historically been at odds with each other on the subject of consciousness. So I wondered how does philosophy inform your position on consciousness?

Damasio: I have too much respect for the discipline to call myself a philosopher. I was never trained as a philosopher. Of course, I love philosophy. I do think that neuroscientists cannot properly study consciousness if we don't know the history of the treatment of the problem in philosophy. But philosophers have not always considered the neuroscience of consciousness in their own work. And that is a pity because neuroscience has so many unequivocal facts to contribute.

Seth: Maybe 20 or 25 years ago, it was perfectly legitimate to be a philosopher of mind, talking about consciousness without really engaging with the neuroscience. And it was reasonable to be a neuroscientist interested in consciousness without necessarily being literate with the philosophy. But that's changed—and changed for the better. Some philosophers are even doing experiments, and to be a serious neuroscientist of consciousness, you need to be receptive to philosophical perspectives, know the territory and be open to the dialogue. The crossover between neuroscience and philosophy is becoming more baked in, which is very promising for the field. In fact, there's a lot more crosstalk between philosophy, neuroscience, and the medical and clinical sciences as well. Anesthesiologists have only relatively recently got into the game, talking with other neuroscientists and philosophers about what they're doing.

So returning to the question, have we made progress? I'd say this is a reason for optimism.

Damasio: Absolutely. I agree.

KRISTEN FRENCH was raised by Christian Scientists, which inspired a fascination with medicine, biology and the relationship between mind and matter from a young age. She is a science editor and writer with bylines in *Wired*, *New York*, *Proto*, and *The Verge*, among other publications. New York City, where she lived for 16 years, will always feel like home, but she is now based in San Diego, California, land of surfers and sunshine.



Animals Feel What's Right and Wrong, Too

It's time to take moral emotion in animals seriously

BY JAMES HUTTON

CRANACH / SHUTTERSTOCK

AMY'S JOB IS PRETTY REPETITIVE, but normally she doesn't mind doing what she's asked. Today, however, she's working alongside Sidney. Amy can't help noticing that Sidney is receiving a small reward every time he completes a task, while she gets nothing. After a few rounds of this treatment, Amy has had enough. She refuses to go on performing her tasks, disengaging completely.

Amy's behavior probably makes a certain kind of sense to you. Here's what I expect you're thinking: Amy must feel like she's being treated unfairly (and rightly so, you might think). It is this perception about the ethics of the situation—her sense of not being treated fairly—that motivates Amy's refusal to go on working.

But there are a couple of important details about Amy and Sidney that you should know. The first is that they aren't workers in any conventional sense, but participants in an experiment. The experiment in question, which is taking place at the University of Vienna, is part of an effort to understand the motivations that underlie cooperative behavior. The second important detail is that Amy and Sidney aren't human beings, but Border collies. A team of animal psychologists, led by Friederike Range, are closely observing them to determine whether domestic dogs show an aversion to disparities in reward. The behavior of Amy and the other dogs in the experiment suggests that they do.[1] Dogs notice when they aren't getting the same rewards as their peers, and once they've figured this out, they refuse to go on performing tricks on demand.

So, what do you make of Amy's behavior now that you know she's a dog and not a human? Is it still viable to think she's perceiving her treatment as unfair? You might think that attributing a moral attitude like a sense of fairness to a dog is hopelessly romantic, anthropomorphic, and thoroughly unscientific. Moreover, you might think the capacity to experience the world in moral terms is something uniquely human, perhaps bound up with our ability to form abstract principles and engage in complex ethical reasoning. But there are two big ideas—one from moral philosophy and one from animal psychology—which suggest that a sense of fairness, a feeling for what's right and wrong, might really be within reach for some non-human animals.

The first big idea is that the moral attitudes of human beings are thoroughly emotional in nature. Of course, we sometimes draw ethical conclusions by reasoning, carefully thinking through the implications of the principles we believe. This, for example, is how many of us have come to recognize that innocuous-seeming activities like catching a plane have weighty ethical significance. But other times we just have an intuitive sense of what's right and what's wrong. You can often just see that someone's behavior is kind, cruel, uncalled-for, or unfair, without having to derive this conclusion from some deeper moral principle. This intuitive moral sense is what enables us to make moral judgments on the fly, to navigate nuances of rightness and wrongness that are difficult to codify in abstract rules and, sometimes, to see shortcomings in the moral principles with which we have been brought up.

Emotional feelings allow us to expand our moral horizons in the face of lived experience.

Philosophers have long suspected that this intuitive moral sense is essentially a capacity for certain kinds of emotion: We experience positive emotions of satisfaction and admiration toward good conduct and negative emotions of anger, disgust, and guilt toward the bad. The past two decades have witnessed a huge revival of interest in this idea, due to scientific findings about how moral judgment works. When we make fast and intuitive moral judgments, the parts of our brains associated with emotions light up[2]; manipulating people's emotional responses, by exposing them to disgusting smells,[3] for example, seems to lead to corresponding shifts in patterns of moral assessment; individuals with emotional deficits such as psychopathy appear to lack an intuitive sense of rightness and wrongness.[4] All this evidence strongly suggests that emotions are integral to the human sense of right and wrong.

In my own research, I've argued that this conclusion extends to the experiential component of our ability to perceive rightness and wrongness.[5] When a person feels she is being treated unfairly, rather than concluding this on the basis of reasoning, the feeling is an emotional one. These emotional feelings allow us to expand our moral horizons and revise our ethical principles in the face of lived experience. Our inner lives are punctuated by moral experiences, and these moral experiences are constituted by various forms of emotion.

HOW DOES THIS APPLY to non-human animals like the Border collies? Can we justify the idea that creatures like Amy perceive certain kinds of treatment as unfair?

Our first big idea tells us that a *human* in Amy's situation would perceive the unfairness of the situation by experiencing an emotion such as anger or indignation. What if a dog like Amy could experience similar emotions? If so, there would be a clear sense in which she is having a moral experience, just as a human would in her position. It would follow that the ability to perceive the world in moral terms isn't unique to human beings but shared by other members of the animal kingdom.

But are we really justified in ascribing emotions like anger or indignation to other animals? This is where the second big idea comes in.

For a long time, the dominant methodology in animal psychology was to ascribe as little mental life as possible. This began as an attitude of healthy skepticism toward the all-too-human tendency of projecting our own thoughts and feelings onto animals at the slightest provocation. But by the 1950s, this approach had ossified into a blanket refusal to take seriously the idea that non-human animals have inner lives. However, in the decades since then, researchers have built up a huge body of evidence that some non-human animals can perform cognitively demanding tasks, from using tools[6] and recognizing themselves in mirrors to working together to solve complex puzzles.[7] This has forced scientists to become less cagey about exploring the mental lives of animals—to consider claims about animal minds on a case-by-case basis rather than denying them by default.

Part and parcel of this is a new willingness to take seriously the idea that animals have emotions. In an opinion piece published in the journal *Affective Science* in March 2022, cognitive scientist Mariska Kret and her collaborators report that "most contemporary researchers do not deny the existence of emotions in animals."[8] What remains contentious, however, is whether the emotions of non-human animals are accompanied by conscious feelings. Here, too, Kret and her colleagues argue that it's time for scientists to shift to a more permissive approach. "When related species show similar behavior under similar circumstances, these are likely driven by similar psychological processes," they write. "Until the contrary can be demonstrated, we must assume that similar behavior in these species is paired with similar emotions and

THE STING OF UNFAIRNESS
You might think it is hopelessly romantic and unscientific to say your dog can sense unfairness, philosopher James Hutton says. But two big ideas from philosophy and psychology suggest a feeling for what's right and wrong might really be within reach for some non-human animals.

DEZY / SHUTTERSTOCK

Many species show a willingness to give up food to help another creature in distress.

in some cases similar feelings." This means that when another mammal shows signs of emotion—like facial expressions, changes in heart rate, and shifts in cognitive processing—our default hypothesis should be that they are consciously feeling emotions.

The same stance is advocated by primatologist Frans de Waal and philosopher Kristin Andrews in an article that recently appeared in *Science*.[9] "To deny felt emotions [in non-human animals]," they write, "does not seem a reasonable position given the fundamental similarity between the nervous systems of humans and other animal species and the shared evolutionary history that has promoted similar emotionally mediated reactions to the environment and social partners."

De Waal and Andrews note that our evidence that *other human beings* are experiencing conscious feelings is always indirect, too. You don't feel my anger firsthand but detect it via external signs such as how I act and what I say. Some scientists are only willing to ascribe conscious feelings when a human being reports having them, but verbal evidence is no more infallible or indispensable than other external signs. The face can provide a "window on human emotions" just as much as the voice. Indeed, an over-reliance on verbal evidence has led to serious moral failings in the past, such as the practice of performing surgery on babies without anesthetics ("We can't be *sure* the baby is feeling anything just because she is crying and whimpering!"), which only stopped in the 1980s.[10] To do better in the present, we need to take signs that other animals are feeling emotions at face value unless there is substantive evidence to the contrary.

