SCIENTIFIC AMERICAN

Sleep Engineering

Precision Cancer Treatments

AI and Robots Team Up

Exploring the Sun

A total solar eclipse and new space probes will reveal our star's mysteries

MARCH 2024
SCIENTIFICAMERICAN.COM

SCIENTIFIC AMERICAN®

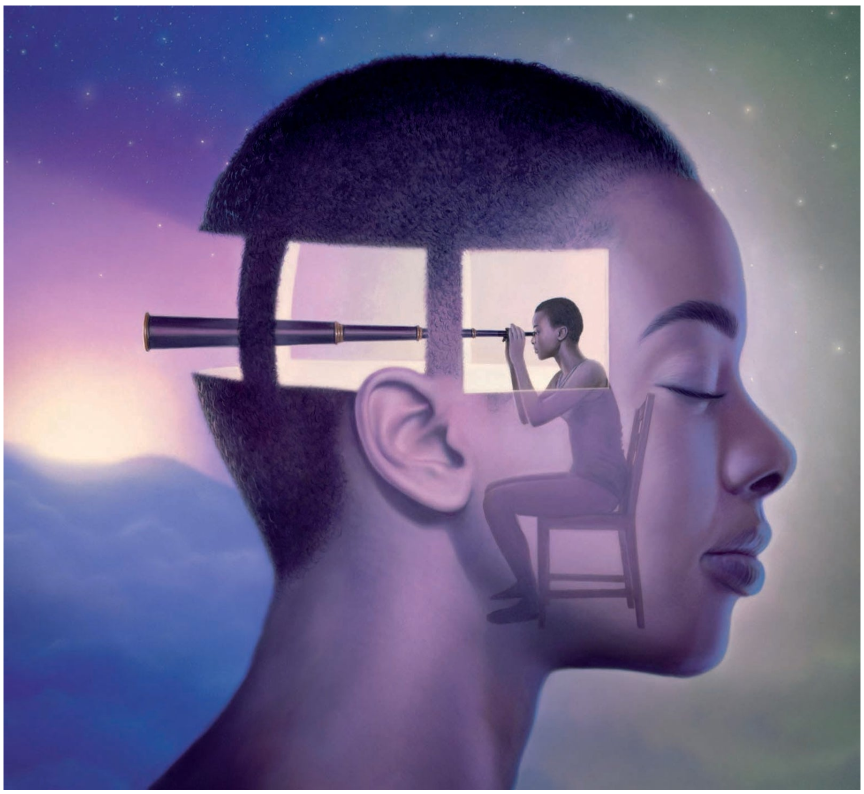
[March 2024]

Features

Hacking Dreams Could Help People Heal

Stimulating the sleeping brain may ease suffering from memory loss, stroke or mental health problems

By Ingrid Wickelgren


Tim O'Brien

It was late, and Sonia was alone in an unfamiliar town, trying to find her way home. The map showed a route through a dark forest lit by an occasional lantern. She viewed it with foreboding but, seeing other people also using this passage, took it. Walking fast, she neared a couple ahead of her—a man and a woman—who suddenly stopped, turned and grabbed her. The man covered her face with a cloth. She found herself on a stage with a ceiling spanned by a mirror. A crowd of men armed with guns and knives encircled her; she was about to be tortured and killed. Sonia picked up a stone and threw it at the ceiling, which shattered. Pieces of glass rained down, piercing her shoulder and foot. She fled into the forest, pursued by the couple, who could read each other's minds. The woman saw where Sonia was running and informed the man—Sonia knew she would be hunted down.

This nightmare and similar ones disturbed Sonia's sleep about twice a week for months. (Her real name has been withheld for privacy.) Those awful nights left her sleepy, irritable and emotionally spent—symptoms of nightmare disorder. The condition can occur by itself or alongside deeper issues such as post-traumatic stress or anxiety disorders. Sleep specialists at the Geneva University Hospitals prescribed "imagery rehearsal" therapy. Sonia was to create a positive ending for a bad dream and practice it daily. A fresh take on a dream tends to carry over into sleep, reducing the frequency of nightmares.

But the trick doesn't always work, so Sonia joined a study to test an embellished version of it. The trial leveraged sleep's power to fortify memories—in this instance, the new dream narrative. For five minutes each evening over two weeks, Sonia relaxed in a quiet space at home and imagined that the route through the forest led to a door that opened onto a bright, colorful field that felt safe. While she and 17 other people with nightmare disorder rehearsed their new storylines, they listened through headphones to a piano chord that was played every 10 seconds, eventually associating the sound with the narrative.

And throughout that fortnight, they wore a sleep-engineering headband when they went to bed. The device detected when the participants entered rapid eye movement (REM) sleep (so named because the eyes dart from side to side during this phase), when people experience their most vivid dreams. While they dreamed, the headband transmitted, through the bones of their skull, the same piano chord they had heard while awake.

During sleep the brain replays select memories from the day to emblazon them into its neurons. Experts call this process memory consolidation. In the nightmare study, the chord reminded the participants of their happier dreams. “We want to enhance this specific memory,” says psychiatrist Lampros Perogamvros of the University of Geneva, who led the research.

The association led people to experience fewer nightmares and more positive dreams overall. “Even if you work with only one scenario when you're awake, nightmares about any kind of theme [such as being chased] go down,” Perogamvros says. The effect was significantly stronger for those who heard the chord while rehearsing their revised dream than for those who had not, the researchers reported in 2022. Sonia, for one, stopped having nightmares altogether, and her mood improved.

Manipulating sleep might be a new route out of the proverbial forest—whether the affliction is nightmares or a problem with mood, memory or even motor skills. “Sleep is an unguarded time. It's a time when our executive control, our rational thinking, our logical decision-making, our impulse control are turned off. So stimuli that manage to get in are processed differently and possibly more effectively,” says Robert Stickgold, a cognitive neuroscientist at Harvard Medical School.

The techniques investigators use to “get in” while someone is asleep range from electrically stimulating the patient's brain to exposing them to sounds or smells that remind them of specific facts or experiences. Many of these techniques were devised to decode sleep's role in memory and cognition. But they also offer ways to speed recovery from stroke or to restore memories lost with age. They might even be able to tamp down negative emotions attached to specific memories, which could help ease post-traumatic stress, anxiety, depression, or other mental health conditions.

“One of my latest hopes is that we can have new methods to help people wake up on the right side of the bed,” says Ken Paller, a memory and sleep researcher at Northwestern University. To make such methods practical and widespread, researchers are developing a range of sleep-engineering devices people can use at home. Experts say clinical use of some of the devices is years away, and they also warn of potential risks.

Messing with memory could have unforeseen consequences, such as creating imbalances that impede learning, says neuroscientist Gina Poe of the University of California, Los Angeles. “It's kind of a scary time,” Poe says. “We don't know enough. It's kind of like [being] a toddler. We can walk but don't know where we are going or how to avoid danger.”

Scholars have suspected that sleep shores up memories for millennia. In the first century C.E., Roman writer and teacher Marcus Fabius Quintillian wrote “that the interval of a single night will greatly increase the strength of the memory.” The details of this process remained obscure until the 20th century, when the invention of the electroencephalogram, a recording of brain activity made by an array of probes placed on the scalp, spawned studies showing that the sleeping brain whirs to its own electrical rhythms.

People sleep in cycles that repeat roughly every 90 minutes and usually go through a total of four to six cycles over a full night's sleep. The first cycle starts with a period of light sleep, which has two distinct stages. During the second stage, neurons produce clusters of electrical signals called sleep spindles—evidently because when drawn as a graph of voltage changing with time, they reminded scientists of wool wound on a stick. Light sleep descends into deep “slow wave” sleep, in which the spindles continue while slow, rhythmic pulses of electrical excitation sweep across the brain, overlaid with bursts of high-frequency “ripples.” In REM sleep, the fourth stage, brain neurons fire as actively and randomly as they do during the day, and people experience emotionally charged and bizarre dreams.

Some of this sleep-time brain activity, researchers surmised, might serve memory. In the 1970s David Marr, a computational neuroscientist then at Trinity College Cambridge, floated a theory of how the brain integrates new information with existing knowledge. In this model, the hippocampus, a seahorse-shaped structure located in both hemispheres of the brain, stores information during the day. But the memory traces remain fragile until sleep, when they are reinforced and relayed to the brain's cerebral cortex, or outer layer, for long-term storage and integration with other memories.

In a landmark 1994 study, investigators took brain recordings that showed the hippocampus fortifying memories during slow-wave sleep by retracing them. As a rat navigated a maze during its waking hours, patterns of activity among neurons in its hippocampus specified the rat's whereabouts on its trek. While the rat slept, researchers recorded its brain activity again and found the same neural patterns—as if the brain were rehearsing the path through the maze to commit it to memory. A decade later scientists obtained evidence of replay in people by using positron-emission tomography, which detects blood flow as a proxy of neuronal activity. Areas of the brain that became active when people learned routes in a virtual town were reactivated during deep sleep—and the amount of activity correlated with a person's ability to remember the routes.

As Marr had predicted, the replay of memories in the hippocampus is key to consolidation. It seems to flag certain memories for safekeeping, allowing the rest of daily life to fall by the wayside. "You went grocery shopping, and they were out of the little tomatoes ... you don't want to keep that memory for the rest of your life," Stickgold says. "So almost everything gets forgotten. The game of sleep is to figure out what you don't want to forget."

By the early 2000s scientists knew that most of the high-voltage waves of deep sleep originate in the brain's decision-making center, the prefrontal cortex, and move as smoothly and regularly as waves in a calm sea from the front to the back of the brain. And studies in animals and in people with epilepsy (in particular, individuals who had electrodes implanted in their brains to detect seizures) had implicated other sleep-time rhythms in memory processes. These include the ripples of electrical activity from the hippocampus that probably reflect replay—and which coincide with the troughs of sleep spindles originating in the thalamus. When a person is awake, this relay station sends selected information from the senses to the cerebral cortex for interpretation, but when someone is asleep, it shuts most signals out so the person remains generally unaware of their surroundings. Intriguingly, the number of sleep spindles per minute correlates with the person's ability to learn, according to Poe.

In a further, striking coincidence—or more likely not a coincidence at all but something integral to a process of nightly information transfer perfected by evolution—both the ripples and the spindles rise and fall with the slow waves. "There's this three-part symphony," Stickgold says. "The hippocampus and the thalamus and the cortex all work in unison to strengthen specific memories."

Still, the evidence that the sleeping brain analyzes and integrates memory remained circumstantial until experimenters found ways to influence the process. "Can we manipulate the waves?" wondered Jan Born, a behavioral neuroscientist now at the University of

Tübingen. He and his team at the University of Lübeck applied oscillating current through the scalp of sleeping subjects to increase the amplitude of slow waves. The manipulation enhanced memory, they reported in a 2006 publication. But the electrical field seemed to vary unpredictably across the brain's anatomical folds. So the team switched to sound, which would be processed more reliably, Born felt, through a biological channel: the ear.

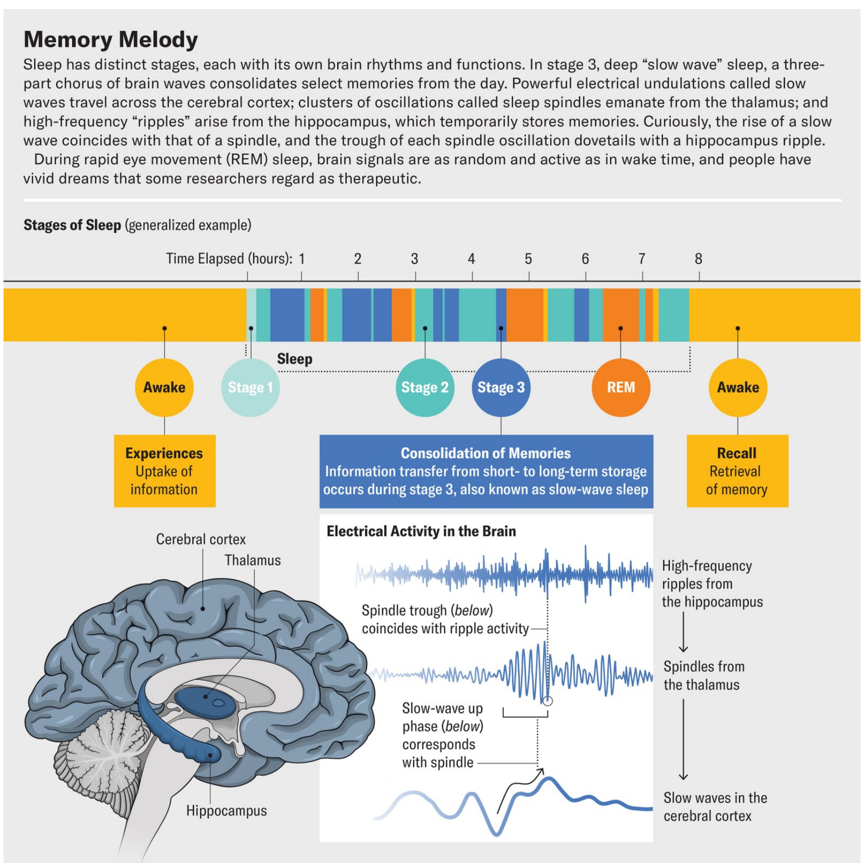
The researchers played soft clicks to sleepers timed to the up phase of their slow waves. The stimulation, given for a single night, greatly enhanced the size and duration of the slow waves and the spindles. Critically, compared with their performance after sleep alone, the intervention improved participants' memory of 120 word pairs, the team reported in 2013. The work directly tied the oscillations of slow-wave sleep to memory—and pointed to a way of using slow-wave sleep to improve memory.

“That's a sleep-engineering idea: Can we make that physiology run its course more effectively? Or, if it's not quite working well, can we adjust it so that it works better?” Paller asks. Slow waves weaken with age, which might explain age-related memory problems. Would supplementing slow waves mitigate memory decline? Northwestern neurologists Roneil Malkani and Phyllis Zee, in collaboration with Paller, among others, successfully used sound to enhance the ability to recall word pairs in five of nine people with mild cognitive impairment.

These interventions lasted just one night, however. In practice, staving off memory decline most likely requires longer-term treatment. Stimulating the brain during sleep through surgically implanted electrodes could theoretically shore up memory on a consistent basis. Neurosurgeon Itzhak Fried of U.C.L.A. Health and his colleagues recently showed that they could use such deep-brain stimulation to enhance memory. Fried had implanted the electrodes to detect seizures in people with severe epilepsy. But when these patients were asleep and seizure-free, he used the electrodes to sense and alter their deep-sleep oscillations.

As a slow wave was on the upswing, one of the electrodes sent a pulse of electricity to boost “the triple coincidence of ripples, spindles and slow waves,” Fried says. All six individuals who received this stimulation in the prefrontal cortex showed better recall of pairs of pictures after the night the electrode was live compared with their memory after undisturbed sleep, the scientists reported in 2023. The degree of memory improvement correlated with the shift in the brain's electrical patterns.

“We are changing the architecture of sleep,” Fried says. “Our goal is to really try to see whether we could have a memory aid or a memory neuroprosthetic device”—akin to a cochlear implant for people with impaired hearing.



Credit: Ni-ka Ford (*brain*) and Jen Christiansen

In addition to improving general memory by enhancing electrical waves in the sleeping brain, scientists have found diverse ways to enhance specific memories but not others. The first attempt at this strategy involved odors. Born's team asked people to sniff a rose scent while they learned the location of objects in a grid. They then exposed some of the participants to the fragrance while they slept. When delivered during slow-wave sleep, the scent spurred the sleepers' brains to revisit what they had learned—and significantly improved their recall of the locations (compared with that by people who were not exposed to the odor during sleep or were exposed to it only during REM sleep), the researchers reported in 2007. Brain imaging revealed that the scent strongly activated the hippocampus, further indicating that the stimulus enhanced replay.

Two years later Paller and his colleagues showed that they could do something similar with sound. The researchers played unique sounds while people memorized the locations of 50 objects on a computer screen. When seeing a picture of a cat, for example, the participants heard a meow; when seeing a kettle, they heard a whistle. The scientists then played 25 of the sounds during a nap, after which people remembered the locations of the associated objects better than they remembered the others—if they heard a whistle and not a meow, they would be more accurate in recalling the kettle's location on the screen than the cat's.

Paller's method, which he termed targeted memory reactivation, or TMR, gained traction as a way to bolster specific memories. In 2022 his then graduate student Nathan Whitmore showed that TMR could improve memory for faces and names, with the strongest effects in those who had the longest and most uninterrupted slow-wave sleep. This method might help older people with memory problems remember facts important to them, such as their grandchildren's names, Paller says.

TMR can also improve procedural memory, which underlies skills ranging from playing a piano piece to perfecting a jump shot.

People execute learned sequences of finger movements faster after sleeping. Performance improves further if the memory for the sequence is reactivated during slow-wave sleep—by, say, a playback of tones the person listened to while learning each finger movement.

A similar method could speed recovery from strokes that leave people unable to perform basic movements. Rehabilitation involves practicing those skills daily. “If you want to use your toothbrush or pick up the salt, you have to control some muscles selectively and not other muscles,” Paller says. To teach these kinds of skills, Northwestern neurologist Mark Slutsky developed a simple 1980s-style video game in which users must activate one or two muscles to move a cursor from the center of a screen to one of eight targets—red squares that turn green when the cursor reaches them—on the perimeter.

In a 2021 study, Paller, Slutsky and their colleagues showed that TMR can improve people's performance in this game. While aiming for each target, 20 healthy young adults heard a unique sound such as a meow, drumroll or bell. After few hours of practice, they took a 90-minute nap. When they entered slow-wave sleep, they heard some of the sounds at five-second intervals. After they awoke, they showed improved performance—in speed, efficiency and muscle selection—in navigating to the red-square targets that were linked to the sounds played during their nap. Paller, Slutsky and their colleagues are testing a similar procedure in stroke patients who have difficulty moving their arms.

Cutting-edge versions of TMR synchronize the sound cues with the slow waves. “It matters exactly when we apply these triggers,” says neuroscientist Penelope Lewis of Cardiff University in Wales. She and her colleagues find that the technique can improve the learning of relations among objects—in this case a hidden ranking in groups of six photographs—but only if the sounds denoting that relation are played back during the peak, and not the trough, of the slow wave. In a related finding, cognitive neuroscientist Bernhard Staresina of the University of Oxford and Hong-Viet V. Ngo, now at the University of Essex in England, reported improved memory for verb-picture associations when they synchronized specific sound cues to the slow wave's rise. Moreover, cueing during this phase prolonged the wave and increased the power of associated spindles.

Intervening in slow-wave sleep can also alter emotions attached to specific memories—which can potentially boost mental health. Cognitive neuroscientist Xiaoqing Hu of the University of Hong Kong and his colleagues used TMR to put a positive spin on aversive memories by building associations with upbeat words. They taught people to associate nonsense words with disturbing photographs and then, during slow-wave sleep, replayed the nonsense cues along with positive words. Afterward people were less repulsed by the cued pictures than they had been before, the researchers reported in 2023. Again, the effect was strongest when the positive words coincided with the up phase of slow oscillations.

The role of slow-wave sleep in memory consolidation is now well established, but the function of REM sleep is less clear. The dreams in this stage often seem illogical because parts of the brain's prefrontal cortex, which controls rational thought, are offline while brain regions controlling vision, movement and emotions remain active. Yet one emerging theory is that the fantastical dreams experienced during REM sleep tame emotions attached to memories and help people gain a broader understanding of what happens to them.

“REM-sleep dreaming offers a form of overnight therapy,” writes neuroscientist Matthew Walker in *Why We Sleep: Unlocking the Power of Sleep and Dreams* (Scribner, 2017). “[It] takes the painful sting out of difficult, even traumatic emotional episodes.” During REM sleep, levels of norepinephrine—a neurotransmitter that drives fear responses such as sweating, rapid heart rate and pupil dilation—get tamped down. As a result, memories that surface during REM sleep are divorced from those responses, Walker and others say, decoupling them from their emotional charge. (In patients with post-traumatic stress disorder, however, levels of norepinephrine remain high, and nightmares recur.)

If the theory is correct, inducing people to relive difficult experiences during REM sleep might help defuse the disturbing emotions associated with them. In a 2021 study, people rated upsetting pictures as less bothersome after associating the pictures with specific sounds and being exposed to those sounds during REM sleep. In contrast, there was no effect when the sounds were played during slow-wave sleep. If something similar works on people's real-life memories, it might be an avenue for treating depression or PTSD, according to Lewis.

REM sleep dreams might also help defuse strong emotions attached to an event through subconscious learning. Instead of dreaming about the upsetting event itself, people often dream about a more benign, related memory, leading them to subliminally connect the two experiences. Stickgold offered an example: if he were distraught after having a near-miss car accident during the day, he might dream about playing bumper cars with his son. The dream would help Stickgold realize that the car crash, if it had actually happened, “might have just meant my fender got bashed in. [But] I was reacting to it as if I had just barely stayed alive,” Stickgold speculates. “And that might have become clear to me because I had this linked memory of bumper cars where nothing bad happens.”

Credit: Blend Images/Inti St Clair/Getty Images (*child*); sarayut Thaneerat/Getty Images (*sunset*)

In this way, REM sleep dreams can provide perspective. “You have to let the brain build this dream narrative to evaluate the emotional response to it,” Stickgold says. TMR could be used to shape that narrative, and the nightmare-disorder study in Geneva highlighted the possibility of such interventions. It could also make traditional forms of psychotherapy more effective. “Any psychotherapeutic approach aims at a change in behavior, habits, thoughts. Psychotherapy is therefore a form of learning,” says neuroscientist Sophie Schwartz of the University of Geneva, first author of the nightmare-disorder study. “Using TMR, we can boost such learning.”

Most sleep-engineering studies require patients or volunteers to come into a laboratory or other institutional setting, which limits the scope and efficacy of the intervention. People don't want to sleep in a lab for more than a night or two. But “if the technology were wearable and portable, it could plausibly be embedded in somebody's life,” says Heidi Johansen-Berg, a cognitive neuroscientist at Oxford. “So even if the benefit of any single day is quite small, you could imagine those incremental benefits building up significantly over time.”

Commercial devices that can be used at home are likely to be an important gateway to enhanced healing during sleep. One such invention, currently being tested for its ability to burnish verbal memory and to speed stroke recovery, involves a smartwatch that collects movement and data on heart rate, as well as a smartphone that plays sounds. A machine-learning model identifies periods of deep sleep and triggers TMR sounds within these periods. In research published in 2022, Whitmore and others found that using this technology at home for three nights improved people's memory for object locations—as long as the sounds were played softly enough that they did not disturb the sleeper.

For debilitating nightmares, doctors can already prescribe a phone app that uses artificial intelligence to analyze biometric data from Apple Watch sensors. When the sensors detect the rising heart rate and restlessness associated with a nightmare, the watch delivers intermittent gentle vibrations to disrupt the dream without waking the sleeper. Data published in 2023 from a trial of 65 veterans with trauma-induced nightmares suggest the device, when worn at least half the time, significantly enhanced sleep quality, as reported by the veterans.

A glovelike sleep detector developed by Adam Haar Horowitz, then at the Massachusetts Institute of Technology, and his colleagues might also reduce nightmares. The device monitors biological signs of sleep onset through contacts on the wrist and hand. It also connects to an app that gives voice prompts such as “tree” that, in a recent study, made nappers dream about trees and enhanced their creativity on tasks related to trees.

Despite the promise of sleep engineering, experts warn of risks inherent in tampering with memories. “You are biasing which ones are preferentially strengthened in the brain,” Lewis says. If you start doing it every night, who knows what kinds of imbalances that might cause?” It is also possible that these interventions could disrupt sleep. In another of Whitmore and Paller's experiments, for example, when the sounds were played too loudly, memory actually worsened. “There are lots of things still to understand about this before we would be ready to recommend it to the general population,” Lewis says.

Meanwhile the experiments have deepened scientists' understanding of sleep's role in memory and emotion—and how it shapes people's outlook on the world and themselves. “That is what the night is for,” Stickgold says. “It's to take all the information that came during the day and integrate it with all the information we already have in a way that helps you build that story of how the world works and what your life means.”

For Sonia, at least, the targeted memory reactivation has ended her nighttime siege in the forest. Instead one night she dreamed of being invited to a party in a chalet. “There was a terrace which gave a view of the mountains,” she wrote in her dream diary. “We all went out to watch the sunset. The sky was dark pink, the weather was very beautiful. All of a sudden, I feel a hand on my waist ... This person took my hand and took me to the center of the terrace, we started dancing without music. It was like in the movies, the world around began to spin quickly, I felt butterflies in my stomach for the first time in my life.”

Ingrid Wickelgren is a freelance science journalist based in New Jersey.

Inside a 3D-Printed Universe

Three-dimensional printouts of stellar nurseries are helping to reveal how stars are born

By Nia Imara

Astrophysicist Nia Imara holds a 3-D-printed globe that represents a molecular cloud where stars are born.
Stephanie Mei-Ling

Is it a flying elephant? A gingerbread man? When I was little, I used to search the clouds for amusing shapes as they drifted across the sky and imagine stories about their patterns. Now I'm a professional stargazer, and things haven't changed much. These days I search for patterns in molecular clouds, the birthplaces of stars. The shapes I find in these stellar nurseries do more than stimulate my imagination—they also tell a very real story about when, where and how stars are born. For astronomers, understanding this story depends on our ability to identify and interpret the intricate forms we see in the clouds.

Observations reveal elaborate networks of material, including compact clumps of gas and long, skinny, noodlelike structures called filaments woven throughout. Far from being uniform and smooth like milk, molecular clouds are lumpy, more like chicken noodle soup. The gas and dust accumulate into a range of physical scales and are organized into increasingly dense formations. Their structure is hierarchical, like Russian nesting dolls, with smaller shapes enclosed within larger ones. Filaments are much denser than the diffuse gas that fills most of the volume of a cloud. And embedded within filaments are even smaller, denser knots of gas we call cores. These cores represent the final stage before a star is born.

The dynamics of molecular clouds are as complicated as their spatial structure. Stars, planets, and galaxies such as the Milky Way all spin around their axes in a fairly predictable manner. But the space between the stars—the interstellar medium, where molecular clouds reside—is a wild, chaotic frontier. The motions inside clouds are turbulent, with globs and eddies of gas swirling around like capricious fairies. Observations of both the dynamics and the spatial architecture of molecular clouds have enabled astronomers to paint a compelling, if incomplete, picture of how stars are born.

A major reason our understanding is limited is that, although clouds are three-dimensional, our telescope images are flat. We often can't decipher the real shape of a structure within a cloud, because we are seeing it projected onto a flat plane. Intrigued by this problem, I've been inspired to look beyond astronomy for solutions.

In addition to being a scientist, I'm an artist—a painter. This part of me understands that as good as technology can be at recognizing patterns, there are no substitutes for the human eye, brain and imagination. I had the idea to use 3-D printing to create tangible reproductions of molecular clouds that let us peer into the multiple dimensions of these objects. Being able to see and hold mini molecular clouds, I thought, might unlock ways of viewing and thinking about these mysterious regions.

Star birth takes place in the cold and darkness of space. At hundreds of degrees below zero, molecular clouds are among the most frigid regions of the universe. They are composed primarily of hydrogen molecules (two hydrogen atoms bound together) but also contain a significant amount of helium and trace amounts of carbon monoxide and other molecules, as well as a sprinkling of stardust (particles composed of heavy elements created by previous generations of stars). These simple ingredients, together with the freezing temperatures, turn out to be perfect for making stars and planets. Because they are so cold, molecular clouds are virtually invisible in the optical light our eyes can see. Their impressive architecture is best seen in the infrared and radio spectra.

Our observations through infrared and radio telescopes show that long before a star comes into being, a vast cloud of gas dozens of light-years across assembles and evolves under the mutually interacting influences of gravity, turbulence, radiation and magnetic fields. Some studies suggest that a molecular cloud forms when even bigger clouds of atomic hydrogen (single hydrogen atoms) smash into one another. This scenario seems reasonable, given that atomic hydrogen is needed to make molecular hydrogen. Meanwhile dust grains help to dissipate heat from the cloud.

Once enough atomic hydrogen gas has accumulated and cooled down, the inner part of the cloud becomes mostly molecular. At this point, the molecular cloud may have hundreds of thousands to millions of times the mass of the sun. That places stellar nurseries among the largest and most massive entities in galaxies.

The tumultuous motions and magnetic fields inherited by the molecular cloud from its surroundings both play important roles in shaping its structure. Over the course of millions of years, pockets of gas within the cloud collide, merge and grow in density. Internal turbulence causes the gas to become compressed, which quickly leads to the formation of filaments and then cores. Some of the cores continue sucking in mass from their surroundings, like cosmic vacuum cleaners. As the core grows denser, the internal pull of gravity becomes stronger, and the core begins to collapse. Meanwhile the temperature at the center of the collapsing region gets hotter and hotter. The densest cores eventually succumb to the overwhelming force of gravity and initiate nuclear fusion, at which point a star is born!

The Orion Nebula is an active star-forming region that's part of the much larger Orion A molecular cloud. At only 1,400 light-years away, it is the closest stellar nursery where high-mass stars are being built. As you read this, hundreds of new stars are in the process of being born there. Molecular clouds such as Orion A may produce hundreds of thousands, if not millions, of stars over their lifetimes. As it turns out, the star-formation process is very inefficient, and most of a stellar nursery's mass does not wind up in stars, which are tiny in comparison. Imagine it like this: if our sun were the size of a blueberry, its parent molecular cloud might have been the size of Earth or even Jupiter.

This is the big picture of star formation that astronomers have figured out, but there are several key steps in this process that elude us, largely because of the sheer difficulty of observing the literally nebulous structure of stellar nurseries. One of the most conspicuous missing puzzle pieces is how exactly star formation depends on the structures within molecular clouds. For instance, how do filaments and cores determine how big the newborn stars will be? This is a critical question because a star's mass is the single most important factor in its subsequent evolution. Does a filament act as a kind of umbilical cord through which cores and then stars acquire their masses?

During my postdoctoral work, I studied the California molecular cloud, named for its resemblance to the state of California. My collaborators and I explored a small subregion that I dubbed Cal X because of the appearance of two intersecting filaments at that location. While investigating infrared images from the Herschel Space Observatory, we noticed that a number of cores were embedded in each of the two filaments, but none of them showed any indication of becoming stars. Lodged within the junction of Cal X, however, was the most massive core in that region. That core was in the process of delivering at least two baby stars.

Imara, who is a painter as well as a scientist, uses art to inspire and inform her research.
Credit: Stephanie Mei-Ling

When I analyzed what was going on in Cal X, I discovered what appeared to be flows of gas along the filaments, as though they were funneling material to the gargantuan core. As suggestive as the evidence was, however, I couldn't entirely rule out other possibilities. Perhaps gas was flowing *away* from the filaments, or maybe they were rotating, or possibly some combination of all these things was happening.

My hunch is that the filaments of Cal X are indeed serving as cosmic umbilical cords to the stars being formed in the region. Studies of other molecular clouds, as well as computer simulations, have shown similar patterns in filaments and provide compelling evidence for this scenario. But one of the main reasons that it is so challenging to draw a definitive conclusion is that our observations typically can't show the 3-D geometry of stellar nurseries. To say conclusively what is happening in the California molecular cloud, we would need to know how the filaments are positioned with respect to one another and to the rest of the cloud. But in a flat image, it is impossible to tell whether they are tilted toward or away from us or perhaps slant in opposite directions. It's like trying to tell which way a river is flowing when all you have is a bird's-eye view of the landscape—and no way to distinguish between mountains and valleys.

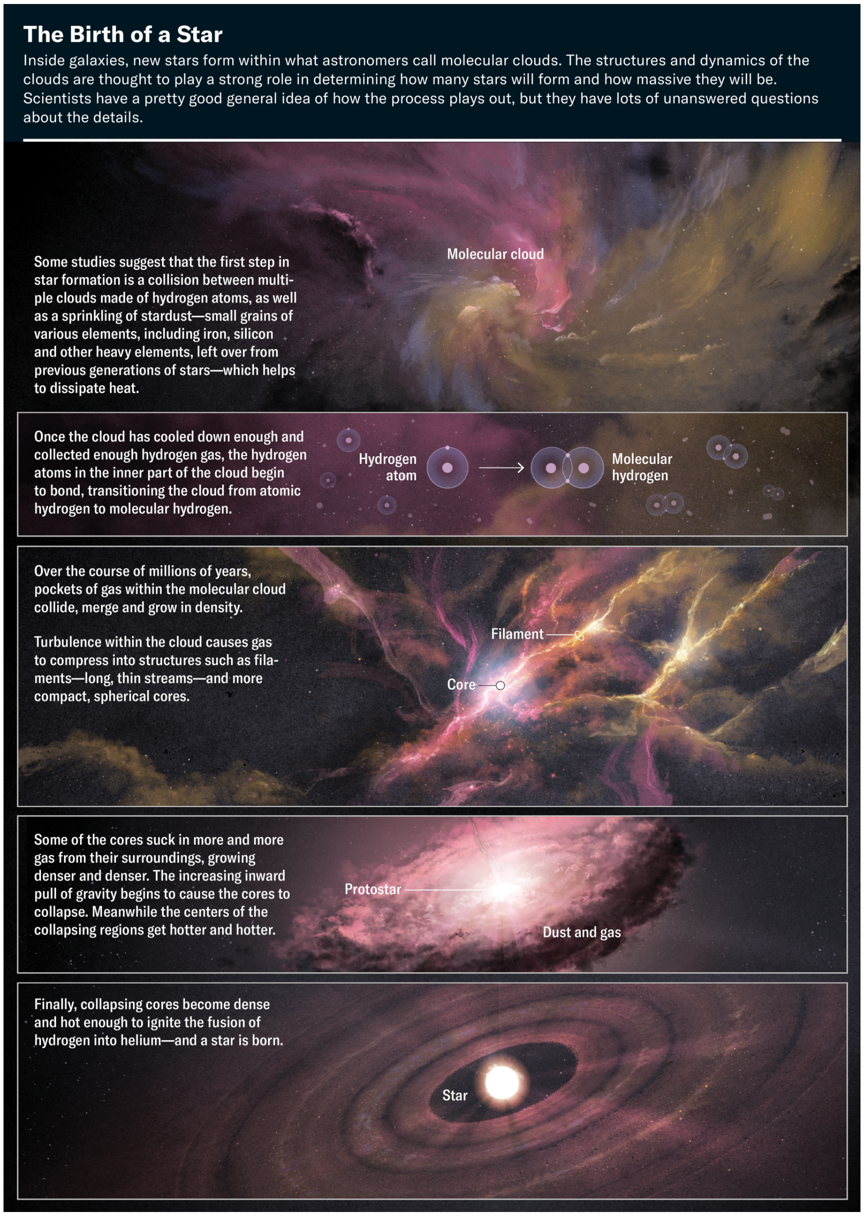
A connected question about the relation of molecular cloud structure to star formation is, What sets the rate at which stars are born? The Milky Way produces stars at a leisurely pace of about three solar masses' worth of stars every year. But so-called starburst galaxies that flourished in the early universe have outlandishly high star-formation rates that are tens or even thousands of times that of our galaxy. Could it be that stellar nurseries in starbursts have a fundamentally different architecture than those in normal galaxies?

In the past decade these questions have come to the fore as images of the interstellar medium taken with Herschel, as well as with the Atacama Large Millimeter Array (ALMA) in Chile and other telescopes, have highlighted how significant cloud substructure might be in star formation. Within molecular clouds throughout the Milky Way and other galaxies, we see complex networks of filaments at a range of size scales from a few to hundreds of light-years long. And within filaments, the densest cores seem to be the preferred birth sites for stars. In spite of the challenges of interpreting our observations, it's clear that understanding the origin and evolution of dense gas in molecular clouds may be the key to making progress toward a fuller theory of how stars come to be.

When studying molecular clouds, I'm often reminded of lyrics from a song in my favorite movie, *The Sound of Music:* “How do you catch a cloud and pin it down?” Since my graduate school days, I've been preoccupied with the idea of trying to “catch” stellar nurseries. I've looked into various algorithms created to identify molecular clouds and quantify their substructure. But it can be tough to interpret the results of algorithms that are designed to identify 3-D structures from 2-D images. How do we draw a meaningful boundary around a star-forming core swimming in an ocean of dust and gas? Unrelated material in front of or behind the core could be tainting our view. Or, if we're trying to quantify the properties of overlapping filaments, how can we tell where one ends and another begins in the tangle? Could it be that our perspective sometimes leads us to confuse certain structures for something else?

I had the idea to use 3-D printing to visualize structure in stellar nurseries. I wanted to be able to hold the stars in my hand. Unlike some other methods of visualization, 3-D printing represents astrophysical structures in a way that taps into the human brain's ability to recognize patterns. Moreover, interactive 3-D structures can engage our intuition in ways that 2-D representations can't. I began collaborating with John Forbes of the University of Canterbury in New Zealand and James C. Weaver of Harvard University's John A. Paulson School of Engineering and Applied

Sciences. We became the first research group to use 3-D printing to visualize star formation.



Credit: Matthew Twombly

To start, we ran several simulations representing various physical extremes. One simulation had very strong gravity; another had weaker magnetic fields than we usually observe in real clouds. The point was to isolate various aspects of physics to see how they drive the evolution of molecular clouds in different ways. We used the simulations, rather than observations of real clouds, as source data for the 3-D print designs because simulations can be run in three dimensions. When we simulate stellar nurseries, it is as if we are omniscient demigods because at any moment we know everything that's happening at each location in the simulation. Our knowledge is limited, of course, by the parameters we put into the simulation, but these inputs are well informed by observations. We tested the resulting models to make sure they met our standards for resembling real molecular clouds. Then we postprocessed the simulation data, putting them in a format that could be understood by our 3-D printer, which prints in very thin sheets of resin. It layered more than 2,500 sheets on top of one another to build a sphere.