So, there is nothing inherently outlandish or unscientific in the suggestion with which we began, that a dog like Amy can experience certain kinds of treatment as unfair.

BUT WHAT POSITIVE EVIDENCE do we have that other animals can feel specifically moral emotions, like indignation at unfair treatment or concern for another's well-being? For all we've said so far, other mammals might be experiencing non-moral emotions, such as fear oriented exclusively toward their own well-being, without experiencing the world in moral terms. That would still be important; we care about animal suffering, so it's important for us to understand the ways they experience emotional as well as physical pain. But it would fall short of the more radical idea that animals have moral experiences.

If animals really do experience the world in moral terms, what does this mean for how we should treat them?

This line of thought leads Frans de Waal to doubt that other animals can see the world in moral terms. "I am reluctant to call a chimpanzee a 'moral being,'" he wrote in *The Bonobo and the Atheist: In Search of Humanism Among the Primates*. "This is because sentiments do not suffice. There is little evidence that other animals judge the appropriateness of actions that do not directly affect themselves. Moral emotions are disconnected from one's immediate situation. They deal with good and bad at a more abstract, disinterested level."

But perhaps de Waal's negative conclusions about animal moral experience are too pessimistic. Think again of Range's dogs. Admittedly, they only respond to disparities in reward when they are the ones getting the short straw. Amy stops giving the paw when she notices the disparity, while Sidney carries on giving the paw and getting the treats. But is the human sense of fairness really so much more disinterested than this? We are much better at noticing when *we* are on the receiving end of unfairness than when others are. Sure, we can do things that dogs can't do to move beyond this—we can listen to different points of view and use reasoning to bracket our own interests. But at the experiential level, our moral emotions seem hardly less partial than the Border collies' reactions. It's therefore unclear that de Waal's focus on disinterestedness marks a genuine difference between humans' moral emotions and the emotions of which animals are capable.

What's more, we do see signs of selflessness or disinterestedness in the behavior of other non-human species. An experiment by Sarah Brosnan and colleagues found that chimpanzees become agitated by disparities in reward even when they are the ones getting the better deal, refusing to work for tasty grapes when another chimpanzee is only getting mediocre carrots.[11] Moreover, in the wild, high-ranking chimpanzees break up fights they aren't involved in, even when this means punishing their own friends. And many species, from rats[12] to pigeons[13] to rhesus monkeys[14] to chimpanzees, show a willingness to give up food to help another creature in distress. These behaviors are evidence that these animals are experiencing emotions directed at the well-being of others rather than at their own lot, removing any rationale for denying that these are genuinely moral emotions.

Based on evidence like this, cognitive ethologist Marc Bekoff and philosopher Jessica Pierce recommend a cautious but permissive approach to ascribing moral emotions to animals. In *Wild Justice: The Moral Lives of Animals*, they write, "Can we draw a line that separates species in which morality has evolved from those in which it hasn't? Given the rapidly accumulating data on the social behavior of numerous and diverse species, drawing such a line is surely an exercise in futility, and the best we can offer is that if you choose to draw a line, use a pencil. For the line will certainly shift 'downwards' to include species to which we would never have dreamed of attributing such complex behaviors."

We need to be somewhat tentative in ascribing moral experiences to any given species, but there are compelling reasons to take this possibility extremely seriously.

If animals really do experience the world in moral terms, what does this mean for how we should treat them? It would require us to rethink the range of harms we must avoid causing. Suppose that mother pigs feel moral concern for the well-being of their piglets. That would mean that, when a mother witnesses her piglets undergoing painful operations like tail-docking and castration, her interests are harmed, too. Those of us who work with pigs—and those who choose which farming practices to support with our wallets—would need to consider the moral bonds pigs feel toward one another as well as the physical and emotional pain they can feel.

Even more radically, animals who are motivated by a sense of right and wrong are, in a very real sense, moral agents. Respecting moral agency is one of the cornerstones of modern ethics, so this might require us to consider more extreme forms of animal liberation. Philosopher Susana Monsó and colleagues argue that training a dog to fight constitutes an unacceptable stunting of its moral agency, because this involves "eliminating any potential caring response to a conspecific in distress."[15] Similarly, lab experiments in which scientists intentionally destroy rats' capacity for empathy might be analogous to the long-discredited practice of lobotomizing children with behavioral problems.

A broader lesson, increasingly acknowledged by experts in this field, is that there is much to be gained when scientists and philosophers work together. Scientists' experimental findings have driven philosophers to make conceptual and ethical innovations. In turn, these innovations have led to improvements in methodology and a better grasp of what the scientific findings mean. With continued cross-pollination between science and philosophy, we will continue to deepen our understanding of our own moral minds and the moral lives of animals.

JAMES HUTTON is a Leverhulme Trust postdoctoral researcher in philosophy at the University of Edinburgh. Follow him on Twitter @JamesHuttonPhil.

REFERENCES

1. Range, F., Horn, L., Viranyi, Z., & Huber, L. The absence of reward induces inequity aversion in dogs. *Proceedings of the National Academy of Sciences* **106**, 340-345 (2009).

2. Greene, J.D., *et al.* An fMRI investigation of emotional engagement in moral judgment. *Science* **293**, 2105-2108 (2001).

3. Schnall, S. Haidt, J., Clore, G.L., & Jordan, A.H. Disgust as embodied moral judgment. *Personality and Social Psychology Bulletin* **34**, 1096-1109 (2008).

4. Aaltola, E. Affective empathy as core moral agency: psychopathy, autism and reason revisited. *Philosophical Explorations* **17**, 76-92 (2014).

5. Hutton, J. Moral experience: Perception or emotion? *Ethics* **132**, 570–597 (2022).

6. Finn, J.K., Tregenza, T., & Norman, M.D. Defensive tool use in a coconut-carrying octopus. *Current Biology* **19**, R1069-R1070 (2009).

7. Plotnik, J.M., Lair, R., Suphachoksahakun, W., & de Waal, F.B.M. Elephants know when they need a helping trunk in a cooperative task. *Proceedings of the National Academy of Sciences* **108**, 5116-5121 (2011).

8. Kret, M.E., Massen, J.J.M., & de Waal, F.B.M. My fear is not, and never will be, your fear: On emotions and feelings in animals. *Affective Science* **3**, 182–189 (2022).

9. de Waal, F.B.M. & Andrews. K. The question of animal emotions. *Science* **375**, 1351-1352 (2022).

10. Rodkey, E.N. & Riddell, R.P. The infancy of infant pain research: The experimental origins of infant pain denial. *Journal of Pain* **14**, 338-350 (2013).

11. Brosnan, S.F., Talbot, C., Ahlgren, M., Lambeth, S.P., & Schapiro, S.J. Mechanisms underlying responses to inequitable outcomes in chimpanzees, Pan troglodytes. *Animal Behavior* **79**, 1229-1237 (2010).

12. Church, R.M. Emotional reactions of rats to the pain of others. *Journal of Comparative and Physiological Psychology* **52**, 132-134 (1959).

13. Watanabe, S. & Ono, K. An experimental analysis of "empathic" response: Effects of pain reactions of pigeon upon other pigeon's operant behavior. *Behavioral Processes* **13**, 269-277 (1986).

14. Masserman, J.H., Wechkin, S., & Terris, W. "Altruistic" behavior in rhesus monkeys. *The American Journal of Psychiatry* **121**, 584-585 (1964).

15. Monsó, S., Benz-Schwarzburg, J., & Bremhorst, A. Animal Morality: What it means and why it matters. *The Journal of Ethics* **22**, 283-310 (2018).

How Putin's War Is Sinking Climate Science

An American journalist leaves Russia as war breaks up the international collaboration key to climate research in the Arctic

BY ANDREA PITZER

IN THE END, the war came three days early. It found me in Moscow, where I watched a Russian news anchor on state television call tanks crossing into Ukraine a "special operation."

A Russian friend watched with me. We sat without speaking, dull and blank as the snow outside. Soon after, another Russian friend came over, and we discussed whether the ticket I'd bought for the next day would get me out of the country soon enough, or whether I had to go to the airport immediately.

I'd believed war was coming, but my belief hadn't been shared by most of my Russian acquaintances. Nor had it been shared by the dozens of people I met in February as I worked my way through a series of cities from Ufa in the south to Arkhangelsk in the north.

I am a journalist who has reported on climate change and gone on four research expeditions in the Far North in recent years. My latest published book, *Icebound*, narrates Dutch navigator William Barents' voyages north of Siberia in the 1590s. I've seen for myself the value of international collaboration in science in the Arctic and have studied Russian, cold-water diving, and navigation specifically to do cross-border work on climate and archaeology. But the war, of course, has changed everything.