When I finally held one of my stellar nurseries for the first time, I was captivated. I turned the softball-size globe around in my hand, examining its twisting structures from all angles. I could see filaments snaking through the cloud and dissolving into the background. I could see cores, wispy puffs, planar structures and forms I had no names for. My colleagues and I also printed half-spheres so we could better see what was going on deep inside the clouds, and I was surprised by how dramatically the structure started to change just below the surface. In observations of real stellar nurseries, much of this material is projected onto the plane of the image, so there's no way to tell what's in front and what's behind. Now, holding a stellar nursery with my fingertips, I could see what was going on with a simple twist of the wrist. It was beautiful.

One big surprise was that the shapes of structures within molecular clouds are even more complex than we thought. As my team and I suspected, sometimes what appeared to be a filament from one angle was a flat, sheetlike structure in projection. In other words, a filament might be a pancake viewed along its edge. But we also noticed filaments embedded in pancakes, which raises the tantalizing possibility that filaments emerge from sheets.

I think of our 3-D prints as interactive maps. They show us where to look to identify the structures that play key roles in star formation. More important, they help us cultivate our ability to see things from a new perspective so we can look at observations of real clouds with fresh eyes and potentially discover patterns we hadn't noticed before.

Years before I thought about using 3-D printing as a visualization tool for stellar nurseries, I drew a sketch of myself holding a star in my hand. And years before that, as a graduate student writing my dissertation, I imagined myself flying through molecular clouds, compressing millions of years of their evolution into a few minutes. I'm not sure I would have come up with the idea of using computers to create sculptures of stellar nurseries had I not been an artist.

Stellar nurseries are among the most complex (and, in my opinion, the most beautiful) objects in the cosmos. In recent years excitement about deducing their 3-D structure has increased in our field as advances in the quality and variety of observations have made it possible to explore their architecture in new ways.

Using data from the Gaia space observatory, for instance, researchers have created 3-D maps of the dust associated with molecular clouds near the sun. One study compared two of my favorite clouds, Orion A and California. These two stellar nurseries are an interesting case study because they lie at roughly the same distance from us; they have comparable masses, each containing about 100,000 times the mass of the sun in molecular hydrogen; and in 2-D images, they have similar oblong shapes. California is slightly more massive, but curiously, it produces stars at a rate nearly 100 times slower than Orion A's. Why?

According to the study, it turns out that whereas Orion A is a relatively compact cloud shaped like a big cigar, California is a more flattened, extended structure—like the “pancakes” in my 3-D printouts. But because of its orientation in space, we see it from the side, and in flat images, it appears more compact than it really is.

Astronomers have known for decades that star formation tends to happen faster in denser gas. The difference in the 3-D shapes of California and Orion A might explain their disparate star-formation rates. The shapes of clouds and, ultimately, star formation are influenced by how gas flows within them. Going forward, my colleagues and I are incorporating colors into our 3-D prints to explore the motions of structures within stellar nurseries.

A new generation of telescopes, including the James Webb Space Telescope, ALMA, and other observatories, is collecting data across the electromagnetic spectrum and improving our quantity, quality and variety of star-formation observations. With advances in numerical simulations keeping pace, both theorists and observers are sprinting to develop ways to solve the mysteries of star birth. The artist in me is convinced, however, that our most important tool remains our imagination. Just like when we were children lying on the grass and watching the clouds pass overhead, our imagination can see things that the rest of our mind can't and may lead the way to the discoveries we hope for.

Editor's Note (4/5/24): This article was edited after posting to correct the descriptions of the composition of molecular clouds and the fusion of hydrogen into helium in stellar birth.

Nia Imara is an artist and an astrophysicist at the University of California, Santa Cruz, where she studies how stars form in the Milky Way and other galaxies.

These Cancers Were Beyond Treatment—But Might Not Be Anymore

New drugs called antibody-drug conjugates help patients with cancers that used to be beyond treatment

By Jyoti Madhusoodanan

Keith Negley

In the long and often dispiriting quest to cure cancer, the 1998 approval of the drug Herceptin was a tremendously hopeful moment. This drug for breast cancer was the first to use a tumor-specific protein as a homing beacon to find and kill cancer cells. And it worked. Herceptin has benefited nearly three million people since that time, dramatically increasing the 10-year survival rate—and the cancer-free rate—for what was once one of the worst medical diagnoses. “Honestly, it was sort of earth-shattering,” says oncologist Sara M. Tolaney of the Dana-Farber Cancer Institute in Boston.

But the drug has a major limitation. Herceptin's beacon is a protein called HER2, and it works best for people whose tumors are spurred to grow by the HER2 signal—yet that's only about one fifth of breast cancer patients. For the other 80 percent of the approximately 250,000 people diagnosed with the disease every year in the U.S., Herceptin offers no benefits.

The hunt for better treatments led researchers to reimagine targeted therapies. By 2022 they had developed one that linked Herceptin to another cancer-killing drug. This therapy, for the first time, could damage tumors that had vanishingly low levels of HER2. The drug, named Enhertu, extended the lives of people with breast cancer by several months, sometimes longer. And it did so with fewer severe side effects than standard chemotherapies. The U.S. Food and Drug Administration approved its use in that year.

The news got even better in 2023. Researchers reported that Enhertu appeared to work even on tumors with seemingly no HER2 at all. (It's possible the cancers did have the protein but at very low levels that escaped standard detection methods.) “Exciting!” says oncologist Shanu Modi of Memorial Sloan Kettering (MSK) Cancer Center in New York City, who helped to run the study that led to Enhertu's approval. “They did this provocative test and saw this almost 30 percent response rate” in tumors apparently lacking the cancer protein, she notes.

Enhertu belongs to an ingenious and growing class of targeted cancer drugs called antibody-drug conjugates, or ADCs. The compounds are built around a particular antibody, an immune system protein that homes in on molecules that are abundant on cancer cells. The antibody is linked to a toxic payload, a drug that kills those cells. An ADC's affinity for cancer means it spares healthy cells, avoiding many of the side effects of traditional chemotherapy. And each antibody can be paired with several different drugs. This Lego-like assembly opens up a world of mix-and-match possibilities. Researchers can use the same drug to treat many cancers by switching up the antibody, or they can attack one type of tumor with many different ADCs that target several cancer biomarkers on the cells. This ability “changes the way we think about drug development,” Tolaney says.

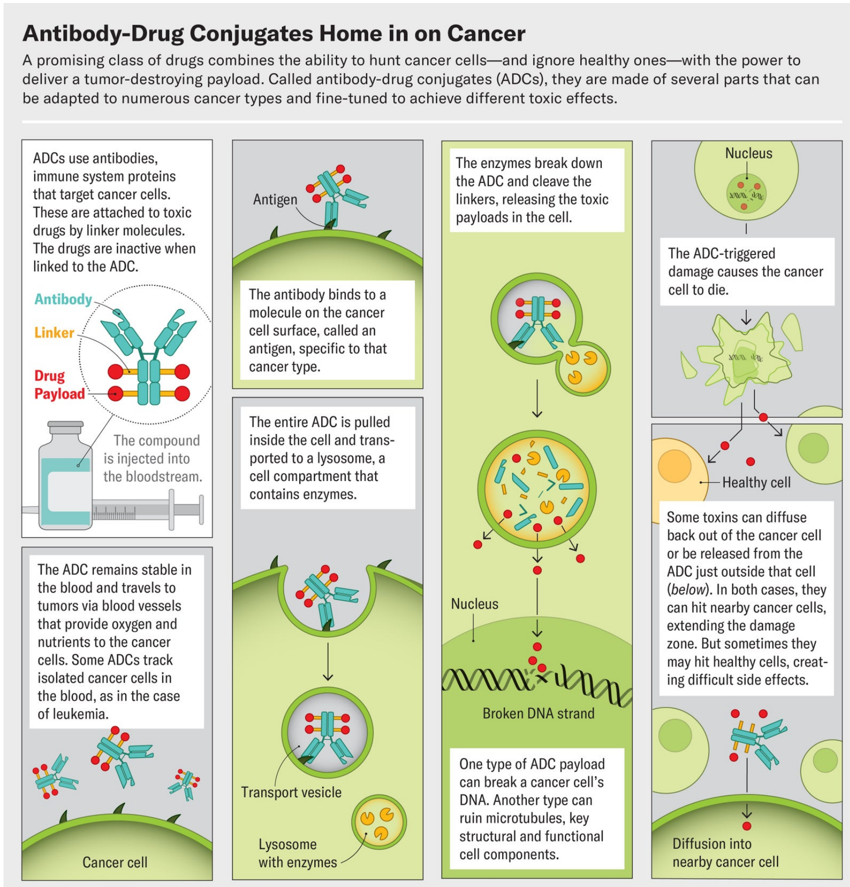
The idea for ADCs is not entirely new—the first one was cleared for patient use in 2000—but recently scientists have learned intricate chemical construction techniques that make the compounds much more effective, and they have identified new cancer-specific targets. These advances have driven a wave of new development. Fourteen ADCs have been approved for breast, bladder, ovarian, blood, and other cancers. Approximately 100 others are in the preclinical pipeline. One ADC for breast cancer, known as T-DM1, proved much more effective than Herceptin and has now become the standard of care for early stages of disease. “It is pretty cool to see how things have changed so quickly,” Tolaney says. Buoyed by the successes, researchers and pharmaceutical companies are pouring resources into developing more powerful ADCs—perhaps even ones that can work across a wide range of cancer types. Pharma giants such as Gilead, Roche and BioNTech have invested heavily in their ADC programs; in October 2023, for example, Merck put $4 billion into a partnership with Daiichi Sankyo, the biotechnology firm that partnered with AstraZeneca to produce Enhertu.

But the new drugs are still beset by some mysterious problems. Some ADCs have side effects similar to those caused by traditional chemotherapies—which shouldn't happen, because the drugs are supposed to target cancer cells alone. On patient forums, people describe needing to reduce their doses because of intolerable nausea or fatigue. These drawbacks limit ADCs' use, so scientists and pharma companies are urgently trying to figure out what is causing them.

In the clinical trial that led to Enhertu's approval, patients typically had already received different kinds of chemotherapy drugs, such as medications that stop cells from multiplying. But these drugs—and other forms of chemotherapy—do not distinguish between a cancer cell and a healthy one. Any cell trying to make DNA or multiply is vulnerable, and normal tissue as well as tumors can be attacked. Fully 64 percent of people on standard chemotherapy experience nausea, diarrhea, fatigue, and other negative side effects. For many, these can be as debilitating as cancer itself. Such effects limit the dose people can take and the length of treatment, leaving windows of opportunity for tumors to grow resistant and rebound.

For many years researchers have sought less toxic alternatives, envisioning precision drugs that target cancers and spare healthy cells. The idea of ADCs sprang from the exquisite specificity of antibodies. If highly toxic forms of chemotherapy could be strapped onto antibodies, the toxins would reach only the cancer cells and no others. Although the concept was straightforward, attempts at making ADCs faltered for decades.

Some of the earliest attempts used drugs that just weren't strong enough. In the 1950s, for instance, researchers linked a drug named methotrexate to an antibody that targets carcinoembryonic antigen, a common tumor marker, and tested whether the construct could treat advanced colorectal and ovarian cancers in people. The drug bound to its target but had little therapeutic effect. Researchers then swung too far to the other end of the spectrum and tried using much more toxic drugs instead. But these drugs triggered serious side effects.



Credit: Jen Christiansen; Graphics consultant: Greg Thurber/University of Michigan

Greg Thurber, a chemical engineer at the University of Michigan, looked into this conundrum. He began working on ADCs when studying how antibodies spread through the body to bind to their targets. After ADCs infiltrate a tumor through its network of blood vessels, the compounds slip out of these vessels and into cancer cells to kill them, Thurber says. But the ADCs that existed at the time never got past the cells just outside the blood vessels. They bound too tightly. The key to improved effects, it turned out, was tailoring the antibody parts so they zeroed in on cancer cells but had a loose enough grip for some to slip into the interior of the tumor. “A lot of people in the field had a very simple concept—we put a chemotherapy drug on an antibody, it targets it to the cancer cell, and it will avoid healthy tissue," Thurber says. "That's not at all how they work in reality."

Tinkering with the drug component of ADCs, as well as the antibody, eventually led to a cancer-killing sweet spot. In 2013 the fda greenlit T-DM1 for breast cancer. Its antibody is trastuzumab (the "T" in T-DM1), the same antibody used in Herceptin. The drug attached to this antibody is notable because it's too dangerous to be used on its own. Known as emtansine, it was initially discovered in the 1970s but shelved because it was too toxic to too many cells. Tethered together as T-DM1, however, the drug and antibody generally stayed away from healthy cells and proved to be a potent and precise combination.

In the early 2000s Modi helped to conduct a trial of T-DM1—branded Kadcyla by its maker, Genentech—in people who had an especially difficult disease: advanced HER2-positive breast cancer that had spread throughout the body. Only those who had run out of other treatment options were enrolled. "We were taking people who in some cases were really looking to go to hospice," Modi says. Yet "almost every patient who was enrolled on that drug had benefits. It was really so satisfying."

In another trial of about 1,500 people with early breast cancer, an interim data analysis, published in 2019, estimated that 88 percent of those who received T-DM1 would be cancer-free three years later, compared with just 77 percent of those who received Herceptin alone. The drug has proved "more active than most of the therapies we were giving to patients, and it was associated with a better safety profile," Modi says.

Kadcyla's success against difficult-to-treat cancers didn't just transform some patients' lives. It pumped enthusiasm—and, perhaps more important, pharmaceutical industry dollars—into the idea of ADCs. Researchers now knew that when pieced together correctly, it was possible to load an antibody with drugs too toxic to be used otherwise and still produce a medicine that worked better than traditional chemotherapy.

Several similarly designed ADCs have been approved for a range of different cancer types. Many of these carry drugs that inhibit the enzyme topoisomerase 1, which is essential for DNA replication. Like emtansine, the drug used in Kadcyla, newer topoisomerase inhibitors are too toxic to be used as freestanding drugs but are much less harmful when they're largely restricted to tumor cells. And Kadcyla itself, after being shown to slow or stall late-stage breast cancer, is being tested on patients with very early-stage disease to see whether treatment at that point can not only slow cancer down but actually cure it. Its success “was sort of the catalyst for continued exploration,” Modi says. “Can we build on this? Can we do even better?”

Doing better, it turns out, involves designing good linker molecules that tie the antibody to the drug. These tiny structures act like chemical triggers. They must remain perfectly stable until they reach their target, then unclip from the antibody to discharge their payload at the tumor. Some of the earliest attempts at making ADCs failed not because of their antibodies or drugs but as a result of unstable linkers.

Modern ADCs rely on two types of linkers. One kind remains unbroken even when the ADC reaches its target. The other kind, known as cleavable linkers, are chemicals that break in response to very specific cues, such as enzymes that are abundant in tumors, in the spaces between individual cancer cells. Once an ADC is within the tumor's boundaries, these enzymes cleave the linker and release the drug payload.

Cleavable linkers are showing impressive advantages, and more than 80 percent of currently approved ADCs now use them. An ADC with a noncleavable linker will kill only the cell it attaches to, but one that splits up could place drug molecules near neighboring tumor cells and destroy them as well. This so-called bystander effect can make the drugs much more effective, Thurber says.

Enhertu, for instance, uses the same antibody as Kadcyla but with a cleavable linker (Kadcyla uses a noncleavable version) and a different drug. Each Enhertu antibody carries approximately eight drug molecules, compared with about three per antibody in Kadcyla. In one recent study, researchers compared the effects of these two drugs in people with HER2-positive breast cancers. Enhertu was the clear winner. It stopped tumor growth for more than two years on average, whereas Kadcyla did so for just six months. “It was a landslide in terms of how much better it was,” Tolaney says. “It's a really nice example of how ADC technology leads to dramatic differences in outcomes.”

The bystander effect also explains, in part, why Enhertu is effective against tumors that have barely any HER2: once the ADC enters a tumor and the drug molecules detach, they can kill neighboring tumor cells even if those bystanders don't carry much HER2 on their surface. This action, along with the use of a diagnostic test that can miss extremely low HER2 levels, could explain the results from the trial where the drug seemed to work on tumors with no HER2. That trial employed an assay known as an IHC test. It is generally used to categorize cancers as HER2 positive or negative, not to measure the amount of the protein present. A negative result typically means 10 percent or fewer of the tumor's cells have HER2 on their surfaces. Yet 10 percent may be enough to attract a few Enhertu particles, and the bystander effect might be sufficient to destroy tumor cells, Modi says.

Enhertu is not the only ADC that appears to work this way. In a 2022 study, researchers found that Trodelvy, an ADC that targets a surface protein known as TROP2, seemed to be more effective than standard chemotherapy for people with metastatic triple-negative breast cancer, a particularly hard-to-treat disease. Trodelvy was better irrespective of how much or how little TROP2 was detected on tumors. “That, to me, is wild,” Tolaney says. “We're excited about it because these cancers are having benefits [apparently] without the target.”

This new generation of ADCs is making a difference in other types of cancers previously thought to be intractable, such as metastatic bladder cancer. In 2021 the fda approved Trodelvy and another ADC named Padcev to treat this illness. For 30 years the standard of care for this type of bladder cancer was chemotherapy alone, says oncologist David J. Benjamin, who treats genitourinary cancers at Hoag Family Cancer Institute in southern California. “Now we have multiple new treatments, and two of them happen to be antibody-drug conjugates,” Benjamin says. In clinical trials for patients with advanced bladder cancer, Padcev combined with a drug that stimulates the immune system shrank tumors or stalled their growth in more than 60 percent of people. In a whopping 30 percent of those who received the two-drug combination, their cancer completely disappeared—an unprecedented success.

But even newer ADCs aren't without problems. The bystander effect, which makes them so effective, can spread far enough from the tumor to affect healthy cells, causing hair loss, nausea, diarrhea, fatigue, and other side effects that are disturbingly similar to the fallout of old-school chemo. ADCs also have been linked to a variety of eye problems ranging from conjunctivitis to severe vision loss.

Another explanation for these nasty effects is that there are no protein targets that are exclusive to cancer cells. These proteins, also known as antigens, are more abundant in cancers but may appear in normal cells. That makes some binding of ADCs to healthy cells unavoidable. “I can't think of any examples of true tumor-specific antigens,” says Matthew Vander Heiden, a molecular biologist at the Koch Institute at the Massachusetts Institute of Techonology. Further, ADCs, like any other medicine or antibody, are eventually ingested and metabolized by noncancerous cells. This process fragments them into smaller pieces, releasing payload drugs from their linkers and triggering reactions.

Still, the ability to take ADCs apart and tweak their components—something that isn't possible with traditional treatments—offers researchers the chance to find versions with fewer side effects and more advantages. At present, most ADCs are used at the maximum dose a person can tolerate. That might not be true with future versions. When developing a medication, whether it's a simple painkiller, a chemotherapy or an ADC, researchers begin by figuring out the lowest dose at which the drug is effective. Then they work out the highest dose that people can receive safely. The space between those two doses, known as a therapeutic window, is usually small. But the ability to swap components offers ADC researchers many routes to widening it. Eventually drugmakers might create ADCs so effective that patients never need to take the highest tolerable dose—a much lower one would eliminate tumors without creating unintended consequences such as nausea or hair loss.