I had flown to Moscow early in February to research Russian Arctic history. By the time the war began, I'd spent more than two weeks in and out of museums, photographing rocks and relics, and learning from experts in everything from anthropology and ornithology to naval history and geology—all material for use in my next Arctic book and for reporting on the state of the planet.

Despite the Kremlin's daily condemnations of Washington, everyone had been friendly. Only on the streets of St. Petersburg had anyone confronted me—and even there, my lone inquisitor seemed to be a member of the international fraternity of harmless drunks. The rhetoric of presidents is far removed from the lived lives of their people.

Drifting through the realms of science and history, I'd traveled in a bubble, until the war altered my immediate plans—and future ones, too. But it would have more profound effects on the researchers I'd met, the citizens of Russia as a whole, countless scientific projects around the world, and, most of all, the people of Ukraine, where Russian forces had already begun their slaughter as I boarded a plane home.

OF MY FOUR EXPEDITIONS to the high Arctic, two have been to Norwegian outposts and two to Russian territory. The Arctic is a natural place to consider this war's effect on scientific collaboration, if only because the region is home to so many partnerships, and the science done there is so essential for the future of the planet.

The farther north one goes, the smaller the distance between countries that border the Arctic, and the greater the pressure and possibilities for cooperation. But in the face of war, international partnerships have frozen or foundered.

By releasing carbon and methane, thawing permafrost runs the risk of turning the Arctic from a haven protecting the globe to a region driving climate change. But due to the war, a global consortium researching permafrost has halted work with Russia. *Science* magazine reported other immediate disruptions in joint Arctic efforts, which include the U.S. Fish and Wildlife Service officials telling its staff not to contact Russian counterparts, leaving existing polar bear and salmon projects on hold. A collaboration I witnessed on an expedition in 2019, when a Norwegian scientist joined a Russian crew to map trash on the shores of the Kara Sea, would be impossible today.

Unlike the Antarctic, the Arctic has no polar continent—it is a scattering of masses and archipelagos. Yet, here, as in lower latitudes, land and water are divided among nations—in this case, those whose shores and territories touch Arctic waters. Composed of eight permanent members (Norway, Canada, Denmark, Sweden, Finland, Iceland, the United States, and Russia), as well as six participants representing indigenous peoples, the Arctic Council was established in 1996 and meets twice a year to establish shared priorities in the region.

A 1987 speech by Soviet leader Mikhail Gorbachev calling for new Arctic initiatives and protections for the region is often credited with inspiring the creation of the Council. But a week into this war, all the members except Russia invoked sovereignty, territorial integrity, and international law to condemn the war in Ukraine and "the grave impediments to international cooperation, including in the Arctic, that Russia's actions have caused." Pausing all meetings as long as Russia is chairing the Council (a role that ends next year), the other members underlined the importance of cooperation while arguing that the attack on Ukraine was egregious enough to suspend it.

Even if the international borders that cleave the ocean are imaginary, their existence has real consequences. And these consequences will only grow more consequential, with the Arctic warming four times more quickly than the planet as a whole. The massive decline in Arctic ice in the last two decades harms animals and the environment, while opening whole new areas to year-round navigation.

These changes have sent countries jostling for dominance in places that already have unsettled history. A

PAGES 76-77: JO CREBBIN / SHUTTERSTOCK

I thought wrong, and the war arrived four days after closing ceremonies instead.

mix of military interests, fossil fuel drilling, expanding tourism, and environmental concerns keep Arctic planning fraught.

Still, irrespective of their country of origin, teams working in the high Arctic often develop their own traditions, superstitions, and negotiated culture. Once per voyage, ice for drinks should come from freshly captured chunks of iceberg; at key points, Neptune requires small but high-end liquor tributes delivered directly to the frigid sea; and a halfway party should be thrown at the farthest northern point on an expedition.

Likewise in the north, nationality sometimes goes by the wayside in pursuit of scientific research. Collaboration can have a multiplier effect that minimizes the impact on the Arctic environment, as well as freeing up time and money for additional projects.

YEARS AGO, on my first trip to the Arctic, Kim Holmén of the Norwegian Polar Institute picked me up at the small airport in Longyearbyen on Svalbard, halfway between Norway and the North Pole. It was the middle of polar night, so we sat inside his office, and he explained the nature of collaboration in the Arctic.

"Together, we can do bigger science," he said, about the kind of research he's done for three decades. "If we meet, we can trust each other. And if we trust each other, we can move forward."

For much of the science that's done in the Arctic, this international mode is the default. Teams often communicate in a language that's not their native tongue, or go to school outside their homeland, or do fieldwork with scientists from many countries. During my expeditions, I've partnered with sailors from a half-dozen countries; biologists from Norway, Russia, and the U.S.; and guides from Belgium, Russia, Norway, the Czech Republic, and Australia. On shore, radio chatter between groups borrows the most useful or specific scientific terms from various vernaculars, and evenings, a referee might run a single shipboard game of Mafia or supervise charades in multiple languages simultaneously.

And never has it been more critical to collaborate. Just days after Russia attacked Ukraine, the United Nations' Intergovernmental Panel on Climate Change—involving the work of 270 researchers from 67 countries—released its first report for 2022. The authors underlined the urgency of acting now to stabilize ecosystems and save existing species, in order to preserve the planet's ability to adapt. Hans-Otto Pörtner, a working group co-chair, noted, "Any further delay in concerted global action will miss a brief and rapidly closing window to secure a livable future."

Meanwhile, scientists outside Russia are in many cases afraid to contact their peers there, for fear of putting them in danger. "Clearly, we do not want to get our scientist friends in trouble," says Holmén, still with the Institute in 2022, now as a special advisor. "And in a totalitarian system you can quickly get in trouble."

With the situation changing from day to day, and seeming to go from bad to worse, some scientists I contacted for this article did not want to speak about their Russian counterparts at all, for fear of harm to them, or of jeopardizing future projects that remain essential from a planetary perspective. For the same reasons, I am not sharing the names or details of the researchers I met with in February.

In the Arctic, everyone understands how contingent survival is. Indigenous peoples have helped the nations that decimated them, sharing knowledge and advocating for more sustainable practices. And not surprisingly, the three agreements made to date by the Arctic Council cover just such collaborative concerns: pollution, search and rescue commitments, and scientific cooperation.

ICEBREAKER The Russian icebreaker 50 Years of Victory passes an iceberg on its way to the North Pole in 2021.

Scientists outside Russia are in many cases afraid to contact their peers there, for fear of putting them in danger.

The Council has weathered prior military crises. As University of British Columbia professor Michael Byers wrote in 2017, council cooperation continued despite Russian condemnation of the U.S. invasion of Iraq in 2003, and also after the Russian war in Georgia in 2008 and its seizure of Crimea in 2014.

According to Byers, it may be that in the Arctic, we see "a combination of remoteness, great technological challenges, and huge expenses pushing in the direction of cooperation and burden-sharing." Nevertheless, he closed his case study on Crimea five years ago by noting that Arctic collaboration could still be tested by events in Ukraine:

> "This is not a celebratory article: The Ukraine crisis and related tensions between Russia and the West are matters of serious concern, and continued Arctic cooperation should not be taken for granted."

A lot has happened in the interim, but just three years ago, MOSAiC, the largest Arctic expedition in history, brought together hundreds of scientists from 20 countries around the world, and sent them to live aboard the Polarstern—a 387-foot, 20,000-horsepower German-built icebreaker—locked in ice and drifting for months. Without the assistance of Russian scientists, who have the most experience living and working on ice floes, as well as aid with emergency refueling from Russian icebreakers, that mission couldn't have taken place. Folke Mehrtens, a press officer with Germany's Alfred Wegener Institute, said on March 18 that it is not yet clear whether or how Russian partners would be permitted to participate in the April conference and workshop to present resulting data.

ANDREA PITZER

Shutting Russian scientists out of research long-term would be self-defeating. Over half the Arctic Ocean's coastline—some 15,000 miles of it—sits in Russian territory. It's a challenge to work in any comprehensive way on a number of problems—including pollutants, currents, climate, or environmental projects—without Russian data.

But even if most funders are in agreement over pausing collaboration during this war, how is it possible to weigh human suffering between two man-made crises: a long-term one that will affect billions of people and render parts of the planet unlivable, and an immediate one that has killed thousands and displaced more than 3 million human beings in less than three weeks?

"It's painful to say, but compared to war and people killing each other, the environment is a lesser issue," said Holmén. "But the aftermath of a conflict of this type will be severe for the environment."

Climate science is not linked to the conflict in Ukraine simply through Arctic partnerships suspended or damaged by war. Russia's post-Soviet power has roots in ongoing European and American consumption of fossil fuels. Understanding climate change and scientific approaches that might mitigate it could, in theory, help reduce global instability that facilitates war.

The scientific questions extend beyond the Arctic in other ways, too. The cataclysm in cross-border partnerships is occurring worldwide. What happens to global science and international collaboration during a war?

The effects have been most profound for Ukrainians. Nuclear power-plant workers at Chernobyl and Zaporizhzhia have effectively become Russian hostages forced to carry out their duties at gunpoint, with little or no relief. At the other end of the planet, safer but

What happens to global science and international collaboration during a war?

immobilized by war, are the dozen Ukrainian researchers now isolated but still working at Ukraine's lab in the Antarctic. Science as a whole and scientists' lives are ruptured in ways that eclipse the issue of postponed collaboration with Russian scientists. (Among more positive developments, academics around the world have been attempting to help Ukrainian researchers in harm's way by creating international partnerships, and these efforts have expanded exponentially.)

Even if the war comes to a sudden end, and Ukraine maintains its independence, researchers say that to build trust again will be a slow process. But for now, while the conflict is still ongoing, the British government's funding agency, U.K. Research and Innovation, has suspended grant payments for all projects involving Russian scientists. (Calls to universally block Russian researchers from publishing in European and American journals have apparently gained little traction so far.)

Even the International Space Station has not been immune from wartime repercussions, with Dmitry Rogozin, the head of Russia's space agency, openly hinting at dire consequences for the U.S. and Europe if Western governments continue to block cooperation with Russia. But so far, Rogozin has only lobbed rhetoric, with no signs of any concrete plan to damage the station or strand NASA astronaut Mark Vande Hei in space.

In the middle of a war, it feels almost obscene to even ponder the long-term effects on research, as if now is not the time to ask such questions. And even if they are asked, we do not yet have answers.

BEFORE I FLEW TO RUSSIA from Washington, D.C., I had weighed the timing of the trip and the wisdom of going at all. I knew that historically, Vladimir Putin had launched prior Russian invasions in close proximity to the Olympics, sending forces into Georgia in 2008 and Ukraine in 2014. But after his cordial meeting with China's Xi Jinping in the first days of February, it seemed to me that Putin would not seriously contemplate a strike until the Beijing Olympics had ended. It would surely take a week after that for Russia to launch any significant invasion—and it would probably begin, I thought, with a stealth incursion into Ukraine from Belarus. But I thought wrong, and the war arrived four days after closing ceremonies instead.

By the time I left for Moscow, I'd just recovered from COVID-19, and my visa was still valid. Maybe I could finish my research in Russia before borders closed again, as they had during the pandemic. More importantly, along with the archival work I had planned to do, I wanted to see the researchers and academics who had been with me as I worked on my three prior books, the people who had gone to sea with me, who had translated for me, and who had made possible important parts of whatever success I've had as a writer. I wanted to see people I love, people I might not see again for a year or two or five, or who knew how long it might be, if war came. And then it came.

Some researchers and academics in Russia are now lying low. Others are writing open letters against the invasion or going to protests. Some are signing letters trying to protect their students who have taken to the streets. Others are publicly supporting Putin. Still others hope conditions will ease, so that they can return to pressing work that might preserve the planet. Some fall into multiple categories.

But the future is unclear. Nearly three weeks after the beginning of the invasion, amid talk of a ceasefire, Putin called for a "natural and necessary cleansing" of Russia from "foreign sympathizers" and "fifth columnists," underlining the question that had already been lingering in the background: Will even an end to the war be sufficient to begin to rebuild what has been destroyed? And what if it is not?

In February, just days before the war, I stood with a scientist and a sailor on the shore at Arkhangelsk and looked out to the White Sea, and past that sea, to the Arctic. It was late at night, the banks were treacherous with ice, and, braced against a bitter wind, we could barely make out anything in the darkness. But the sea and everything beyond it—we knew they were still there.

It remains to be seen how deep the current rupture goes and how long it will last. For now, scientific partnerships are shattered. The window to prevent climate catastrophe is closing. And in Ukraine, people are dying.

ANDREA PITZER is a journalist and the author of *Icebound: Shipwrecked at the Edge of the World*, which narrates the three Arctic voyages of Dutch navigator William Barents. She is also the author of *One Long Night: A Global History of Concentration Camps* and *The Secret History of Vladimir Nabokov*.

Painkillers That Don't Kill

Scientists are getting closer to designing safer, more targeted pain drugs to replace opioids

BY CATHRYN JAKOBSON RAMIN

T**RAVIS GUSTAVSON DIED** in February, 2021, in Mankato, Minnesota at the age of 21. The morning of the day he died, he had a tooth pulled at the dentist's office. Due to a drug history, the doctors didn't prescribe him strong painkillers, so he was planning to white knuckle it through the day with ibuprofen, according to his mother. Instead, he called a guy who sold him illegal street heroin and fentanyl. In a text to the dealer, Gustavson sent a photo of the amount he planned to take and asked if he had gotten the dose right. "Smaller bro" and "be careful plz!" the dealer wrote back. Gustavson overdosed.

Gustavson, whose death was reported in the *L.A. Times*,[1] is one of the many casualties of an opioid crisis that has ravaged the United States for over two decades. Opioid overdose deaths have claimed more than 600,000 lives in the U.S. and Canada since 1999, more than were lost in World War I and II combined. In both countries, 2020 was the worst year on record, in terms of total deaths and percentage increase from the previous year, precipitated in part by the anxiety, stress, and isolation of the COVID-19 pandemic.[2]

The urgent need for safer, better painkillers could not be more apparent. In 2016, as the FDA grappled with how to curb the escalating crisis, the agency wrote new rules that limited opioid prescriptions. But given that no powerful clinical alternatives to commonly prescribed opioids exist, these new rules had the unintended effect of driving more people to desperate measures. Some chronic pain patients, after their doses were cut, died by suicide.[3] Others sought relief in illict street drugs like heroin and fentanyl.

Though scientists have tried for decades to design drugs that are both more effective and less addictive than opiates, most efforts have run aground.[4] At the heart of these failures lies a fundamental misunderstanding of the nature of pain, says Clifford Woolf, a leading expert in pain information processing from Boston Children's Hospital. It is not a single phenomenon that can be addressed with a universal analgesic,[5] but consists of distinct types that demand targeted therapeutic approaches.

Woolf and colleagues at Harvard Medical School are convinced they are finally just a few years away from identifying powerful precision painkillers that could not only safely replace opiates, but effectively target distinct pain types. Their end game is to eliminate chronic pain altogether.

PRECISION PAINKILLERS

In 2006, Alex Binshtok, a young post-doc from Israel by way of Odessa, began a fortuitous collaboration with Woolf, already one of the world's foremost pain scientists, and neurobiologist Bruce Bean of Harvard Medical School. The three scientists had a bold idea for a way to create targeted painkillers without side effects. Anxious to produce results, Binshtok worked long days in the lab for months, patiently studying the hindpaws of rats, and sleeping among the rodent cages at night.

DRAWLAB19 / SHUTTERSTOCK

All neurons express multiple types of sodium, potassium, and calcium ion channels that together regulate their ability to translate stimuli from the external world to the rest of the nervous system. But Woolf, Bean, and Binshtok knew that pain-sensing neurons, called nociceptors, feature unique large-pore ion channels not found on other types of neurons—a discovery made in the late 1990s by David Julius and Ardem Patapoutian, who won the Nobel Prize for related work in 2021. Woolf, Bean, and Bishtok soon identified a way to exploit these unique ion channels to shut down signaling on activated mouse nociceptors, while leaving all other neurons untouched.[6]

The implications were vast. Because these compounds worked exclusively on nociceptors that were sensing painful stimuli, they could squelch pain at the source without disabling neurons engaged in other functions. By comparison, most clinically used analgesics tend to hit diverse targets in the central and peripheral nervous systems. Opioids, for example, act not just on nociceptors, but the brain, spinal cord, and the nervous system in the gut, which can lead to sedation, physical dependence, respiratory depression, and death.[7]

The scientists knew that making the leap from animal models and basic science—the research at the heart of human knowledge—to clinical drug development would be challenging, but they also understood that their discovery opened a major new avenue for the treatment of pain. Within a month, Woolf and Bean submitted their first collaborative paper on ion channel behavior to *Nature*, where it was published in 2007.[6]

A decade and a half later, their finding remains a big deal to the field of pain science, says Allan Basbaum, a professor at University of California, San Francisco, who studies the biology of chronic pain. "The big problem with drugs is that they don't know where to go," he says. Woolf, Bean, and Binshtok had devised a way to direct drugs to their intended pain targets.

I NEED A NEW DRUG

Despite the risks of addiction and overdose, opioids have been used to treat pain for millennia. Egyptian medical papyri from 1550 BC show that healers used poppy plants to cure breast abscesses and to calm crying children, and that tinctures of opium were mixed into eye drops and ointments.[8] But over the long term, opioids aren't an effective treatment for pain, not just because of the potential for deadly side effects, but because the brain demands higher and higher doses to produce the same effect. A recent review published in the *Journal of the American Medical Association* found that, across a variety of chronic conditions, opioids reduced pain by an average of less than one point on a 10-point scale.[9]

Less ancient options for pain mitigation—NSAIDs, acetaminophen, ibuprofen, corticosteroids, and anticonvulsant drugs that are prescribed to suppress nerve pain—are clinically effective in only a minority of patients with chronic pain and are also known to have deleterious effects on cells that are unrelated to their pain targets, including cardiovascular tissue.