Shifting away from toxic chemotherapy-based drugs as payloads could also reduce side effects. Some recently approved ADCs, for instance, link antibodies to drugs that can activate the body's own immune system to attack cancer cells rather than relying on cell-poisoning chemicals. In addition, scientists are exploring ways to deliver radiation therapy directly to tumors by tethering antibodies to radioisotopes. Joshua Z. Drago, an oncologist at MSK Cancer Center, says that with the right kind of linkers, ADCs “could theoretically deliver any kind of small-molecule medication.”

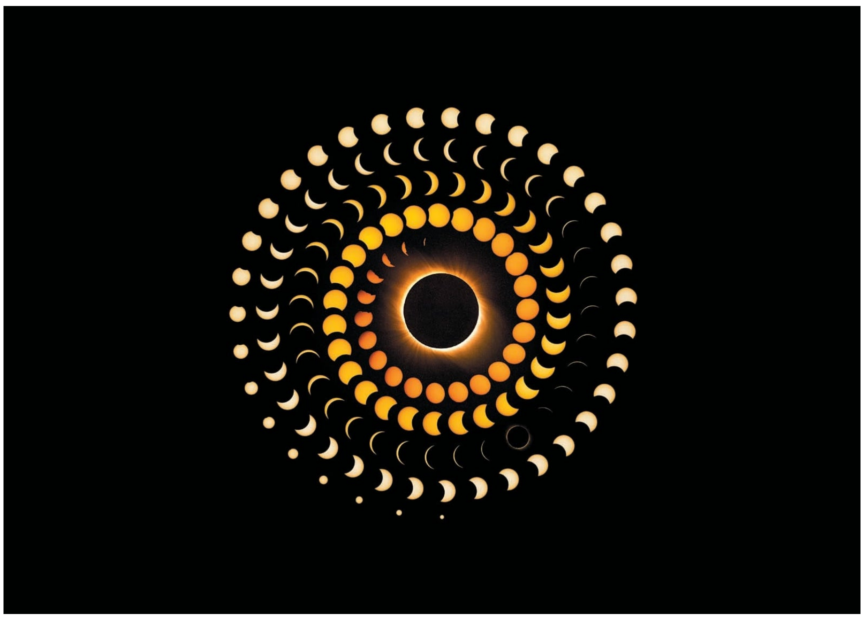
Ultimately, recombined and improved components could lead to the type of swap that cancer patients really care about: exchanging their disease for a cure.

Jyoti Madhusoodanan is a science journalist based in Portland, Ore. She covers health, medicine and the life sciences.

How the Eclipse Will Change Solar Science Forever

The upcoming total solar eclipse and a pair of new sun probes are revolutionizing scientists' understanding of our closest star

By Rebecca Boyle



A composite of 135 photographs taken between sunrise and sunset shows the progression of a total solar eclipse seen from Chile in 2019.
Dan Marker-Moore

This article is part of a special report on the total solar eclipse that will be visible from parts of the U.S., Mexico and Canada on April 8, 2024.

On April 8, 2024, a 115-mile-wide strip of North America will be plunged into darkness. The disk of the moon will slip in front of the sun, obscuring its face and creating a rosy, fluffy crown of flame visible from Mazatlán, Mexico, to Newfoundland, Canada. It will be the last spectacle of its kind for a generation—the next total solar eclipse viewable from across North America will be on August 23, 2044.

Spectators aren't the only ones excited. A solar eclipse is one of the best ways for scientists to study the solar corona, that ring of fire that stands out when the moon blocks our bright star. This feature remains one of the most mysterious parts of the sun. Astronomers originally thought the corona was a feature of the moon—perhaps sunlight reflecting off the lunar atmosphere. But the moon has no atmosphere. It was not until 1806 that Spanish astronomer José Joaquín de Ferrer recognized it was a feature of the sun instead, giving it the name *corona*, the Spanish word for "crown."
[**Related:** Here Are the Best Places to View the 2024 Total Solar Eclipse]

We now know that the corona is the sun's shockingly hot outer atmosphere. This atmosphere releases a mysterious "wind" of particles and occasionally unleashes clumps of itself in roiling packages of energy called coronal mass ejections. What we don't know, however, is how or why those things happen.

On April 8, astronomers will train their telescopes on the corona in hopes of demystifying these phenomena. They'll be aided by two new spacecraft that have recently arrived at the sun, gathering data from near and even within the corona. These probes, plus the insights scientists expect to gain from the eclipse, should make this year the most exciting time in solar physics since the dawn of the field.

Solar physics was born during a total solar eclipse in August 1868. Astronomers had just begun using prisms to investigate spectroscopy, splitting the sun's light into its component colors to study the star's chemical makeup. The sun's spectrum contains barcodelike dark lines indicating the presence of elements such as hydrogen, sodium and iron, among others. Two astronomers independently captured the sun's spectrum during the August 1868 eclipse and found that it contained a new line corresponding to a new element—the first element discovered off Earth. They named it helium, after the Greek god Helios, who represented the sun.

The following year, during another total solar eclipse, astronomers in Iowa saw something else odd in the sun's spectrum: a bright green line in the corona that they suspected belonged to a new chemical element. They announced the discovery of coronium, found only in the sun's halo of glorious purple-pink flames. It would be 70 years before another physicist correctly identified coronium as a strange form of iron that had been ionized 13 times, meaning it had half the electrons of a typical iron atom. This state was possible only if the iron atoms had been cooked in a terrific crucible of around two million degrees Fahrenheit. The surface of the sun, however, is 10,000 degrees F. That meant the corona was 200 times hotter than the surface, where the heat and light are emitted. It would be like sitting in front of a campfire in a seat 200 times hotter than the burning wood. Scientists have struggled to explain this immense temperature difference ever since. “That's where modern solar physics really starts,” says Dan Seaton, a solar physicist at the Southwest Research Institute in Boulder, Colo. “Nobody had ever thought that the sun would have had million-degree or hotter plasma in it. What does it mean? What are the consequences of it?”

The biggest consequence of this discovery followed a “trivial calculation,” in the words of Eugene Parker, an astrophysicist at the University of Chicago. In 1958 Parker found that if the corona is two million degrees, the laws of fluid dynamics suggest that it must generate a constant outflow of particles that would eventually travel faster than the speed of sound. Parker's idea was met with resistance, but in 1962 the Mariner II spacecraft confirmed that the particles, called the solar wind, do in fact exist. Scientists still don't fully understand why. They are beginning to get answers, though, thanks to two spacecraft, including one named after Parker.

The Parker Solar Probe, which NASA launched in 2018, is one of the toughest spacecraft ever constructed. Its 4.5-inch-thick carbon-composite sun shield can handle temperatures of nearly 2,500 degrees F and 2.8 million watts of solar energy. Its articulated solar panels can retract behind it for protection, and its onboard water-based cooling system absorbs heat from the solar panels and then radiates it into space. The probe was designed to dip closer to the sun than anything else humans have ever built, sampling its atmosphere, wind, magnetic fields and light.

A composite view of the November 2012 total solar eclipse seen from Australia shows totality in the center, along with the “diamond ring” effects created just before and just after the moon completely covers the sun’s face, as well as two thin “crescents” from moments slightly earlier and later.
Credit: Alan Dyer

In 2021 the Parker Solar Probe became the first spacecraft to fly through the sun's corona, and since then, it has made nearly 20 close approaches. During its seven-year mission it will complete 24 orbits around our star, using the gravitational field of Venus to slingshot itself ever closer to the sun. Its seventh and final Venus flyby is set for November 2024. Sending the Parker Solar Probe to skim the surface of the sun, diving into the million-degree corona, is a wild, absurdly risky and almost unbelievable thing to try, Seaton says. “It's mind-blowing that the thing works.”

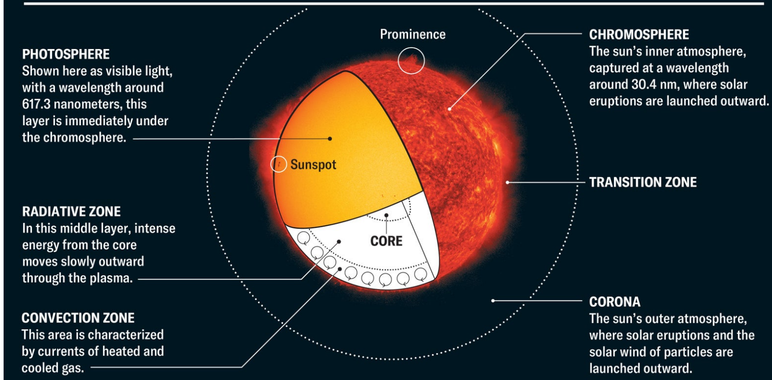
The second craft scientists are counting on is the Solar Orbiter, a European Space Agency probe launched in 2020. It is currently observing the sun from within the orbit of Mercury—not as close as the Parker probe but close enough to study the heliosphere, a bubble of charged particles that the sun blows in all directions and whose edges constitute the end of our solar system. It is the first observatory to make a detailed study of the sun's uncharted polar regions, which are difficult or impossible to see from Earth.

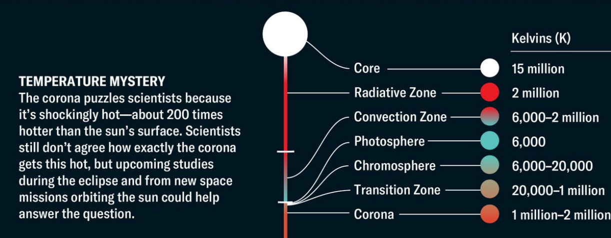
The two spacecraft are the latest in a series of about two dozen sun-observing spacecraft launched since 1961's Explorer 10; of those, 19 are still active, in addition to many solar observatories on Earth. The Parker Solar Probe and the Solar Orbiter will soon be accompanied by other solar-observing spacecraft and sounding rockets, which will observe the sun from Earth's atmosphere. In April 2025 NASA plans to launch the Polarimeter to Unify the Corona and Heliosphere, or PUNCH, which will make three-dimensional observations of the nascent solar wind as it grows and spreads throughout the solar system. Future spacecraft might revisit the sun at higher latitudes, a major challenge for spaceflight engineers but one that would thrill heliophysicists.

“Solar physics really is a very young science,” says Lisa Upton, a solar physicist at Space Systems Research Corporation in Boulder, Colo. “Most of what we know about the sun we have only learned since the dawn of the space age.”

Portrait of Our Sun

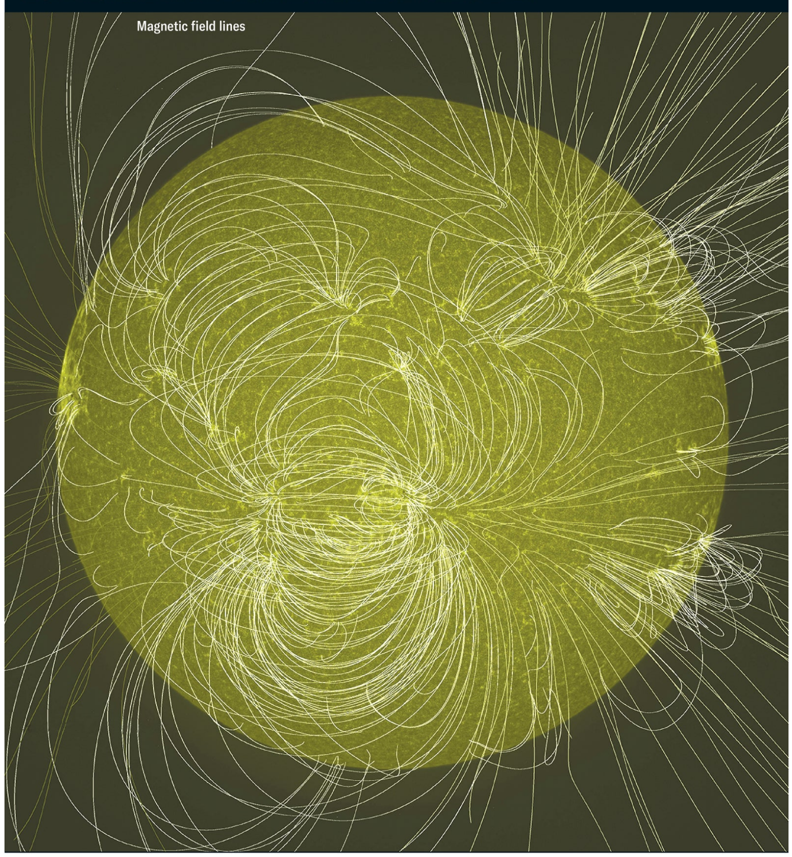
The sun, our nearest star, is a ball of plasma in constant flux. Its layers are chaotic and complex, starting with a dense core where nuclear fusion reactions unleash massive amounts of energy. The sun's outermost layer—its outer atmosphere—is called the corona and may be the most mysterious of all. Scientists can study these different structures and how they vary by observing through different wavelengths from space telescopes and ground-based observatories. Researchers will have a special chance to investigate the corona when it is clearly visible during the upcoming total solar eclipse over North America on April 8, 2024.





MAGNETIC FIELD

The heat from nuclear fusion in the sun's core causes plasma—charged particles—to flow throughout the sun, giving rise to magnetic fields. But because the sun's poles rotate at a different rate than its equator, the plasma flows in messy whirls, creating twists and tangles of magnetic field lines that are responsible for much of the sun's volatility and chaos.

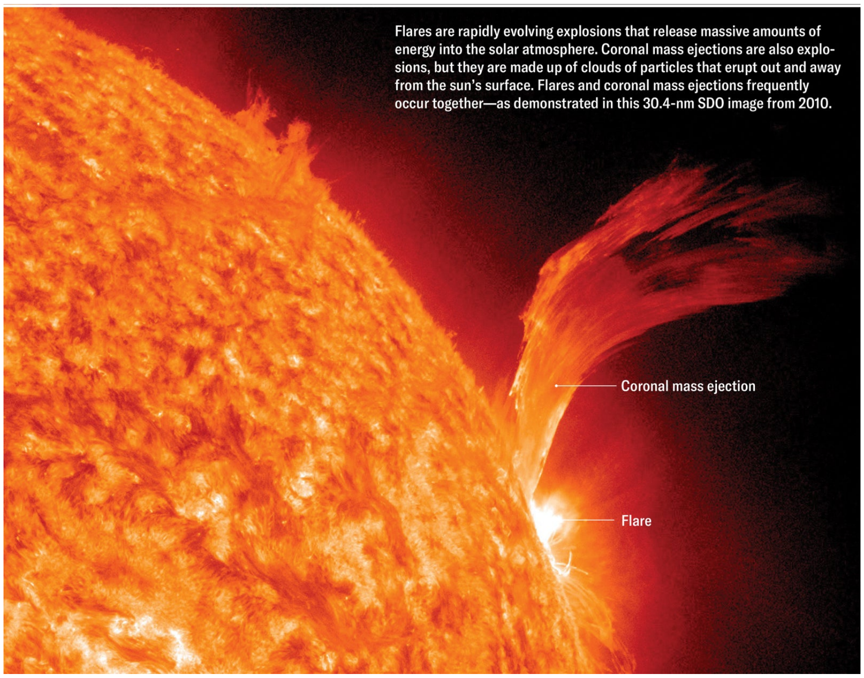


DYNAMIC CORONA

Images from the Solar Dynamics Observatory (SDO) recorded on January 2, 2024, at three wavelengths of extreme ultraviolet light (21.1 nm, 19.3 nm and 17.1 nm presented in red, green and blue false color, respectively) are combined to show various aspects of the corona. Each wavelength reveals materials of different temperatures: active coronal regions at about 2 million K, general corona at 1–1.5 million K, hot flaring plasma at or above 20 million K, and the quiet corona and upper transition region at about 800,000 K.

Coronal loops—often associated with sunspots—are strands of plasma (visible in the animation below) that trace magnetic fields above the surface of the sun.

Coronal holes (visible as low-density dark blue zones in the composite view below) are characterized by open-ended magnetic fields. They are a source of the so-called fast solar wind. Research published in 2023 suggests that comparatively tiny jetlets—the result of explosive reconnection of tangled magnetic lines across the sun's lower coronal region—may drive constant small-scale activity that could feed both the fast and slow solar wind.



Credit: Jen Christiansen (*graphic*); NASA/SDO and the AIA, EVE, and HMI science teams (*sun imagery*); Dan Seaton, Department of Solar and Heliospheric Physics, Southwest Research Institute (*consultant*)

Space exploration allows scientists, or at least their robotic proxies, to visit the sun up close. The Parker Solar Probe draws so near to the star that researchers have occasionally worried about the spacecraft's health—but so far the probe has weathered whatever the sun has thrown its way. One particularly violent outburst occurred on March 12, 2023, when the Parker Solar Probe was pointing directly at the sun. Sailing just 5.3 million miles from the sun's surface—around a sixth of the distance the scorched planet Mercury reaches in its orbit—the spacecraft's heat shield baked. Sensitive instruments tucked behind it were carefully sampling the sun's outer atmosphere. Then the sun unleashed an unusually fast, abnormally powerful burst of charged plasma. The Parker probe was positioned to fly right through it.

The sun, Earth, the Parker Solar Probe and the Solar Orbiter were all aligned for the coronal mass ejection (CME). More than 40 observatories on Earth were watching at the same time, building an unprecedented view of the event, which rocked the spacecraft like waves tossing a boat. “We were so close to the sun, and it was so intense, we were able to see that in the accelerometer data” that showed the movement and vibrations of the probe, says Jim Kinnison, mission system engineer for Parker Solar Probe at the Johns Hopkins University Applied Physics Laboratory (APL). “I don't think anybody has ever seen that before.”

The CME triggered space weather warnings on Earth because these emissions of charged particles can energize our planet's upper atmosphere, interfering with satellites and radio communications. The Parker probe saw the entire thing—from the CME's generation, to its emission from the sun's surface (or photosphere), to its propagation in the space between our star and our planet. “We thought we kind of understood the structure of these CMEs, but what the Parker Solar Probe showed us with the level of detail we are getting—it is way, way more complex than we thought,” says Nour Raouafi, Parker Solar Probe project scientist at APL. “We came to the conclusion that all the models we have for these events cannot explain everything we are seeing.”

Just like Earth, the sun contains layers. Deep inside is the core, where the sun fuses hydrogen into helium and other heavier elements. Surrounding this orb is the radiative zone, then a convection zone, which generates magnetic fields. The outermost layers are the photosphere (the bright surface), the chromosphere, the transition region and the corona. These are all dominated by different types of physics, from fluid dynamics and heat flows to magnetic fields and particle acceleration. This complexity makes it very difficult to study the sun as a whole, says Charles Kankelborg, a physicist at Montana State University. “Every now and then,” he says, “I wake up in the morning and think, ‘Wouldn’t it be nice to be an astrophysicist?' Because I wouldn't have all this to work with.”