In 2006, as Binshtok studied his rats, drug development had come to a standstill. Pharmaceutical companies introduced Cox-2 inhibitors in the late 1990s, only to have them black-boxed or banished from the market after reports of heart attacks and strokes.[10] When Merck paid out $4.5 billion in legal settlements related to Vioxx, Big Pharma grew leery of taking further chances.

Although few could foresee the oncoming tsunami of opioid-related overdoses and deaths that would unfold over the coming decades, it was already evident that in most developed nations, chronic pain was a serious problem, and that opioid prescriptions meant to treat conditions like bad backs and sore knees were flowing too freely. No one was more aware of this than Clifford Woolf, who had spent much of his career studying the biological mechanisms that underlie chronic pain.

Having revealed, in 1983, the principles behind central sensitization—in which neurons in the central nervous system continue to send pain signals to the brain long after an injury has healed—Woolf spent years searching for therapeutic interventions, compounds that would not carry opioids' liability for abuse and addiction and whose effects would apply narrowly to a specific target. He found one promising compound in the 1990s, but because it produced psychotic and hallucinogenic effects, it was shelved.[11]

That failure left Woolf even more determined to find a way to replace highly addictive opioids with safer, better drugs.

THE VARIETIES OF PAIN

For many decades, the medical profession saw pain exclusively as a symptom of other underlying problems, and treatment choices were determined only by its intensity: mild, moderate, or severe. But pain is various and multifaceted, which is what makes it so difficult to treat. Causes can include cancer, inflammation or tissue injury, injury or lesions of the nervous system, or, as in central sensitization, abnormal amplification of pain signaling in the central nervous system. Or it can manifest as a combination of these subtypes. In some cases, pain is not just a symptom but the disease itself.[5]

With the diversity of pain in mind, over the past decade Bean and Woolf have devoted much of their energy to learning about different nociceptor cell types. Nociceptors, specialized peripheral sensory neurons that sense noxious stimuli, were discovered in 1906 by Charles Scott Sherrington, an English scientist who later won the Nobel Prize for related work.[12] These neurons convert potentially damaging stimuli into electrical signals—action potentials—that are then relayed to higher brain centers. But different types of nociceptors respond to different types of stimuli. Thus, the nociceptors that send messages notifying us of the onset of appendicitis are distinct from those that make our knees and hips feel as creaky as the Tin Man when we get out of bed.[13]

Their discovery opened a major new avenue for the treatment of pain.

Woolf and Bean have collaborated in their work on nociceptor types with eminent neurobiologist David Ginty, at Harvard Medical School. Ginty's research has been focused on the neural circuits that are involved in the perception of touch, in all its subtle variations: how the brain perceives the difference between a warm breeze on the cheek and a blow with a blunt object. Of the 14 to 20 neuron subtypes involved in touch, Ginty has found six sensory neurons (thus far) that are specifically tied to pain. These are the nociceptors that Woolf and Bean are studying.

"We study these sensory neurons in great detail," says Ginty. If they can figure out exactly what kinds of stimuli different types of nociceptors respond to and how they send signals to the spinal cord and the brain, this will help them figure out how to narrowly treat specific kinds of pain. For example, they could selectively inhibit the pain associated with colitis, or with cancer, without blocking pain from the skin, which is protective. Or they could block all nociceptors, and therefore all pain signals, at once, which could be useful during surgery or after major trauma. "That will allow us to consider translational opportunities and think about druggable targets," says Ginty.

The compounds Woolf and Bean identified in 2007, which they call nocions, are currently being shepherded through clinical trials via a well-funded joint venture called Nocion Therapeutics. But nocions, which can be used to treat pain, cough, and itch, must be administered topically, which limits their therapeutic application. To address the full spectrum of pain, Woolf, Bean, and Ginty are on the hunt for novel analgesic compounds that can work selectively but *systemically*—drugs that treat specific pain types, but can be taken in pill form.

FATAL MISTAKES Opioid overdose has killed more people in the U.S. and Canada since 1999 than World Wars I and II combined. Efforts to design replacement drugs for opioids have run aground because of a fundamental misunderstanding of the nature of pain.

GLOBALNEWSART.COM / SHUTTERSTOCK

To really move forward with this work, the three scientists needed a way to eliminate the variables associated with studying rodent models.

In 2006, after the publication of a paper in *Nature* that explored genetic insensitivity to pain in a Pakistani family with a mutation in the gene *SCN9A*, which affects the behavior of a specific ion channel,[14] pharmaceutical companies invested heavily to develop drugs that targeted that ion channel. The compounds they developed performed very well in rodents. But when those same compounds were tested on humans, the analgesic effect disappeared. “We still don’t really understand why it didn’t work,” says Bean. “It could be that it’s because human nerve cells are larger than those of rats, or that the barrier between blood and nerve cells is different.”

The Harvard scientists knew that they’d have better outcomes if they could conduct their experiments on human nociceptors gleaned from induced pluripotent stem cells, which begin as skin or blood cells and can be reprogrammed into any type of cell in the body. This “disease in a dish” model was rapidly becoming the standard in laboratory science, leaving rodent models behind in favor of stem cells in a Petri dish. But despite many attempts, Woolf and Bean had never achieved the enormous supply of high-quality cells required to do large-sale drug screening.

> *“The big problem with drugs is that they don’t know where to go.”*

FROM DISH TO DRUG—THE LONG ROAD AHEAD

In February of 2019, at a meeting in Bethesda of the HEAL Initiative (HEAL stands for Helping to End Addiction Long-term) the NIH’s half-billion-dollar plan meant to stem the tide of the opioid crisis, Woolf ran into Don Lo, an old friend from academic scientific circles. Lo, who’d recently been appointed to a key role in drug innovation at the National Center for Advancing Translational Sciences (NCATS), said that instead of heading back to Boston, Woolf would do well to show up at the post-meeting poster session in the NIH cafeteria. There was someone he should meet.

While other attendees chatted over coffee, Illyas Singeç, who directs the Stem Cell Translation Laboratory at NCATS, stood alone in front of a poster covered with unpublished data. After a few minutes of conversation, it was obvious to Woolf that Singeç knew how to produce the high-quality stem cells that he and Bean so desperately needed, at scale. "It was an enormously important moment for pain science. We needed the wheel," Woolf says simply, "and he had invented it."

The first batch of neurons arrived in Boston in June of 2019. Within a few months, in cooperation with a large Harvard Medical School consortium called STOP PAIN,[15] Woolf and Bean could do industrial-scale drug screening in an academic setting, making it possible to study hundreds of thousands of compounds. The cells, made at NCATS, were frozen and shipped, and the final stage of the differentiation process was completed in the lab in Boston.

The Harvard consortium, composed of scientists and laboratories from multiple disciplines, gave Woolf and Bean access to a team of medicinal chemists, who could help them select and optimize the most promising drug targets. "We're the biologists," Woolf says. "We devise the screens, and when we get a hit—and we're getting them steadily—the chemists tell us whether it has what it needs to be a drug: that it's stable and doesn't break down in a few minutes; that it doesn't have adverse effects." STOP PAIN's computer engineering arm added computational modeling to the arsenal, making it possible for Woolf and Bean to identify the exact molecular and biological mechanisms by which the screened compounds work.

The urgent need for safer, better painkillers could not be more apparent.

"If I think back to where I was when I started my career," Woolf says, "the possibilities that are now available and the pace at which the technology advances—what took 20 years, we can now do in two months. It's an incredibly exciting time to be working in the nervous system. When I began, it was a black box. Now we can begin to identify the way it changes and functions and identify new therapeutic strategies."

Although Woolf and Bean have a solid head start, the protocol for making these differentiated cells, developed at NCATS, is not exclusive to Harvard's STOP consortium, or to any organization. Last year, Bean, Singeç, and Woolf published a paper in *Science Translational Medicine* that details the process of finding nociceptor-selective treatments.[16] "Our goal," Singeç says, "is to disseminate the technology as much as we can, to academic labs and pharmaceutical companies over the next several years."

DIRTY DESIGNER DRUGS

Woolf and Bean hypothesize that previous compounds intended to work in ion channels—including ones in which pharmaceutical companies invested millions—failed because they targeted a single channel, allowing another channel on the same cell to step in and do the work of the one that

The overriding goal is to eliminate the need for opioids for pain relief.

they'd successfully disabled. For that reason, the investigators are designing intentionally "dirty" drugs—ones that will hit multiple ion channel targets—sodium, potassium, and calcium channels that are responsible for different aspects of function on different types of nociceptors.

The big win, says Bean, would be to develop a compound that is useful for the treatment of persistent pain. The statistics back him up: A 2021 paper in *Pain* noted that in 2019, 20.5 percent of U.S. adults suffered from chronically painful conditions.[17]

Once these drug candidates have been fully evaluated—the scientists believe that this will take two or three years—they'll be submitted to the FDA for new drug designation, at which point hungry pharmaceutical companies will hopefully scoop them up and shepherd them through the multi-year phases of clinical trials, which for pain and anesthesia drugs routinely cost more than $100 million.[18] (For the drug companies, it should be worth it: In 2021, the market for pain drugs was worth $31 billion, with projections of $39 billion in sales by 2029.)[19]

The overriding goal, Bean and Woolf agree, is to eliminate the need for opioids for pain relief, by developing compounds that are far more effective, and still safe, with no liability for abuse. Bean says that if they're successful in the next two to three years, it would represent the ultimate achievement in a lifetime of work. But even if they're not, they will develop a better understanding of the biology, and a better foundation for the next try. "Each time you try, you set the stage—you help scientists with the next step. We're convinced that selective inhibiting of these cells can be done. It will work."

CATHRYN JAKOBSON RAMIN is an investigative journalist and the author of two books: *Crooked*, about the back pain industry, and *Carved in Sand*, about what happens to memory and attention in midlife and beyond. Both are published by HarperCollins.

REFERENCES

1. Stobbe, M. U.S. drug overdose deaths top 100,000 in a year for the first time, officials say. *Los Angeles Times* (2021).

2. Humphreys, K. Responding to the opioid crisis in North America and beyond: recommendations of the Stanford–Lancet Commission. *The Lancet Commissions* **399**, 555-604 (2022).

3. Szalavitz, M. Cracking down on opioids hurts people with chronic pain. *Vice* (2017).

4. Safavi-Hemami, H. Pain therapeutics from cone snail venoms: From Ziconotide to novel non-opioid pathways. *Journal of Proteomics* **190**, 12-20 (2019).

5. Woolf, C.J. Capturing novel non-opioid pain targets. *Biological Psychiatry* **87**, 74–81 (2020).

6. Binshtok, A., Bean, B.P., & Woold, C.J. Inhibition of nociceptors by TRPV1-mediated entry of impermeant sodium channel blockers. *Nature* **449**, 607–610 (2007).

7. Jayakar, S. *et al.* Developing nociceptor-selective treatments for acute and chronic pain. *Science Translational Medicine* **13**, eabj9837 (2021).

8. Ross, A.M. Poppy and opium in ancient times: Remedy or narcotic? *Biomedicine International* **1**, 81-87 (2010).

9. Busse J.W., *et al.* Opioids for chronic noncancer pain: A systematic review and also meta-analysis. *Journal of the American Medical Association* **320**, 2448–2460 (2018).

10. Halpern, G.M. COX-2 inhibitors: a story of greed, deception and death. *Inflammopharmacology* **13**, 419-425 (2005).

11. Woolf, C.J. Evidence for a central component of post-injury pain hypersensitivity. *Nature* **306**, 686-688 (1983).

12. Woolf, C.J. Nociceptors–noxious stimulus detectors. *Neuron* **55**, 353-364 (2007).

13. Dubin, A. & Patapoutian, A. Nociceptors: the sensors of the pain pathway. *Journal of Clinical Investigation* **120**, 3760–3772 (2010).

14. Cox, J.J., *et al.* An *SCN9A* channelopathy causes congenital inability to experience pain. *Nature* **444**, 894–898 (2006).

15. Jiang, K. Stopping pain: New consortium aims to transform the treatment of pain. Harvard Medical School (2019).

16. Jayakar, S., *et al.* Developing nociceptor-selective treatments for acute and chronic pain. *Science Translational Medicine* **13**, eabj9837 (2021).

17. Yong, R.J., Mullins, P. M., & Bhattacharyya, N. Prevalence of chronic pain among adults in the United States. *Pain* **163**, e328–e332 (2022).

18. Ledesma, P. How much does a clinical trial cost? Sofpromed (2020).

19. Telugunta, R., Srivastava, A., & Sumant, O. U.S. pain management drugs market. Allied Market Research (2021).

The Power of Narrative

Global solutions, like the greening of the world's financial system, begin with the right story

BY KLAUS SCHWAB & THIERRY MALLERET

WHAT FUTURE DO WE FACE? What future do we want? What must we do to get there?

These three questions preoccupy us all. We can't predict the future. However, we can imagine it and even design it; no outcome is predetermined and, as cognitive human beings, we retain the agency to shape the world we want. Perhaps most critically, we can also prepare for the future, by confronting both the risks that we can mitigate and the things that will surprise us. As the most effective conduits for ideas, narratives have the unique power to help us determine what's going on, what lies ahead, and what needs to be done.

Defined in the simplest possible terms, a narrative is a story about something. Stories are essential to us because as human beings and social animals, we are storytelling creatures. In his 1938 novel Nausea, Jean-Paul Sartre wrote, "A man is always a teller of tales, he lives surrounded by his stories and the stories of others, he sees everything that happens to him through them; and he tries to live his life as if he were recounting it." Robert Shiller, the "father" of narrative economics, goes one step further, linking narratives to the decisions we make: "The human brain has always been highly tuned towards narratives, whether factual or not, to justify ongoing actions."

The rich scholarly literature about narratives makes it clear that we think, act, and communicate in terms of narratives, and each interpretation, understanding, or model of how the world operates begins with a story. Narratives provide the context in which the facts we observe can be interpreted, understood, and acted upon. In that sense, they equate to much more than the stories we tell, write, or illustrate figuratively; they end up being the truths, or the ideas we accept as truths, that underpin the perceptions that shape our

"realities" and in the process form our cultures and societies. Through narratives, we explain how we see things, how these things work, how we make decisions and justify them, how we understand our place in the world, and how we try to persuade others to embrace our beliefs and values. Narratives shape our perceptions, which in turn form our realities and end up influencing our choices and actions. They are how we find meaning in life.

Our narratives have carved the paths of history, some down the darkest roads. But our narratives also bring out the best in us. In his book, *The Better Angels of Our Nature*, Harvard professor of psychology Steven Pinker makes the vibrant plea that the main line of history is one of improvement and that today's world is richer, healthier, and safer than it's ever been. This is correct: On most metrics, if we had to choose throughout history the best time to be alive, it would indeed have to be today. And if we had to choose one quality that powers the better angels of our nature? Imagination.

Imagination is a glorious attribute. When its infinite possibilities are harnessed, it corresponds to a form of "superpower" from which every human being can benefit, individually or collectively. It's easy to understand why: Every resolution of a problem begins with a bit of imagination. Conversely, many problems occur because of the failure to imagine that they could.

The expression "failure of imagination" captures this by describing the expectation that future opportunities and risks will resemble those of the past. Novelist Graham Greene used it in *The Power and the Glory*, but the 9/11 Commission made it popular by invoking it as the main reason why intelligence agencies had failed to anticipate the "unimaginable" events of that day.

Ever since, the expression has been associated with situations in which strategic thinking and risk management are stuck in unimaginative and reactive thinking. Considering today's wide and interdependent array of risks, we can't afford to be unimaginative, even though, as the astrobiologist Caleb Scharf points out, we risk getting imprisoned in a dangerous cognitive lockdown because of the magnitude of the task. "Indeed, we humans do seem to struggle in general when too many new things are thrown at us at once. Especially when those things are outside of our normal purview. Like, well, weird viruses or new climate patterns," Scharf writes. "In the face of such things, we can simply go into a state of cognitive lockdown, flipping from one small piece of the problem to another and not quite building a cohesive whole."

Imagination is precisely what is required to escape a state of "cognitive lockdown" and to build a "cohesive whole." It gives us the capacity to dream up innovative solutions to successfully address the multitude of risks that confront us. For decades now, we've been destabilizing the world, having failed to imagine the consequences of our actions on our societies and our biosphere, and the way in which they are connected. Now, following this failure and the stark realization of what it has entailed, we need to do just the opposite: rely on the power of imagination to get us out of the holes we've dug ourselves into. It is incumbent upon us to imagine the contours of a more equitable and sustainable world. Imagination being boundless, the variety of social, economic, and political solutions is infinite.

Narratives shape our perceptions, which form our realities and influence our choices.

WITH RESPECT TO THE ASSERTION that there are things we don't imagine to be socially or politically possible, a recent book shows that nothing is preordained. We are in fact only bound by the power of our own imaginations. In *The Dawn of Everything*, David Graeber and David Wengrow (an anthropologist and an archaeologist) prove this by showing that every imaginable form of social and economic organization has existed from the very beginning of humankind. Over the past 300,000 years, we've pursued knowledge, experimentation, happiness, development, freedom, and other human endeavors in myriad different ways. During these times that preceded our modern world, none of the arrangements that we devised to live together exhibited a single point of origin or an invariant pattern. Early societies were peaceful and violent, authoritarian and democratic, patriarchal and matriarchal, slaveholding and abolitionist, some moving between different types of organizations all the time, others not. Antique industrial cities were flourishing at the heart of empires while others existed in the absence of a sovereign entity.