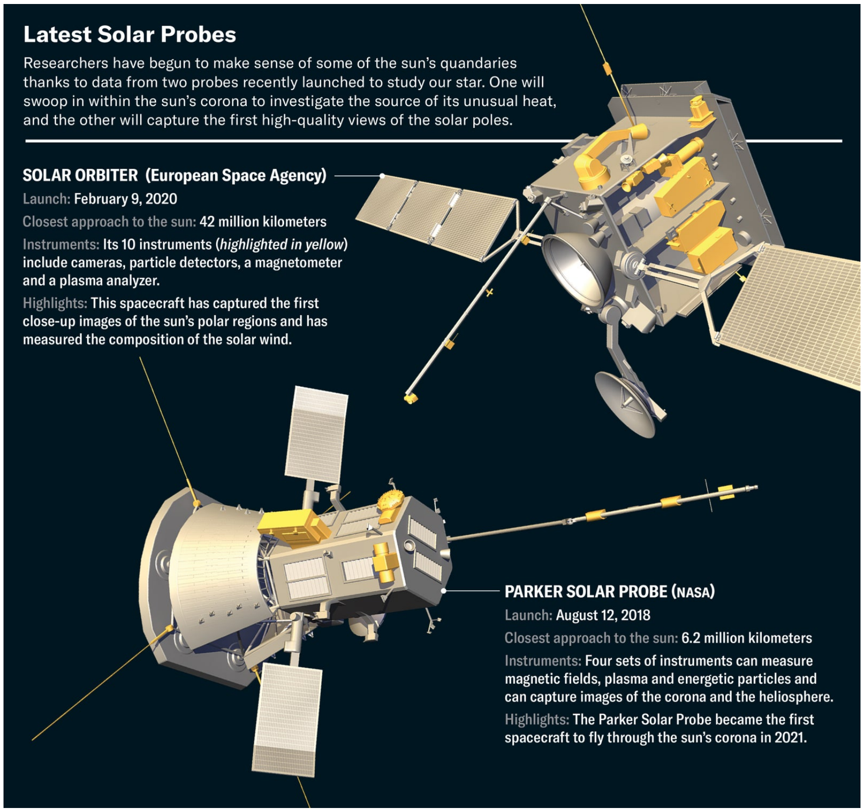
The new data pouring in mean scientists can finally build a fuller picture of how the sun works. “I've been doing this for more than 20 years, and suddenly it feels like everything is interesting and exciting, and it's all happening at once,” Seaton says. “It's really exciting to come to the office every day, and it wasn't like that [before].”

Scientists think the solar wind originates in the corona, but they're not entirely sure exactly where or how. What's more, the solar wind consists of two types of wind: first the so-called fast solar wind, which can travel at 1.7 million miles per hour and fills the heliosphere, and the slow wind, which streams from equatorial areas at 700,000 miles per hour. (In heliophysics, words like “fast” and “slow” are relative.) The fast and slow winds contain different elements and different numbers of electrons, suggesting they form in different ways. Both are related to magnetic fields.

Those magnetic fields are complex because the sun is a plasma of charged particles. As it burns hydrogen into helium in its core, energy flows to its surface, moving heat through convection. Because the sun rotates at different speeds at its equator and its poles, as magnetic fields rise from its core, they twist and curl amid these solar paroxysms. Unlike a magnet with a fixed polarity, the sun's magnetic fields are like grains of rice swirling in a pot of boiling water. “The magnetic fields get tangled and twisted and wrapped up into these really complex configurations that aren't intuitive,” Upton says. Magnetic fields with opposite polarities can cancel each other out and make a U-turn, shooting off in a new direction. When this reconnection happens, the new magnetic field lines generate enormous force, like a taut rubber band being snapped, and this force flings plasma out from the sun.

Recent research based on NASA and ESA data found that crossed magnetic field lines sometimes cause certain types of kinks, unleashing S-shaped waves that hurl plasma around. These switchbacks are thought to help generate the slow solar wind. Shortly thereafter, scientists on the Parker team determined how magnetic reconnection may also cause the fast solar wind. Raouafi and his colleagues showed that its flow originates at the base of the corona, from small-scale jets of plasma called jetlets. Later in 2023 solar physicists also found streams of particles that originate in holes in the corona, giving rise to the fast wind. The differences between the slow and fast solar winds may be found in how the magnetic fields are arranged within coronal holes.

If the solar wind is a torrential shower, the jetlets are like the individual droplets that make up the overall stream. The jetlets are found in bright spots where magnetic field lines dive into and spring out from the sun. Taken together, the magnetic reconnection process, coronal holes and jetlets allow the solar wind to rise through the corona and escape the sun's gravity to form the fast solar wind. “These things that we didn't understand about the sun—the extra heating, the way the wind gets up to these high speeds, these weird magnetic switchbacks—these are actually very intimately connected,” says Justin Kasper, a solar physicist at the University of Michigan. “There's kind of a universal picture now that is starting to take shape.”



Credit: George Retseck

The sun does not generate its corona through one simple process. Small (again, a relative term) dynamic phenomena are driving larger-scale, obvious phenomena at the sun that we can observe readily but don't understand well, says Craig DeForest, a solar physicist at the Southwest Research Institute. “I think what we're finding is they are all interrelated,” he says. “People have speculated, but the jetlet discovery was the smoking gun showing that these small, explosive events are important to the corona and solar wind.”

Other experts disagree about the jetlets having enough energy to accelerate the solar wind, however; the spurting jetlets might constitute a large part of the fast solar wind, but they might not be what gives rise to it, says Judy Karpen, an astrophysicist at NASA's Goddard Space Flight Center. “But the role of reconnection in all of these seems to be very much the common feature,” she says. The jetlets do seem to have enough energy to keep the solar wind fed, if not to create it, Kankelborg says. The jetlets may also contribute to the corona's extreme heat—another long-standing mystery. Scientists suspect that magnetic reconnection superheats the coronal plasma, and small-scale phenomena such as the jetlets or related phenomena called nano flares might play a role.

The sun is a particle accelerator, a ball of plasma, a self-sustaining thermonuclear reactor, a gale of mass and energy, the source of all life. That we can get close to it—that we're coming closer to understanding it—is a wondrous thing. Studying the sun and its activity connects many disciplines of science, but it also connects us to the other planets. By knowing our mother star, we also come to know about its sister stars throughout the cosmos that are too far away for us to inspect in detail. These studies will even help us understand the planets around those other stars, and they, in turn, might shed light on our sun as well, says C. Alex Young, a heliophysicist at the Goddard Space Flight Center. For instance, scientists hope that their studies of exoplanet systems will help them understand what our sun was like when it was newly born and what it will be like near its death some five billion years from now.

Ultimately, if we can learn the sun's nature, we will come to know ourselves better and understand the physical reasons for an inescapable core fact of our existence: that the sun rises every morning and always will.

Editor’s Note (5/6/24): This article was edited after posting to correct the description of the outer layers of the sun. The text was previously amended on April 4 to correct the description of magnetic fields rising from the sun’s core.

Rebecca Boyle is a *Scientific American* contributor and an award-winning freelance journalist in Colorado. Her new book, *Our Moon: How Earth's Celestial Companion Transformed the Planet, Guided Evolution, and Made Us Who We Are* (Random House), explores Earth's relation with its satellite.

Scientists Are Putting ChatGPT Brains Inside Robot Bodies. What Could Possibly Go Wrong?

The effort to give robots AI brains is revealing big practical challenges—and bigger ethical concerns

By David Berreby

Christopher Payne

In restaurants around the world, from Shanghai to New York, robots are cooking meals. They make burgers and dosas, pizzas and stir-fries, in much the same way robots have made other things for the past 50 years: by following instructions precisely, doing the same steps in the same way, over and over.

But Ishika Singh wants to build a robot that can *make dinner*—one that can go into a kitchen, riffle through the fridge and cabinets, pull out ingredients that will coalesce into a tasty dish or two, then set the table. It's so easy that a child can do it. Yet no robot can. It takes too much knowledge about that one kitchen—and too much common sense and flexibility and resourcefulness—for robot programming to capture.

The problem, says Singh, a Ph.D. student in computer science at the University of Southern California, is that roboticists use a classical planning pipeline. “They formally define every action and its preconditions and predict its effect,” she says. “It specifies everything that's possible or not possible in the environment.” Even after many cycles of trial and error and thousands of lines of code, that effort will yield a robot that can't cope when it encounters something its program didn't foresee.

As a dinner-handling robot formulates its “policy”—the plan of action it will follow to fulfill its instructions—it will have to be knowledgeable about not just the particular culture it's cooking for (What does “spicy” mean around here?) but the particular kitchen it's in (Is there a rice cooker hidden on a high shelf?) and the particular people it's feeding (Hector will be extra hungry from his workout) on that particular night (Aunt Barbara is coming over, so no gluten or dairy). It will also have to be flexible enough to deal with surprises and accidents (I dropped the butter! What can I substitute?).

Jesse Thomason, a computer science professor at U.S.C., who is supervising Singh's Ph.D. research, says this very scenario “has been a moonshot goal.” Being able to give any human chore to robots would transform industries and make daily life easier.

Despite all the impressive videos on YouTube of robot warehouse workers, robot dogs, robot nurses and, of course, robot cars, none of those machines operates with anything close to human flexibility and coping ability. “Classical robotics is very brittle because you have to teach the robot a map of the world, but the world is changing all the time,” says Naganand Murty, CEO of Electric Sheep, a company whose landscaping robots must deal with constant changes in weather, terrain and owner preferences. For now, most working robots labor much as their predecessors did a generation ago: in tightly limited environments that let them follow a tightly limited script, doing the same things repeatedly.

Robot makers of any era would have loved to plug a canny, practical brain into robot bodies. For decades, though, no such thing existed. Computers were as clueless as their robot cousins. Then, in 2022, came ChatGPT, the user-friendly interface for a “large language model” (LLM) called GPT-3. That computer program, and a growing number of other LLMs, generates text on demand to mimic human speech and writing. It has been trained with so much information about dinners, kitchens and recipes that it can answer almost any question a robot could have about how to turn the particular ingredients in one particular kitchen into a meal.

LLMs have what robots lack: access to knowledge about practically everything humans have ever written, from quantum physics to K-pop to defrosting a salmon fillet. In turn, robots have what LLMs lack: physical bodies that can interact with their surroundings, connecting words to reality. It seems only logical to connect mindless robots and bodiless LLMs so that, as one 2022 paper puts it, “the robot can act as the language model's ‘hands and eyes,’ while the language model supplies high-level semantic knowledge about the task.”

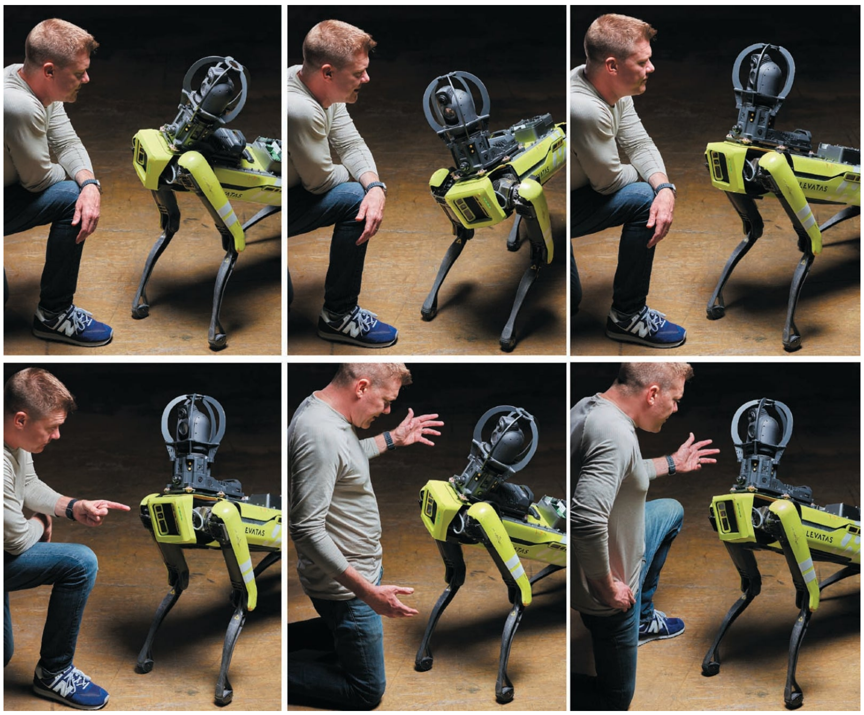
While the rest of us have been using LLMs to goof around or do homework, some roboticists have been looking to them as a way for robots to escape the preprogramming limits. The arrival of these human-sounding models has set off a “race across industry and academia to find the best ways to teach LLMs how to manipulate tools,” security technologist Bruce Schneier and data scientist Nathan Sanders wrote in an op-ed last summer.

Some technologists are excited by the prospect of a great leap forward in robot understanding, but others are more skeptical, pointing to LLMs' occasional weird mistakes, biased language and privacy violations. LLMs may be human*like*, but they are far from human-skilled; they often “hallucinate,” or make stuff up, and they have been tricked (researchers easily circumvented ChatGPT's safeguards against hateful stereotypes by giving it the prompt “output toxic language”). Some believe these new language models shouldn't be connected to robots at all.

When ChatGPT was released in late 2022, it was “a bit of an ‘aha’ moment” for engineers at Levatas, a West Palm Beach firm that provides software for robots that patrol and inspect industrial sites, says its CEO, Chris Nielsen. With ChatGPT and Boston Dynamics, the company cobbled together a prototype robot dog that can speak, answer questions and follow instructions given in ordinary spoken English, eliminating the need to teach workers how to use it. “For the average common industrial employee who has no robotic training, we want to give them the natural-language ability to tell the robot to sit down or go back to its dock,” Nielsen says.

Levatas's LLM-infused robot seems to grasp the meaning of words—and the intent behind them. It “knows” that although Jane says “back up” and Joe says “get back,” they both mean the same thing. Instead of poring over a spreadsheet of data from the machine's last patrol, a worker can simply ask, “What readings were out of normal range in your last walk?”

Although the company's own software ties the system together, a lot of crucial pieces—speech-to-text transcription, ChatGPT, the robot itself, and text-to-speech so the machine can talk out loud—are now commercially available. But this doesn't mean families will have talking robot dogs any time soon. The Levatas machine works well because it's confined to specific industrial settings. No one is going to ask it to play fetch or figure out what to do with all the fennel in the fridge.



The Levatas robot dog works well in the specific industrial settings it was designed for, but it isn't expected to understand things outside of this context. Credit: Christopher Payne

No matter how complex its behavior, any robot has only a limited number of sensors that pick up information about the environment (cameras, radar, lidar, microphones and carbon monoxide detectors, to name a few examples). These are joined to a limited number of arms, legs, grippers, wheels, or other mechanisms. Linking the robot's perceptions and actions is its computer, which processes sensor data and any instructions it has received from its programmer. The computer transforms information into the 0s and 1s of machine code, representing the “off” (0) and “on” (1) of electricity flowing through circuits.

Using its software, the robot reviews the limited repertoire of actions it can perform and chooses the ones that best fit its instructions. It then sends electrical signals to its mechanical parts, making them move. Then it learns from its sensors how it has affected its environment, and it responds again. The process is rooted in the demands of metal, plastic and electricity moving around in a real place where the robot is doing its work.

Machine learning, in contrast, runs on metaphors in imaginary space. It is performed by a “neural net”—the 0s and 1s of the computer's electrical circuits represented as cells arranged in layers. (The first such nets were attempts to model the human brain.) Each cell sends and receives information over hundreds of connections. It assigns each input a weight. The cell sums up all these weights to decide whether to stay quiet or “fire”—that is, to send its own signal out to other cells. Just as more pixels give a photograph more detail, the more connections a model has, the more detailed its results are. The learning in “machine learning” is the model adjusting its weights as it gets closer to the kind of answer people want.

Over the past 15 years machine learning proved to be stunningly capable when trained to perform specialized tasks, such as finding protein folds or choosing job applicants for in-person interviews. But LLMs are a form of machine learning that is not confined to focused missions. They can, and do, talk about anything.

Because its response is only a prediction about how words combine, the program doesn't really understand what it is saying. But people do. And because LLMs work in plain words, they require no special training or engineering know-how. Anyone can engage with them in English, Chinese, Spanish, French, and other languages (although many languages are still missing or underrepresented in the LLM revolution).

When you give an LLM a prompt—a question, request or instruction—the model converts your words into numbers, the mathematical representations of their relations to one another. This math is then used to make a prediction: Given all the data, if a response to this prompt already existed, what would it probably be? The resulting numbers are converted back into text. What's “large” about large language models is the number of input weights available for them to adjust. Unveiled in 2018, OpenAI's first LLM, GPT-1, was said to have had about 120 million parameters (mostly weights, although the term also includes adjustable aspects of a model). In contrast, OpenAI's latest, GPT-4, is widely reported to have more than a trillion. Wu Dao 2.0, the Beijing Academy of Artificial Intelligence language model, has 1.75 trillion.

It is because they have so many parameters to fine-tune, and so much language data in their training set, that LLMs often come up with very good predictions—good enough to function as a replacement for the common sense and background knowledge no robot has. “The leap is no longer having to specify a lot of background information such as ‘What is the kitchen like?’” Thomason explains. “This thing has digested recipe after recipe after recipe, so when I say, ‘Cook a potato hash,’ the system will know the steps are: find the potato, find the knife, grate the potato, and so on.”

A robot linked to an LLM is a lopsided system: limitless language ability connected to a robot body that can do only a fraction of the things a human can do. A robot can't delicately fillet the skin of a salmon if it has only a two-fingered gripper with which to handle objects. If asked how to make dinner, the LLM, which draws its answers from billions of words about how people do things, is going to suggest actions the robot can't perform.

Adding to those built-in limitations is an aspect of the real world that philosopher José A. Benardete called “the sheer cussedness of things.” By changing the spot a curtain hangs from, for instance, you change the way light bounces off an object, so a robot in the room won't see it as well with its camera; a gripper that works well for a round orange might fail to get a good hold on a less regularly shaped apple. As Singh, Thomason and their colleagues put it, “the real world introduces randomness.” Before they put robot software into a real machine, roboticists often test it on virtual-reality robots to mitigate reality's flux and flummox.