The point is this: Prior to our "modernity" that culminated with the idea of the Enlightenment, there was an incredible variety of social possibilities that most modern people didn't imagine were possible. Rousseau's idealism about the "state of nature" in a pre-civilization world or Hobbes' conviction that life before the state existed was "nasty, brutish and short" look more like a hypothesis than the scientific canon they often incarnate. If ancient societies did not obey any specific organizational trajectory or evolutionary model, why should it be different in today's world? What prevents us from imagining different pathways and new forms? What is it that constrains our ability to imagine better ways of dealing with our problems and why can't we organize ourselves differently? Are we not capable of creating entirely new narratives?

The answer is: Yes, we are! The definition of imagination as "the ability to confront and deal with a problem" involves creativity and an openness to new ways of thinking, plus of course large amounts of disciplined analysis and the prospect of a business or policy application. Otherwise, to paraphrase the Brazilian economist Carlos Braga's favorite expression, "vision or imagination without implementation is just hallucination."

PAGES 92-93: PHOTOCREO MICHAL BEDNAREK / SHUTTERSTOCK

RICHARD WHITCOMBE / SHUTTERSTOCK

ECONOMIC CUT "Nature is not free," write Klaus Schwab and Thierry Malleret. "It represents an indispensable input to economic activity. It is an asset. We need to treat it as such and in the process reconsider our measures of economic prosperity."

Today, all sorts of people are engaged in elaborating novel and imaginative ideas, products, and strategies. They do so by developing new ventures, start-ups, economic policies, or mammoth projects and, in so doing, create and shape the future. Their original ideas translate into narratives that produce models which in turn influence behavior and help construct the future. Ultimately, they become instruments of policy and project market power.

When its infinite possibilities are harnessed, imagination corresponds to a superpower.

In the environmental sector, the stakes for our future are the highest. Imaginative nature-based solutions are leading to a bloom of start-ups. It doesn't take much imagination to realize that nature gives freely. For centuries, not only have we taken this generosity for granted, but we've also exploited it to such an extent that we are now on the edge of a precipice.

Of course, nature is not "free"; it is priceless, and a degree of imagination is needed to grasp what this

The idea is far-fetched, but it's already been picked by hard-nosed investment bankers.

means in terms of policy. For economists and policymakers, valuing nature and assessing its contribution to our economies and societies is a recent endeavor because, for centuries, we've been overlooking the fundamental role nature plays in our lives and underestimating the risks that environmental degradation poses to human welfare and economic growth. Without taking care of the complex ecosystems that ensure that the temperatures remain tolerable, the air breathable and the water drinkable, we simply cannot function as societies. It follows that nature represents an indispensable input to economic activity. It is an asset. We need to treat it as such and in the process reconsider our measures of economic prosperity.

Recent economic studies conclude that more than half of the world's GDP depends on nature. It renders services that are obvious (like fish stocks and bee pollination), and others that are much more difficult to comprehend (like the work of complex ecosystems within the soil that recycle nutrients, purify water, and absorb atmospheric carbon). These are unfamiliar topics for economists and require a lot of imagination and disciplined analysis on their part to properly account for nature's contributions to economic growth. As stated recently in the Dasgupta Review, "humanity has prospered immensely in recent decades, the ways in which we have achieved such prosperity means that it has come at a devastating cost to nature. Estimates of our total impact on nature suggest that we would require 1.6 Earths to maintain the world's current living standards."

Global solutions are going to come from undiscovered shores, and already we see them on the horizon. The Network for Greening the Financial System (NGFS) is a group of 91 central banks and supervisors committed to mobilizing mainstream finance to support the transition toward a sustainable economy. It is investigating many bold financial innovations that could (and most likely will) one day revolutionize the way in which climate-related risks are accounted for in central banking and banking supervision. In short, alongside governments (which have a much broader and more effective range of tools and policies available to prevent and mitigate climate-related risks), central banks will adapt their monetary policy operational frameworks to reflect climate-related risks. This will involve the mitigation of balance sheet risks that stem from climate change and environmental degradation, but also the active support of the transition to a non-carbon, green economy.

Imagining what form this might take and devising policy tools and instruments to get there is the task of the NGFS, and largely depends on how climate risks will affect the economy and financial system through a range of different transmission channels. The menu of options available is extensive and encompasses changes in all three most important policy fields of a central bank: credit operations, collateral policies, and asset purchases. Some of the options represent a radical departure from standard central bank operational policies. They are, in short, the product of central bankers' imagination.

Some ideas go into uncharted territory, well beyond the scope of what the NGFS is devising in terms of possible policies. Creating "carbon quantitative easing" policies is one of them. It's a novel, untested, and somewhat outlier narrative that already sounds familiar because it plays a key role in Kim Stanley Robinson's novel *The Ministry for the Future*. In the novel, as the climate crisis gets dramatically worse, the world's top central bankers end up cooperating by collectively abandoning caution to the wind and deciding to create a global "carbon coin" to fund decarbonization. Robinson's inspiration is grounded on some academic papers published less than five years ago, and lately on the work of the NGFS, that doesn't yet venture into such wild territory. But asserting novelists' "poetic license," Robinson's imagination runs wild:

> [The NGFS] suggested that possibly nations, companies and individuals who draw carbon from the atmosphere could be paid for it directly. Possibly petrostates could be compensated for the fossil fuels they keep in the ground. Possibly oil companies could be paid to suck carbon from the air and then pump it back into the ground; they could also be paid to pump water from under the great glaciers of Antarctica and Greenland, which are currently sliding into the sea on newly melted subterranean water slides. Of course, legislatures and citizens will need to urge their central banks, and ultimately to instruct or order them, to do these things. But the good news is that with these new strategies in hand, even in our current political economy, awkwardly suited at best to the task at hand, we might be able to pay ourselves to do the necessary things, and thus dodge the coming mass-extinction event.

This idea is far-fetched, but it's already been picked up by some hard-nosed investment bankers. Besides, surely the very purpose of imagination is to venture into unknown territories!

KLAUS SCHWAB is the founder and executive chairman of the World Economic Forum. His latest books are *The Fourth Industrial Revolution*, a worldwide best-seller translated into 30 languages, and *Shaping the Future of the Fourth Industrial Revolution*.

THIERRY MALLERET is the managing partner of the Monthly Barometer, a succinct predictive analysis provided to private investors, global CEOs, and opinion and decision-makers. He has written several business and academic books and has published four novels.


JACKSON WILD SUMMIT
SEPT 26
2022
SEPT 30
Lake Neusiedl - Seewinkel National Park, Burgenland, AUSTRIA
JACKSONWILD.ORG
© St. Martins Therme & Lodge

CULTURAL PROGRAMS OF THE NATIONAL ACADEMY OF SCIENCES

Convergence II

The Art Collection of the National Academy of Sciences

Science relies on the rational. Art celebrates the intuitive.

This second volume catalogue of the art collection of the National Academy of Sciences illustrates by example how the art exhibits and collection provide a platform for thinking about the impact of science on our society. Through a cultural and personal lens, the topics found within the collection range from diversity and equity to climate change and collaboration.

This catalogue contains beautiful reproductions of more than 30 original artworks from the NAS collection including such artists as **Tim Rollins and K.O.S.**, **Mike and Doug Starn**, **Tavares Strachen**, **Neri Oxman**, **Camille Seaman**, **Diane Burko**, **Renée Stout**, **Alfredo Arreguín**, **Jeffrey Kent**, **David Maisel**, and many more. Essays from leading scholars and curators are included that illuminate the work.

Pre-order and make a donation to receive a gift!