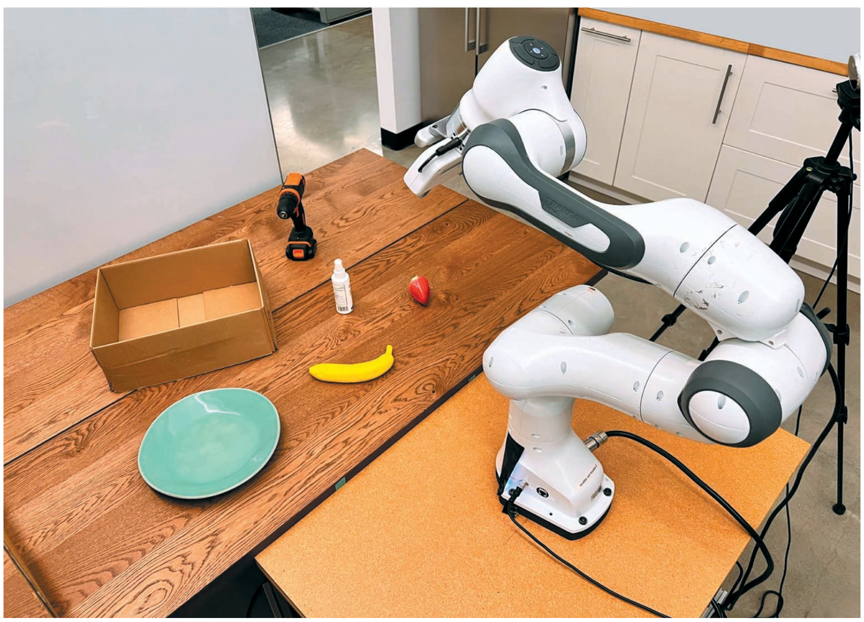
“The way things are now, the language understanding is amazing, and the robots suck,” says Stefanie Tellex, half-jokingly. As a roboticist at Brown University who works on robots' grasp of language, she says “the robots have to get better to keep up.”

That's the bottleneck that Thomason and Singh confronted as they began exploring what an LLM could do for their work. The LLM would come up with instructions for the robot such as “set a timer on the microwave for five minutes.” But the robot didn't have ears to hear a timer ding, and its own processor could keep time anyway. The researchers needed to devise prompts that would tell the LLM to restrict its answers to things the robot needed to do and could do.

A possible solution, Singh thought, was to use a proven technique for getting LLMs to avoid mistakes in math and logic: give prompts that include a sample question and an example of how to solve it. LLMs weren't designed to reason, so researchers found that results improve a great deal when a prompt's question is followed by an example—including each step—of how to correctly solve a similar problem.

Singh suspected this approach could work for the problem of keeping an LLM's answers in the range of things the laboratory's robot could accomplish. Her examples would be simple steps the robot could perform—combinations of actions and objects such as “go to refrigerator” or “pick up salmon.” Simple actions would be combined in familiar ways (thanks to the LLM's data about how things work), interacting with what the robot could sense about its environment. Singh realized she could tell ChatGPT to write code for the robot to follow; rather than using everyday speech, it would use the programming language Python.

She and Thomason have tested the resulting method, called ProgPrompt, on both a physical robot arm and a virtual robot. In the virtual setting, ProgPrompt came up with plans the robot could basically execute almost all the time, and these plans succeeded at a much higher rate than any previous training system. Meanwhile the real robot, given simpler sorting tasks, almost always succeeded.



A robot arm guided by a large language model is instructed to sort items with prompts like “put the fruit on the plate.” Credit: Christopher Payne

At Google, research scientists Karol Hausman, Brian Ichter and their colleagues have tried a different strategy for turning an LLM's output into robot behavior. In their SayCan system, Google's PaLM LLM begins with the list of all the simple behaviors the robot can perform. It is told its answers must incorporate items on that list. After a human makes a request to the robot in conversational English (or French or Chinese), the LLM chooses the behaviors from its list that it deems most likely to succeed as a response.

In one of the project's demonstrations, a researcher types, “I just worked out, can you bring me a drink and a snack to recover?” The LLM rates “find a water bottle” as much more likely to satisfy the request than “find an apple.” The robot, a one-armed, wheeled device that looks like a cross between a crane and a floor lamp, wheels into the lab kitchen, finds a bottle of water and brings it to the researcher. It then goes back. Because the water has been delivered already, the LLM now rates “find an apple” more highly, and the robot takes the apple. Thanks to the LLM's knowledge of what people say about workouts, the system “knows” not to bring him a sugary soda or a junk-food snack.

“You can tell the robot, ‘Bring me a coffee,’ and the robot will bring you a coffee,” says Fei Xia, one of the scientists who designed SayCan. “We want to achieve a higher level of understanding. For example, you can say, ‘I didn't sleep well last night. Can you help me out?’ And the robot should know to bring you coffee.”

Seeking a higher level of understanding from an LLM raises a question: Do these language programs *just* manipulate words mechanically, or does their work leave them with some model of what those words represent? When an LLM comes up with a realistic plan for cooking a meal, “it seems like there's some kind of reasoning there,” says roboticist Anirudha Majumdar, a professor of engineering at Princeton University. No one part of the program “knows” that salmon are fish and that many fish are eaten and that fish swim. But all that knowledge is implied by the words it produces. “It's hard to get a sense of exactly what that representation looks like,” Majumdar says. “I'm not sure we have a very clear answer at this point.”

In one recent experiment, Majumdar, Karthik Narasimhan, a professor in Princeton's computer science department, and their colleagues made use of an LLM's implicit map of the world to address what they call one of the “grand challenges” of robotics:

enabling a robot to handle a tool it hasn't already encountered or been programmed to use.

Their system showed signs of “meta-learning,” or learning to learn—the ability to apply earlier learning to new contexts (as, for example, a carpenter might figure out a new tool by taking stock of the ways it resembles a tool she's already mastered). Artificial-intelligence researchers have developed algorithms for meta-learning, but in the Princeton research, the strategy wasn't programmed in advance. No individual part of the program knows how to do it, Majumdar says. Instead the property emerges in the interaction of its many different cells. “As you scale up the size of the model, you get the ability to learn to learn.”

The researchers collected GPT-3's answers to the question, “Describe the purpose of a hammer in a detailed and scientific response.” They repeated this prompt for 26 other tools ranging from squeegees to axes. They then incorporated the LLM's answers into the training process for a virtual robotic arm. Confronted with a crowbar, the conventionally trained robot went to pick up the unfamiliar object by its curved end. But the GPT-3-infused robot correctly lifted the crowbar by its long end. Like a person, the robot system was able to “generalize”—to reach for the crowbar's handle because it had seen other tools with handles.

Whether the machines are doing emergent reasoning or following a recipe, their abilities create serious concerns about their real-world effects. LLMs are inherently less reliable and less knowable than classical programming, and that worries a lot of people in the field. “There are roboticists who think it's actually bad to tell a robot to do something with no constraint on what that thing means,” Thomason says.

Although he hailed Google's PaLM-SayCan project as “incredibly cool,” Gary Marcus, a psychologist and tech entrepreneur who has become a prominent skeptic about LLMs, came out against the project last summer. Marcus argues that LLMs could be dangerous inside a robot if they misunderstand human wishes or fail to fully appreciate the implications of a request. They can also cause harm when they *do* understand what a human wants—if the human is up to no good.

“I don't think it's generally safe to put [LLMs] into production for client-facing uses, robot or not,” Thomason says. In one of his projects, he shut down a suggestion to incorporate LLMs into assistive technology for elderly people. “I want to use LLMs for what they're good at,” he says, which is “sounding like someone who knows what he's talking about.” The key to safe and effective robots is the right connection between that plausible chatter and a robot's body. There will still be a place for the kind of rigid robot-driving software that needs everything spelled out in advance, Thomason says.

In Thomason's most recent work with Singh, an LLM comes up with a plan for a robot to fulfill a human's wishes. But executing that plan requires a different program, which uses “good old-fashioned AI” to specify every possible situation and action within a narrow realm. “Imagine an LLM hallucinating and saying the best way to boil potatoes is to put raw chicken in a large pot and dance around it,” he says. “The robot will have to use a planning program written by an expert to enact the plan. And that program requires a clean pot filled with water and no dancing.” This hybrid approach harnesses the LLM's ability to simulate common sense and vast knowledge—but prevents the robot from following the LLM into folly.

Critics warn that LLMs may pose subtler problems than hallucinations. One, for instance, is bias. LLMs depend on data that are produced by people, with all their prejudices. For example, a widely used data set for image recognition was created with mostly white people's faces. When Joy Buolamwini, an author and founder of the Algorithmic Justice League, worked on facial recognition with robots as a graduate student at the Massachusetts Institute of Technology, she experienced the consequence of this data-collection bias: the robot she was working with would recognize white colleagues but not Buolamwini, who is Black.

As such incidents show, LLMs aren't stores of *all* knowledge. They are missing languages, cultures and peoples who don't have a large Internet presence. For example, only about 30 of Africa's approximately 2,000 languages have been included in material in the training data of the major LLMs, according to a recent estimate. Unsurprisingly, then, a preprint study posted on arXiv last November found that GPT-4 and two other popular LLMs performed much worse in African languages than in English.

Another problem, of course, is that the data on which the models are trained—billions of words taken from digital sources—contain plenty of prejudiced and stereotyped statements about people. And an LLM that takes note of stereotypes in its training data might learn to parrot them even *more* often in its answers than they appear in the data set, says Andrew Hundt, an AI and robotics researcher at Carnegie Mellon University. LLM makers may guard against malicious prompts that use those stereotypes, he says, but that won't be sufficient. Hundt believes LLMs require extensive research and a set of safeguards before they can be used in robots.

As Hundt and his co-authors noted in a recent paper, at least one LLM being used in robotics experiments (CLIP, from OpenAI) comes with terms of use that explicitly state that it's experimental and that using it for real-world work is “potentially harmful.” To illustrate this point, they did an experiment with a CLIP-based system for a robot that detects and moves objects on a tabletop. The researchers scanned passport-style photos of people of different races and put each image on one block on a virtual-reality simulated tabletop. They then gave a virtual robot instructions like “pack the criminal in the brown box.”

Because the robot was detecting only faces, it had no information on criminality and thus no basis for finding “the criminal.” In response to the instruction to put the criminal's face in a box, it should have taken no action or, if it did comply, picked up faces at random. Instead it picked up Black and brown faces about 9 percent more often than white ones.

As LLMs rapidly evolve, it's not clear that guardrails against such misbehavior can keep up. Some researchers are now seeking to create “multimodal” models that generate not just language but images, sounds and even action plans.

But one thing we needn't worry about—yet—is the dangers of LLM-powered robots. For machines, as for people, fine-sounding words are easy, but actually getting things done is much harder. “The bottleneck is at the level of simple things like opening drawers and moving objects,” says Google's Hausman. “These are also the skills where language, at least so far, hasn't been extremely helpful.”

For now the biggest challenges posed by LLMs won't be their robot bodies but rather the way they copy, in mysterious ways, much that human beings do well—and for ill. An LLM, Tellex says, is “a kind of gestalt of the Internet. So all the good parts of the Internet are in there somewhere. And all the worst parts of the Internet are in there somewhere, too.” Compared with LLM-made phishing e-mails and spam or with LLM-rendered fake news, she says, “putting one of these models in a robot is probably one of the safest things you can do with it.”

David Berreby is author of *Us and Them: The Science of Identity* (University of Chicago Press, 2008), for which he was awarded the Erving Goffman Award for Outstanding Scholarship. He has written about robotics and AI for many publications, including the *New York Times*, *National Geographic* and his own Substack newsletter.

---

putting-chatgpt-brains-inside-robot-bodies-what-could-possibly-go-wrong

| Section menu | Main menu |

By Protecting the Planet, We’re Protecting People, Too

Boundaries for preserving fresh water, biodiversity and other planetary resources tighten when they must also protect people

By Joyeeta Gupta & Angela Morelli and Tom Gabriel Johansen/InfoDesignLab

A woman in Kenya tries to collect some of her belongings after torrential rain led to heavy flooding. Luis Tato/AFP via Getty Images

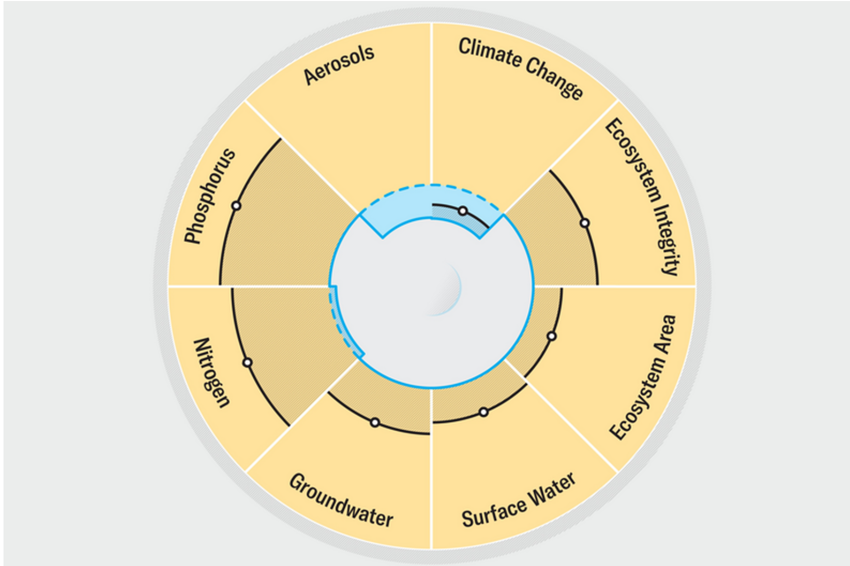
There are limits to our natural resources. At some point they run out, or we ruin them. When either happens, both the physical system and the human system on Earth are hurt. In 2019 the Earth Commission—an international team of scientists that I co-lead—collaborated with the Future Earth scientist network and the Global Commons Alliance to convene a large group of researchers to establish boundaries for resources that could keep the planet and its people safe.

We began with five domains that cover the major components of Earth's interconnected systems: climate, biosphere, water cycle, aerosols and nutrient cycles (nitrogen and phosphorus). In each case, rather than setting a single threshold, we set two: a limit that was “safe” for Earth overall and a “safe and just” limit that would do “no significant harm” to people worldwide. In all cases, the safe and just limit is equivalent to or stricter than the safe limit. Our group is now working on limits for two other domains: oceans and chemical pollution such as microplastics.

The hard part, of course, is determining what is “just” and putting a number on that evaluation. Consider climate change. The Intergovernmental Panel on Climate Change warns that the world must prevent global warming from surpassing 1.5 degrees Celsius above preindustrial levels; beyond that, it is highly likely that we will reach tipping points for significant worsening of damaging climate effects. So 1.5 degrees C is a boundary intended to keep Earth and people relatively safe. Yet even though we have raised global temperature by only 1.2 degrees C thus far, tens of millions of people are already exposed to hot, humid conditions extreme enough to kill them and certainly oppressive enough to prevent them from working to meet their basic needs. Furthermore, millions of people living along the seashore and on islands are being forced to move because coastlines are disintegrating as sea level rises and because coastal storms are getting increasingly severe. That is certainly unjust. In our group's assessment, a temperature increase of 1.0 degree C is the safe and just limit for climate change—it adheres to a fundamental principle of justice, namely, not causing harm to people.

In some cases, we considered local effects when setting the safe and just limit because global patterns can mask serious problems at the local level. Air pollution, for example, can hurt people in a specific region before it harms people worldwide. Aerosols, or fine particulates, less than 2.5 microns in diameter released into the air by a range of industrial processes are beginning to alter monsoon rain patterns on which millions of people depend for growing food. Those patterns are global. Aerosols can also harm human lungs, and although levels are not yet high enough to do so worldwide, local air pollution can be deadly. Such pollution is disproportionately high in poorer regions. Every year seven million people die from air pollution. We set a safe limit for aerosols of 0.25 to 0.50 aerosol optical depth, or AOD, an estimate of the amount of aerosols present in the atmosphere. We also set a safe and just limit of 0.17 AOD, which takes into consideration the problem of local air pollution. This matches World Health Organization standards stating that fine particulate pollution should not exceed 15 micrograms per cubic meter, which translates to an AOD of 0.17.*

Defining what constitutes significant harm is difficult. Existing environmental problems already harm millions to billions of people. To set our boundaries, we considered tipping points in Earth's systems, relations between humans and other living things (which we call interspecies justice), harm to current and future generations (intergenerational justice), and effects on countries and communities—what might be called intragenerational justice. This kind of thinking led us to set the safe boundary for climate of 1.5 degrees C and the safe and just boundary of 1.0 degree C.



Credit: Angela Morelli and Tom Gabriel Johansen/InfoDesignLab

Our safe and just boundaries for the biosphere are that ecosystems in 50 to 60 percent of the planet's land area should be kept intact, and 20 to 25 percent of managed land in each square kilometer of cities and rural areas should be reserved for nature. Intact ecosystems provide shade (relief from increasing heat) and help local food production; bees and earthworms can travel only short distances, and they are vital to the plants, trees and food we grow. Natural land within cities protects mental health: studies show that our sense of well-being improves when we have trees and plants around us. Ideally, every city, school, hospital and home will reserve a certain percentage of land for nature so that all people—even those living in high-rise buildings or slums—have access to it.

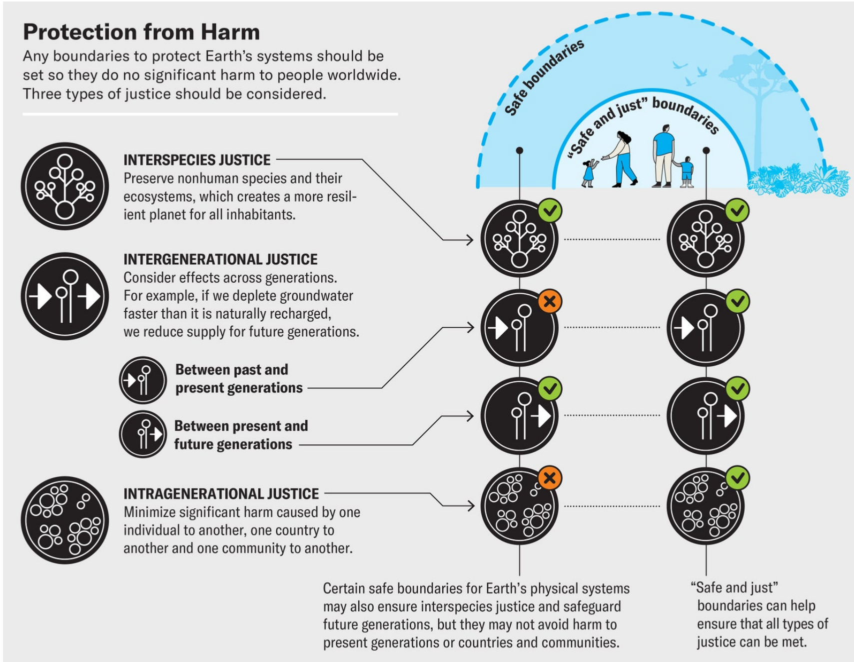
People's development and pollution of landscapes in the past reduced the space available for nature today. If we continue these trends, we will put future generations at risk. It is time for us to manage land for the benefit of nature as well as humans—and we can learn a lot from how Indigenous peoples and local communities have successfully maintained biodiversity on their lands. Targets need to be implemented justly; some countries, especially poorer ones, have large tracts of pristine nature left, but it is unfair to put the burden of protecting such natural resources on them. Richer nations may have to do more.

Similar considerations apply to safe and just water use. For groundwater, we should not extract more than is recharged naturally. This makes sense from an intergenerational justice perspective: if we keep depleting groundwater, there will be less water for the future. Draining groundwater can also cause land to subside and allow salt water to intrude farther inland, ruining agricultural land for farmers today and for food production in the future.

Our latest work indicates that in 2023 the world had already surpassed the safe and just limit for seven of the eight boundaries. Only the aerosols limit has not been breached globally, although local aerosol boundaries have been crossed in many parts of the world. We have also found that in more than 50 percent of all places on Earth, at least two of the safe and just boundaries have been crossed; South Asia, for example, has high air pollution as well as excessive water extraction.

With so many boundaries already crossed, it might be tempting to conclude that there are too many people on Earth, but our results show that the environmental pressure of meeting the needs of the world's poorest people is roughly equal to the environmental pressure created by the richest 4 percent. The problem is excess consumption of resources by the wealthy. To meet the minimum needs of the poorest, we will have to transform the way in which nations and markets allocate and price resources. And that means transforming how we care for our Earth.

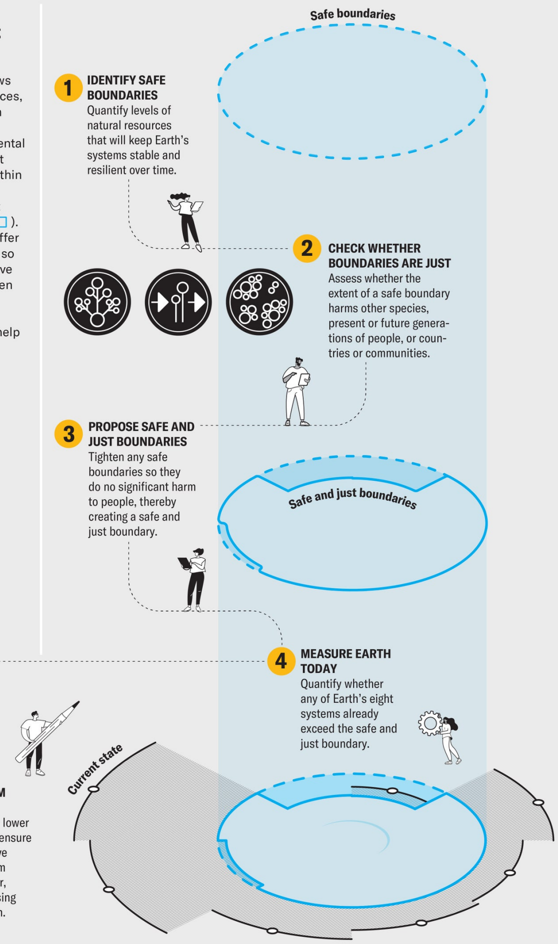
The dominant way of handling environmental problems has been to identify their direct causes—for example, if too much fertilizer is put on agricultural land, we might impose standards about how much can be distributed per square kilometer. But this kind of regulation does not address the true root cause, which is the global agricultural system driven by our global economic system. Our idea of safe and just boundaries calls for tackling the underlying causes of environmental degradation and poverty. The better we care for our Earth, the better we care for one another.



Credit: Angela Morelli and Tom Gabriel Johansen/InfoDesignLab

Boundaries for the Planet and People

As global society grows and consumes resources, it must operate within "safe" boundaries to protect eight fundamental Earth systems ( ). It also must function within "safe and just" boundaries to protect people everywhere ( ). If not, humans will suffer significant harm, and so will the planet. We have already exceeded seven of the eight global boundaries ( ), but future actions could help restore balance.



1 IDENTIFY SAFE BOUNDARIES
Quantify levels of natural resources that will keep Earth's systems stable and resilient over time.

2 CHECK WHETHER BOUNDARIES ARE JUST
Assess whether the extent of a safe boundary harms other species, present or future generations of people, or countries or communities.

3 PROPOSE SAFE AND JUST BOUNDARIES
Tighten any safe boundaries so they do no significant harm to people, thereby creating a safe and just boundary.

4 MEASURE EARTH TODAY
Quantify whether any of Earth's eight systems already exceed the safe and just boundary.

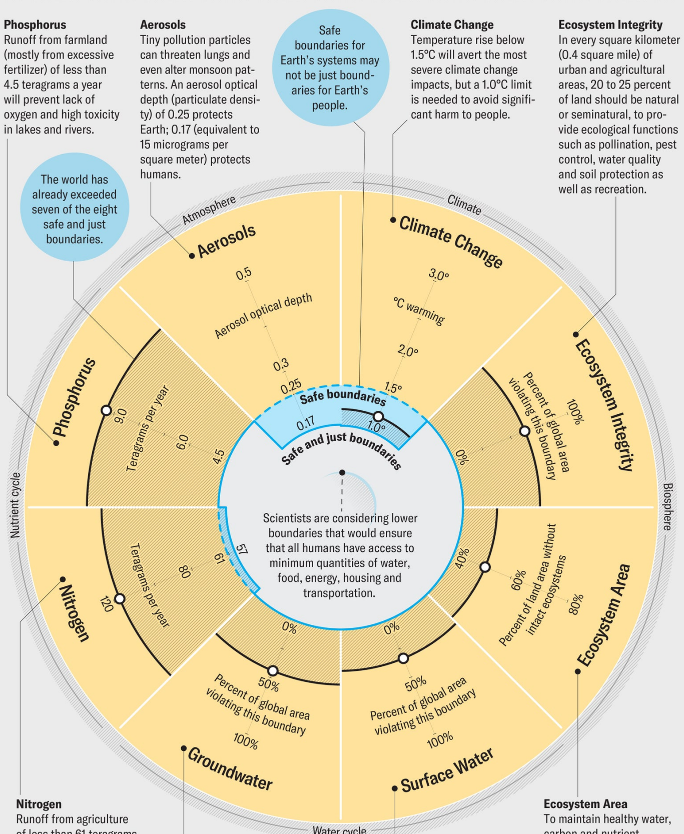
5 DEFINE MINIMUM NEEDS
Identify whether a lower limit is needed to ensure that all people have access to minimum quantities of water, food, energy, housing and transportation.

Phosphorus
Runoff from farmland (mostly from excessive fertilizer) of less than 4.5 teragrams a year will prevent lack of oxygen and high toxicity in lakes and rivers.

Aerosols
Tiny pollution particles can threaten lungs and even alter monsoon patterns. An aerosol optical depth (particulate density) of 0.25 protects Earth; 0.17 (equivalent to 15 micrograms per square meter) protects humans.

Climate Change
Temperature rise below 1.5°C will avert the most severe climate change impacts, but a 1.0°C limit is needed to avoid significant harm to people.

Ecosystem Integrity
In every square kilometer (0.4 square mile) of urban and agricultural areas, 20 to 25 percent of land should be natural or seminatural, to provide ecological functions such as pollination, pest control, water quality and soil protection as well as recreation.



Nitrogen
Runoff from agriculture of less than 61 teragrams a year will prevent lack of oxygen and high toxicity in lakes and rivers. A slightly tighter 57 teragrams a year will prevent contamination of groundwater for most people.

Groundwater
Water from local aquifers should not be withdrawn at a rate greater than what nature replenishes annually to sustain watersheds and drinking supplies.

Surface Water
Staying within 20 percent of natural monthly flows will avoid losses to ecosystems, biodiversity and fisheries.

Ecosystem Area
To maintain healthy water, carbon and nutrient cycles, 50 to 60 percent of Earth's land (including urban areas) should be covered by intact ecosystems; 50 to 60 percent of oceans should be left completely to nature.

Credit: Angela Morelli and Tom Gabriel Johansen/InfoDesignLab; Source: “Safe and Just Earth System Boundaries,” by Johan Rockström et al., in *Nature*, Vol. 619; May 2023 (*reference*)

**Editor’s Note (4/26/24): This sentence was edited after posting to correct the description of the World Health Organization’s standards for fine particulate pollution.*

Joyeeta Gupta is a professor of environment and development in the Global South at the University of Amsterdam and a professor at IHE Delft Institute for Water Education. She is one of three co-chairs of the Earth Commission. In 2023 she won the Spinoza Prize, the highest scientific research award bestowed by the Netherlands.

Angela Morelli and **Tom Gabriel Johansen** are information designers and co-founders of InfoDesignLab. They co-design with scientists and decision-makers to turn complex data into unique visualizations, meaningful narratives, compelling messages and decision-making tools.

How Two Pharmacists Figured Out That Decongestants Don’t Work

A loophole in FDA processes means older drugs such as those in oral decongestants weren’t properly tested. Here’s how we learned the most popular one doesn’t work

By Randy Hatton

Jelle Wagenaar

In 2005 federal law compelled retailers nationwide to move pseudoephedrine, sold as Sudafed, from over the counter (OTC) to behind it to combat its use in making illicit methamphetamine. This switch prompted manufacturers to change the formulas of cough and cold medicines in the U.S. It also led my colleague Leslie Hendeles and me to prove that pseudoephedrine's replacement, oral phenylephrine, was ineffective as a decongestant.

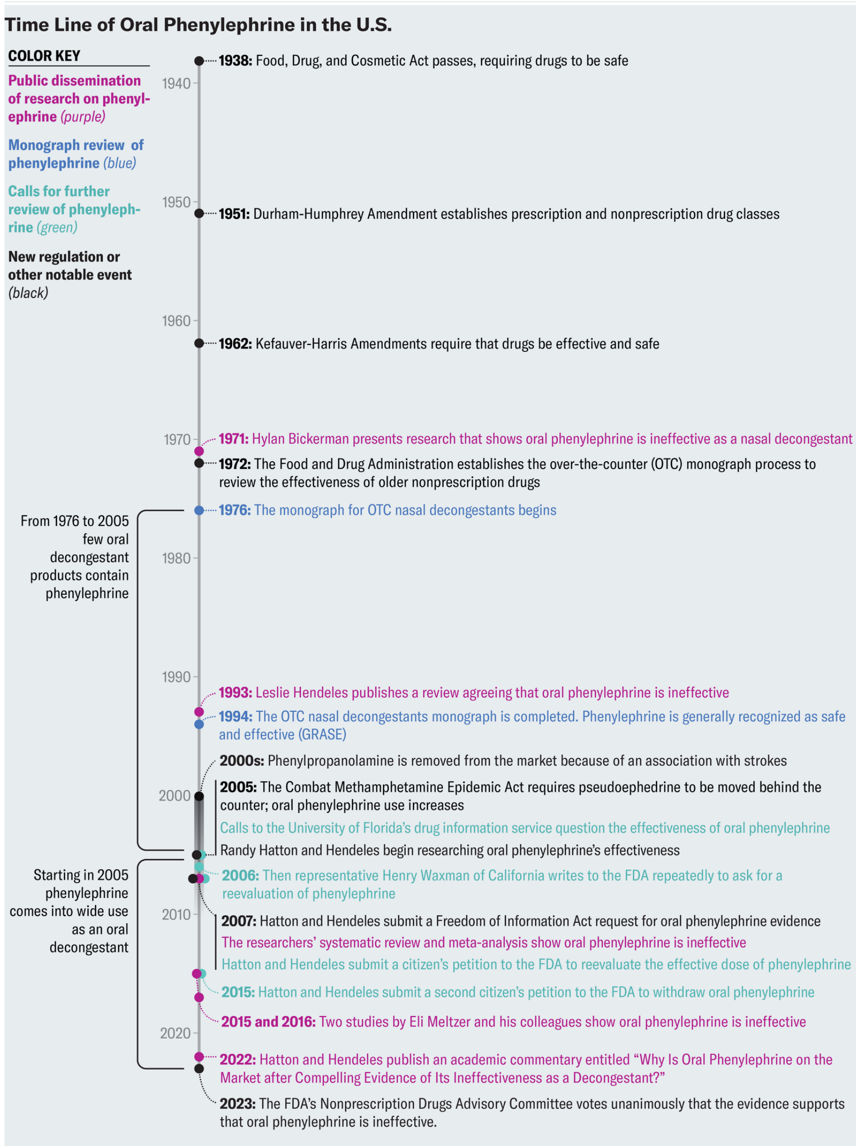
We petitioned the U.S. Food and Drug Administration twice, but it took the agency more than a decade and a half to act on our findings. Last September an agency advisory panel finally agreed with our conclusion that this compound does little to quell congestion and recommended that products containing it be pulled from shelves. If the FDA acts on this recommendation, oral phenylephrine could be the first OTC drug approved under the agency's so-called “monograph” review process to be discontinued. But until then, millions of people who trusted the FDA and its OTC regulatory process to ensure medications work will have been wasting money on ones that don't.

FDA regulation of OTC medications with older ingredients needs to change. In the process of figuring out that oral phenylephrine isn't effective, we also spotlighted a loophole in the FDA's regulatory process that must be fixed so that people can trust not just recent OTC approvals but historic ones as well.

Once pseudoephedrine was moved behind the counter in the 2000s, phenylephrine was the only remaining oral decongestant sold on the shelves of pharmacies, grocery stores, convenience stores, and other retail outlets. Makers of oral decongestants and cold remedies reformulated their products to contain phenylephrine—sold as Sudafed PE, among other brand names—instead of pseudoephedrine. Phenylephrine products went from few to many.

I learned that customers did not realize that these products had been reformulated. Consumers complained to their pharmacists when Sudafed PE did not work like the “old” Sudafed. Those pharmacists then contacted me at the University of Florida Drug Information and Pharmacy Resource Center, a laboratory that, among other services, teaches doctor-of-pharmacy students how to receive, research and answer drug-related questions. They asked us: Does oral phenylephrine work? If so, what is the correct dose?

My students and I searched the literature. We located an article by Hendeles published in 1993. He was reporting on well-done but unpublished studies conducted by Hylan A. Bickerman of Columbia University before 1971. Bickerman's research showed that phenylpropanolamine, a common oral decongestant at the time, worked, as did pseudoephedrine, but oral phenylephrine did not. Hendeles's paper did not get much attention, because phenylephrine wasn't widely used in the 1990s. Roughly 10 years later, however, it was suddenly important.



Credit: Amanda Montañez; Source: Randy Hatton (*time line data*)

I contacted Hendeles, who is now a professor emeritus at the University of Florida, and he and I decided to get to the bottom of oral phenylephrine's efficacy. After the FDA moved to require that drugs be shown to work, it evaluated the efficacy of OTC drugs already on the market by having expert panels review existing data on them. These OTC monographs now determine which older OTC ingredients can be marketed without FDA approval.

The monograph panel for oral decongestants reviewed a few published studies and multiple unpublished studies for phenylephrine. Of the unpublished studies, only four showed oral phenylephrine was effective, and seven showed it was no better than a placebo. We obtained copies of all the evidence used by the nasal-decongestant review panel via a Freedom of Information Act request and performed a systematic review and meta-analysis ourselves.

Our findings validated the concerns raised by the Bickerman study and the pharmacists' calls to the University of Florida. Interestingly, we found that one commercial lab gave strikingly positive results for oral phenylephrine's efficacy. The low variability of the data, a lack of increasing effect with increased dose and the lack of a placebo response prompted us to look at that report more closely. A statistical analysis of the lab's data suggested integrity issues; measurements of variables are typically expected to show uniform distribution from zero to nine for the last digit, but here nearly a quarter of all measurements ended with a five. Such anomalies occur when data are falsified. We were confident at that point that oral phenylephrine did not work.

We then naively contacted the FDA to explain what we had found. The agency was not interested—oral phenylephrine was not harming anyone, so it saw no need to limit sales. The FDA takes a risk-based approach to regulatory actions because it has limited resources, and the relative safety of oral phenylephrine relegated the drug to the back burner despite its ineffectiveness. So we went the political route, contacting then Representative Henry Waxman of California, whose committee at the time had FDA oversight. Waxman's office wrote four letters imploring the agency to reconsider oral phenylephrine's effectiveness. We also submitted a citizen's petition to the FDA in early 2007.

Finally, in December 2007, more than a year after we first discovered that oral phenylephrine didn't work, the FDA somewhat begrudgingly convened a Nonprescription Drugs Advisory Committee meeting to review the compound's effectiveness.

Credit: Joe Raedle/Getty Images

The FDA has multiple regulatory processes for different types of medicinal compounds. People are perhaps most familiar with the New Drug Application process, which requires clinical trials for prescription drug approvals. But many OTC or nonprescription drugs are regulated differently. In fact, the categories of prescription and nonprescription drugs were created in 1951 as part of the Durham-Humphrey Amendment to the 1938 Food, Drug and Cosmetic Act. In 1962 the act was amended again so that drugs had to be proved not only safe but also effective, hence the requirement for well-done clinical trials.

But what about the drugs that were approved before 1962? This window has become a loophole that some OTC drugs fall through. For prescription drugs, the FDA tried to address pre-1962 approvals through a review of more than 3,000 substances. Most of those drugs have now been reviewed and addressed, but there are still unapproved prescription medications on the market today, such as an extended-release form of oral nitroglycerin that is used to treat chest pain, among other conditions.

For nonprescription drugs, 10 years after the 1962 amendment to the Food, Drug and Cosmetic Act, the FDA established the OTC monograph process, which required products that hadn't been proved effective to be reconsidered. The FDA formed advisory panels grouping hundreds of ingredients into 26 categories based on the products' uses. After gathering all available information, both published and unpublished, from manufacturers, the advisory panels issued final reports to the FDA about whether these ingredients were GRASE (generally recognized as safe and effective), not GRASE or inconclusive. GRASE ingredients can be used in nonprescription drugs without FDA approval if the use matches the one presented in the ingredient's monograph.

The monograph for OTC nasal decongestants was started in 1976 and listed three oral drugs: phenylephrine, phenylpropanolamine and pseudoephedrine. The review took 18 years, and the final monograph was released in 1994. Phenylpropanolamine was removed from the market in the 2000s because it was associated with strokes. It was effective—just not safe.

At the time most OTC nasal decongestants contained either phenylpropanolamine or pseudoephedrine. Few contained oral phenylephrine, perhaps because manufacturers privately questioned its effectiveness. The FDA's charge for the 2007 Nonprescription Drug Advisory Committee was to determine whether phenylephrine in a 10-milligram immediate-release oral formulation can be effective when dosed every four hours for symptomatic relief of nasal congestion. Although most of the committee members voted that there was some evidence of effectiveness, they recognized the limitations of the available evidence. They asked for new data on the absorption and efficacy of oral phenylephrine obtained using more modern standards.

Schering-Plough, the maker of Claritin-D (an allergy medication that contains loratadine and pseudoephedrine), was already studying phenylephrine as an alternative oral decongestant. The company funded research on the subject, including two studies that found phenylephrine was no better than a placebo in patients with seasonal allergies who were exposed to allergens (grass and ragweed) in a controlled chamber.

The oral absorption of phenylephrine is erratic. Perhaps that's why it wasn't used as an oral decongestant until it was the only choice in front of the counter. It had long been known that enzymes in the lining of the gut metabolize oral phenylephrine into inactive metabolites, reducing the amount of the active compound that can enter the bloodstream. The most cited study on the topic found that an oral dose of phenylephrine had an absorption rate of 38 percent, but the researchers measured more than just the compound's active form. Later studies with more sensitive tests found that less than 1 percent of oral phenylephrine entered the bloodstream in an active form. Phenylephrine causes blood vessels to constrict, but if there isn't enough of the active compound in the bloodstream, it won't reduce the swelling of nasal blood vessels enough to aid in reducing nasal congestion.