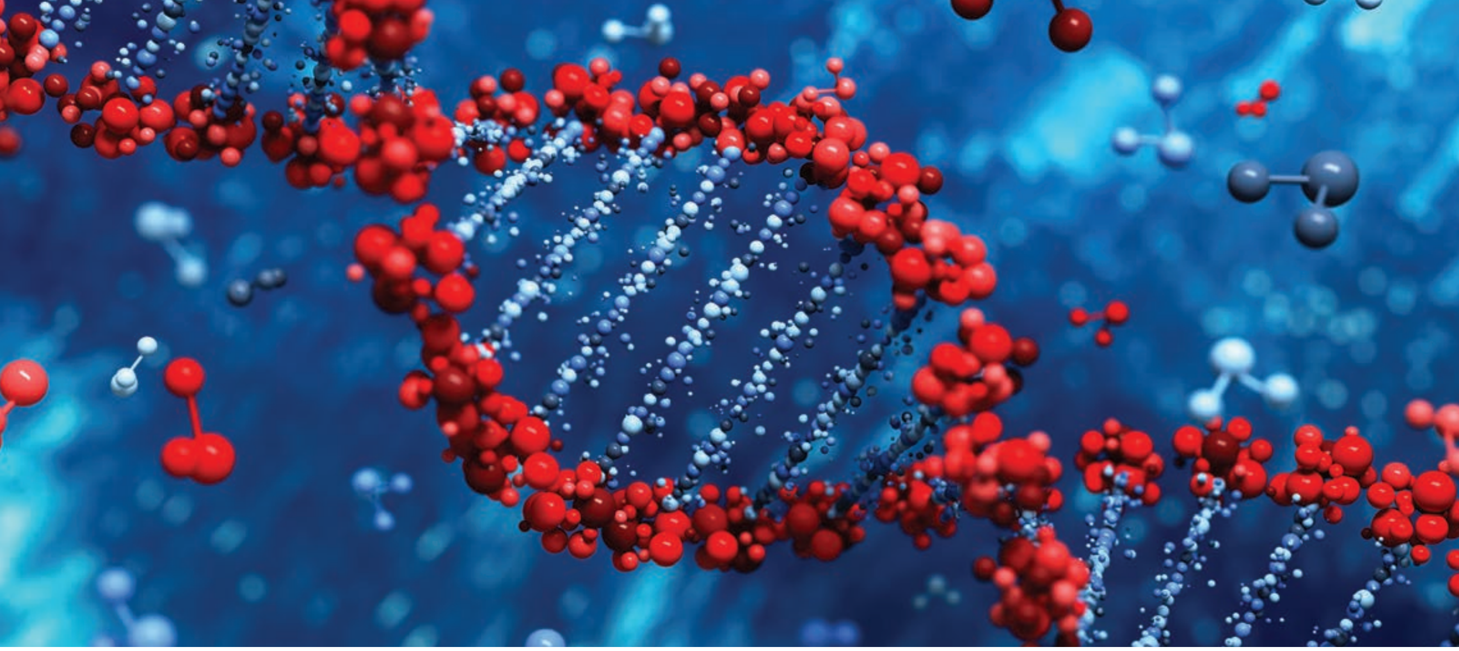
Pre-order now with a donation to Cultural Programs of the NAS and receive a gift of appreciation – a copy of *Convergence II* in a beautiful clam shell case along with a selection of one or more, depending on the donation level, limited-edition prints by: **Tim Rollins and K.O.S.**, **Diane Burko** and **Alfredo Arreguín**.

Limited edition prints are available while supplies last.

CPNAS CULTURAL PROGRAMS OF THE NATIONAL ACADEMY OF SCIENCES

FOR MORE INFORMATION, VISIT WWW.CPNAS.ORG.

@CPNAS



SCIENCE PHILANTHROPY ALLIANCE

Advancing basic science—The foundation of our future

sciphil.org

Pocket is where everything you save goes

Pocket scours the internet to bring you a curated mix of fascinating articles like these and a beautiful space to collect them in.

Sign up and save anything that sparks your curiosity in your Pocket to absorb when you're ready.

www.getpocket.com

Issue Index

43

MARCH / APRIL 2022

SUBSCRIPTIONS
Address subscription and address changes to:
Nautilus
360 W 36th Street, 7S
New York, NY 10018;
email: order@nautil.us

Visit shop.nautil.us for more information.

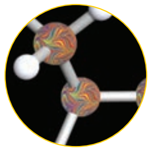
POSTMASTER
Send address changes to:
Nautilus
360 W 36th Street, 7S
New York, NY 10018

28

2019
75 STORY
76 LANGUAGE

27

2019
72 QUANDARY
73 PLAY
74 NETWORKS



26

2019
68 CONTEXT
69 PATTERNS
70 VARIABLES
71 FLOW

25

2018
A SELECTION OF STORIES FROM THE APR - DEC 2018 ONLINE ISSUES.

24

2018
56 PERSPECTIVE
57 COMMUNITIES
58 SELF

23

NOV/DEC 2017
54 THE UNSPOKEN
55 TRUST

22

SEPT/OCT 2017
52 THE HIVE
53 MONSTERS

21

JULY/AUG 2017
50 EMERGENCE
51 LIMITS

20

MAY/JUNE 2017
48 CHAOS
49 THE ABSURD

19

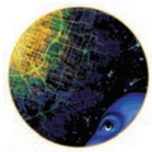
MAR/APR 2017
46 BALANCE
47 CONSCIOUSNESS

18

JAN/FEB 2017
44 LUCK
45 POWER

17

NOV/DEC 2016
42 FAKES
43 HEROES



16

SEP/OCT 2016
40 LEARNING
41 SELECTION

15

JULY/AUG 2016
38 NOISE
39 SPORT

14

MAY/JUNE 2016
36 AGING
37 CURRENTS

13

MAR/APR 2016
34 ADAPTATION
35 BOUNDARIES

12

JAN/FEB 2016
32 SPACE
33 ATTRACTION

11

NOV/DEC 2015
30 IDENTITY
31 STRESS

10

SEP/OCT 2015
28 2050
29 SCALING

09

JULY/AUG 2015
25 WATER
26 COLOR
27 DARK MATTER

08

MAY/JUNE 2015
22 SLOW
23 DOMINOES
24 ERROR

07

MAR/APR 2015
19 ILLUSIONS
20 CREATIVITY
21 INFORMATION

06

WINTER 2015
16 NOTHINGNESS
17 BIG BANGS
18 GENIUS

05

FALL 2014
13 SYMMETRY
14 MUTATION
15 TURBULENCE

04

SUMMER 2014
10 MERGERS & ACQUISITIONS
11 LIGHT
12 FEEDBACK

03

SPRING 2014
07 WASTE
08 HOME
09 TIME

02

WINTER 2014
04 THE UNLIKELY
05 FAME
06 SECRET CODES

01

FALL 2013
01 HUMAN UNIQUENI
02 UNCERTAINTY
03 IN TRANSIT

Alva Noe

The philosopher on the narrow lens of neuroscience

INTERVIEW BY KEVIN BERGER

BRAIN IN A VAT

The conception of the human being that neuroscience works with is ridiculously impoverished. "The brain is a computer." "Our minds are inside of us." "Each of us is a brain in a vat." There's just a whole series of pictures that represent us as solitary aliens, islands unto ourselves. Neuroscience desperately needs to develop richer, more positive ways of understanding what it is to be human.

MY LIMBIC SELF

Neuroscience is the new jargon. I hear people all the time say things like, "Oh, yeah, I have prefrontal issues." "My limbic self is fine but prefrontally I need a lot of help." These just become pictures. And it's not as though we really understand rationality or decision-making in terms of the prefrontal cortex. Thanks to popular science, neuroscience offers an on-the-shelf account of personality, emotions, memory, perception, consciousness, sex, gender, love. But these are all profoundly fraught areas. There's a kind of cultural hegemony of neuroscience—even though this is probably a disease of the educated classes—that's a form of ignorance. It's a lack of irony, a lack of sophistication about the fact that science is a work in progress. Neuroscience is a work in progress, the human being is a work in progress.

THE TRUTH ABOUT YOU

Art is revolutionary. It's the seed. It's where humans break the molds, which we have created for ourselves. Science is where we test hypotheses. That requires precision, that requires definitions, that requires regimentation. In the domain of aesthetic experience, there's something intellectually bullying about neuroscience. A priest tells you, "Here's the truth about you." Neuroscience says, "Here's the truth about you." What is it about the prestige of science that makes so many of us want to say, "Oh, yes, that's the truth about me: I really am just a brain. I really am just a collection of snap, crackles, and pops in my nervous system."

MONDRIAN AND A HIGHLIGHTER PEN

Neuroscience doesn't have the resources to differentiate works of art from other perceptual stimuli. It essentially investigates a class of stimuli that it arbitrarily calls art. If looking at a Mondrian is interesting because it activates the color experience in my brain, well, so is looking at this highlighter pen. It's true that I couldn't see a Mondrian if not for my color-sensitive perceptual capacity. But I want to know, "What makes the Mondrian artistically interesting and the pen not?" I don't think neuroscience has imagined a way to answer that question.

ART QUESTIONS EVERYTHING

When neuroscientists reflect on their own projects, they tend to have individualistic ways of understanding what they're doing. Art has no presuppositions. Art takes the presuppositions that we have and disrupts them. Art questions everything. The distinctive feature of life is that it is organized and regimented—whether at the cellular level, the organismic level, or the cultural level. Art is the spoke in the wheel of organization. It's the most profoundly effective method we have for exploring who and what we are. The ways we explore ourselves through art not only unveil us to ourselves but do so in ways that make us different. Art is the work of self-understanding that carries in its wake the possibility of change.

SERENA CAMPANINI


We can do this.
GenOcean.org
Join the campaign to **restore** and **protect** the ocean we need.
courageous
unesco
Intergovernmental Oceanographic Commission
2021 2030 United Nations Decade of Ocean Science for Sustainable Development