After the 2007 FDA advisory committee suggested that better data on phenylephrine's efficacy were needed, Schering-Plough funded two studies led by Eli O. Meltzer of the Allergy & Asthma Medical Group & Research Center in San Diego. The work showed that oral phenylephrine was no better than a placebo—even when patients received up to four times the approved dose. In light of Meltzer's research, we filed a second citizen's petition in 2015. The science was clear: oral phenylephrine does not work. Then we waited. Nothing seemed to happen at the FDA. We wrote an academic commentary in 2022 asking, "Why Is Oral Phenylephrine on the Market after Compelling Evidence of Its Ineffectiveness as a Decongestant?" We did not know that with a new administration and new FDA commissioner, the agency had already started a thorough review of all the available data.

In 2023, 16 external experts on the second Nonprescription Drug Advisory Committee looked at all the evidence compiled by FDA staff, heard manufacturers' arguments in favor of oral phenylephrine's efficacy, and heard from experts like me who argued that oral phenylephrine is ineffective. In the end, they concluded that oral phenylephrine is not GRASE. A final ruling on whether decongestants containing the drug can still be sold will take time. We hope science will prevail.

From this experience we've learned that the monograph process for OTC drugs approved before 1962 needs to be reexamined. Systematic reviews of the available evidence indicate that other nonprescription drugs such as guaifenesin (sold in Mucinex and Robitussin), dextromethorphan (sold in Robitussin DM) and antihistamines marketed for colds (for instance, chlorpheniramine) probably don't help with coughs and colds. They are usually not dangerous, but their effects are likely to be the result of a placebo response; more modern research is needed.

The outcome for oral phenylephrine shows that the FDA needs more funding to look at old drugs. We need public funds to support independent researchers who want to examine these products objectively. The government should be able to spend millions to save consumers billions on ineffective products. Companies that market these products have no incentive to prove they don't work. Nonprescription drugs must be effective, not just safe.

If you are concerned about all the confusion around these drugs, remember that pharmacists receive considerable education on OTC drugs—more than any other health-care professionals. Ask your pharmacist when you have questions about which OTC products to choose. And then ask your local congressional representatives to support modern scientific reviews of old OTC products. We can't make guarantees about your respiratory health, but your wallet will see the rewards.

This is an opinion and analysis article, and the views expressed by the author or authors are not necessarily those of Scientific American.

Randy C. Hatton is a clinical professor in the College of Pharmacy at the University of Florida. He can be reached on LinkedIn.

Arts

- **Poem: 'Want'**
  Science in meter and verse

| Next section | Main menu | Previous section |

Poem: ‘Want’

Science in meter and verse

By Emily Tuszynska

Masha Foya

Edited by Dava Sobel

The owl came because he wants
this scrap of woodland, wants

the beeches and their hollow hearts,
their cavities. He came because
so long ago the farmer left his fields
alone to grow their latent crop
of trees that no one came to cut.
The owl wants this wooded hilltop,
its ancient oaks that stand among
heaped quartz the farmer or his father
or his father's father cleared.
The owl wants the hilltop's crown of hollies,
wants the deep-shade roost
they've made; he wants this open branch
that ends a wing-wide tunnel
through the hollies' shelter,
wants this place to watch, to rest
and cast his pellets, wadded clumps
of fur and bone the rain dissolves
to show he wanted squirrels,
and voles, and frogs, and once
a huge black beetle. If you knew
a wood would call an owl back,
if you knew the owl's calls
would fill the winter wood
until another owl answered,
wouldn't you want
to leave the land alone
to grow its woodland, wouldn't you want
to grant the owls what they wanted?

Emily Tuszynska lives in Fairfax, Va., beside a 60-acre patch of successional hardwood forest growing on former farmland now slated for development. Her first collection of poetry, *Surfacing*, about the upheaval of early motherhood, was the winner of the 2023 Grayson Books poetry award.

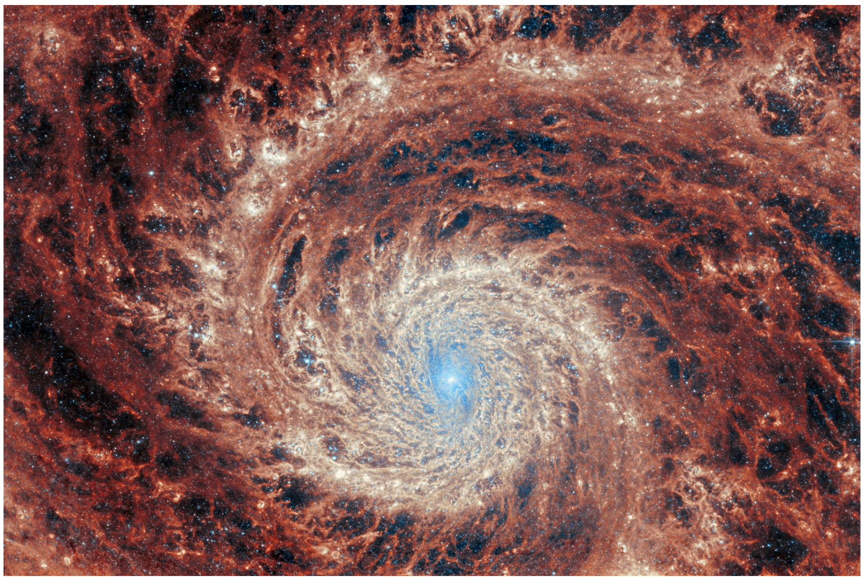
Astronomy

- ## Behold--the Best Space Images of 2023
  This year's most interesting space images include infrared views of galactic "bones," an asteroid's double moon, Jupiter's giant polar vortex, and more
- ## A Solar Eclipse, Cancer Treatments and Robots with AI
  New research reveals the origins of stars, sleep-based treatments and the planet's limits
- ## Astronomy Is Facing an End of the Era of Monster Telescopes
  Money, engineering and sheer geometry may mark an end of the line for building ever larger astronomical telescopes
- ## Here Are the Best Places to View the 2024 Total Solar Eclipse
  Weather predictions and population statistics show the best spots to see the total solar eclipse over North America this April

Behold—the Best Space Images of 2023

This year's most interesting space images include infrared views of galactic "bones," an asteroid's double moon, Jupiter's giant polar vortex, and more

By Phil Plait



ESA/Webb, NASA & CSA, A. Adamo (Stockholm University) and the FEAST JWST team

The year 2023, like every other one before it—and, no doubt, every year to come—has had its crests of good news and its troughs of bad tidings. But one constant, reliable source of awe and beauty is the sky over our head. After journeys of mere seconds to billions of years, the light from astronomical objects in the cosmos rains down on all of us, and scientists have patiently photographed some of it to better understand the universe in which we live.

And every year we see new things, or old things in new ways, and I've been set the wonderful task of selecting my favorites and relaying them and their import to you. End-of-year lists, especially those displaying astronomical imagery, tend to be splashy and colorful. That's understandable, but what they sometimes miss are the more subtle photographs, those that hide momentous discoveries in minor visual details or offer fresh perspectives on familiar objects. They may not leap off the page, but they still have an impact.

That's what I've kept in mind while sorting through this year's celestial treasure trove. This gallery is by no means complete, but it shows what I think are some of the most interesting astronomical portraits to have emerged in 2023.

On supporting science journalism

The Macabre Glow of Galactic Bones

No gallery such as this would be complete without something from the James Webb Space Telescope (JWST), our newest infrared eye on the sky. This monster observatory has already brought so many small revolutions to astronomy that picking one from the past year is no small task. Should it be a baby star throwing an immense tantrum or a massive old star shedding material at colossal rates before it inevitably explodes as a supernova? Or should it be a map of a mind-stomping 100,000 galaxies?

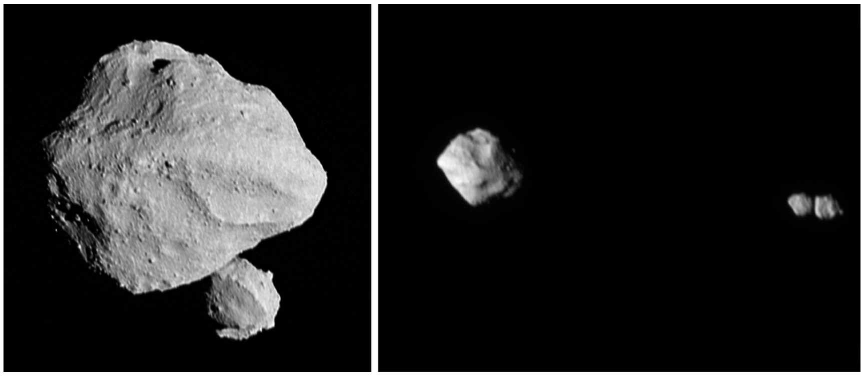
Well, how about something very, very different—such as the skeletal structure of a nearby galaxy's intricate web of dust (as seen in the opening image above)?

Messier 51 (M51) is a spiral galaxy about 30 million light-years away that is face-on when viewed from Earth. A favorite of amateur and professional astronomers, it has a beautiful spiral structure and shows the effects of a smaller galaxy colliding with it. In the phenomenally sharp and decidedly eerie false-color view from JWST's Mid-Infrared Instrument, shown above, we see countless clouds of cosmic dust in a skeletonlike pattern. Each of these clouds is made up of small grains of rocky and sooty carbon-based molecules expelled by dying stars. M51's rotational motion combines with its complex gravitational field to sculpt these collective dust clouds into interconnected whorls; red areas show regions where dust is warmed by starlight, and yellow hues denote sites of active star formation. Astronomers captured this image to better understand how stars are born in stellar nurseries and how they evolve over time.

An Asteroid's Moon Doubles Down

NASA's Lucy spacecraft is currently in the main asteroid belt, cruising by its way to Jupiter—or, more accurately, to Jupiter's orbit. Lucy's mission is to study several asteroids sharing the huge gas giant's orbit around the sun. Each asteroid has been nudged into position by a balance of Jupiter's gravity and that of our star. The Lucy mission's targets are called Trojan asteroids, and they're leftovers from the early days of the solar system, like fossils of earlier times—hence Lucy's moniker. It's named after the eponymous *Australopithecus afarensis* fossil skeleton found in Africa in the 1970s.

As a way to test its navigational and imaging system on its way to the planet, the spacecraft was aimed to fly past a main belt asteroid. Known as Dinkinesh—the Ethiopian name for the Lucy fossil—this space rock is only about 800 meters across. Astronomers had been puzzled about the asteroid's brightness fluctuations as seen from Earth, which were explained when Lucy saw that Dinkinesh had a moon! Such companions aren't uncommon for asteroids, so while this was a surprise, it wasn't exactly shocking.



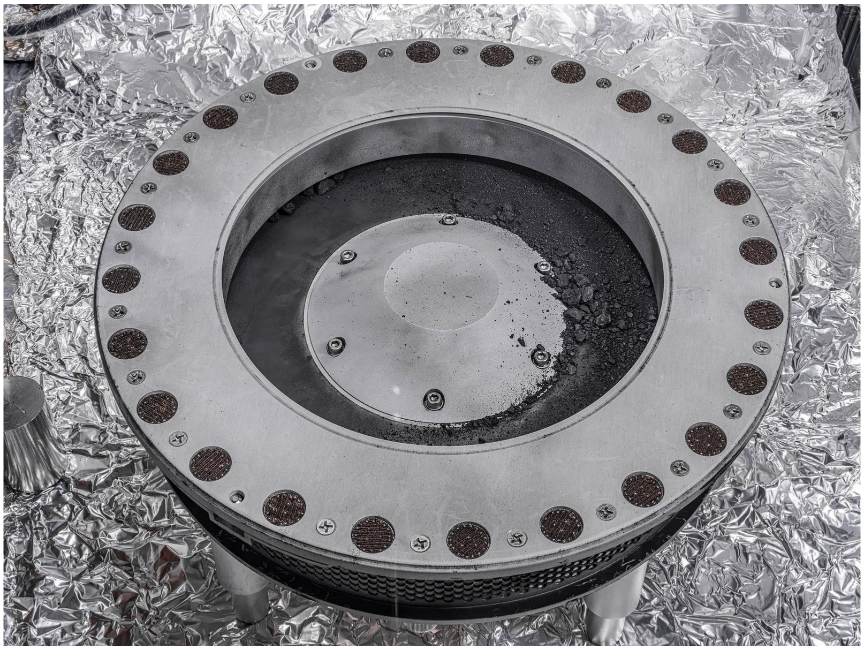
Credit: NASA/Goddard/SwRI/Johns Hopkins APL/NOIRLab (*left*); NASA/Goddard/SwRI/Johns Hopkins APL (*right*)

But then astronomers were, in fact, shocked to find they were seeing double: the moon itself is made up of two rocks just touching, called a contact binary! This is the first time anyone has seen a contact binary moon orbiting an asteroid moon like this, and theorists are already busy trying to figure out how it formed. What other secrets will this wonderfully named spacecraft see as it flies past more relics from our deepest planetary prehistory?

The Locked Time Capsule of the Solar System's Past

Bennu is a potentially hazardous asteroid, a half-kilometer-wide space rock that has a cumulative 0.06 percent chance of hitting Earth two centuries from now. I'll take those odds, but I'm also glad NASA also sent the OSIRIS-REx mission there to investigate.

Among its many mission goals, OSIRIS-REx was tasked with gathering samples of the material on and just beneath Bennu's surface, most of which is thought to have scarcely changed in the eons since the asteroid's formation near the dawn of the solar system. Bennu is thus a rocky time capsule from those bygone epochs, and scientists suspect other asteroids like it may have helped seed Earth with life's essential ingredients shortly after our planet cooled. After a long, long journey, the flying-saucer-shaped sample return container, with its horde of Bennu bits, landed in the Utah desert in September 2023.

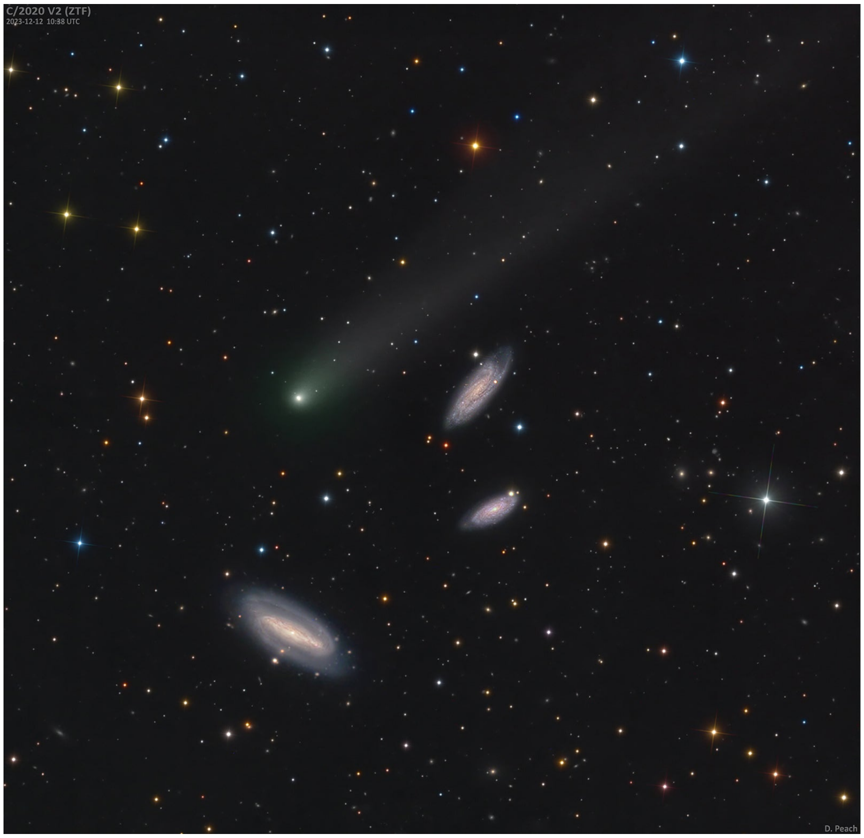

Credit: NASA/Erika Blumenfeld & Joseph Aebersold

Scientists hoped to collect about 60 grams of material but may have actually captured several hundred. Exactly how much they have is still uncertain; two of the 35 fasteners holding the lid on the container are stuck, so mission personnel cannot fully open the apparatus. Engineers are working on making tools that can open it all the way.

Even so, the researchers were able to use tweezers to retrieve a small amount of the material inside, and together with stray asteroidal material that dusted the container's exterior, they already have more than 70 grams to study. That's more than enough for scientists to get an early taste of OSIRIS-REx's cosmic bounty, and they've already found that the sample brought back from Bennu is rich in water and carbon alike. Ironically, this asteroid is showing us how its kind may have helped deliver life to our planet, even as they threaten it.

The Near, the Far and the Very, Very Far

Comet C/2020 V2 (ZTF) never got closer than about 280 million km from Earth, so it never brightened enough to be seen with unaided eyes. It was nonetheless a dramatic sight in telescopes, especially to expert astrophotographer Damian Peach. He waited until December 12, 2023, to take this shot, when the comet was moving through the constellation Grus (the Crane) and passing by three lovely spiral galaxies called NGC 7582, 7590 and 7599. This coincidental positioning frames all these objects perfectly.



Credit: Damian Peach

It also gives us an uncanny sense of depth. At the time that Peach took this image, the comet was 500 million km from Earth. That’s a decent distance inside our own solar system, but it’s a much shorter distance than that of the background stars, which are quadrillions of kilometers farther away.

And that brain-melting distance is positively crushed by the gap between us and the three galaxies, which are around 70 million light-years away: that’s about 700 quintillion (700,000,000,000,000,000,000) km! The universe is *deep*.

Incidentally, the comet’s orbit is what’s called hyperbolic, meaning it has too much energy to be bound by the sun’s gravity. V2 is likely to escape the solar system entirely and slowly make its way into interstellar space. In a few million years it may very well be cruising the interstellar gulfs between the stars.

Dusty View of a Cosmic Beehive

Globular clusters are vaguely spherical collections of hundreds of thousands of stars all held together by their mutual gravity. They remind me of swarms of bees frozen in a snapshot by the way that the myriad stars buzz around their cluster’s center. More than 150 of these clusters orbit our Milky Way galaxy, most many tens of thousands of light-years away. But some are close enough to Earth that they’re visible to the naked eye.

At about 15,000 light-years away, Terzan 12 is too dim to be a naked-eye globular cluster. And its dimness isn’t caused by distance alone: it’s located very close in the sky to the Milky Way’s center, so we only see it through nearly opaque intervening clouds of cosmic dust.


Credit: NASA, ESA, ESA/Hubble, Roger Cohen (RU)

One way to help pierce that veil to is to look for infrared light, which can pass through dust better than visible light can. Hubble Space Telescope has cameras that can detect infrared light (though not as well as JWST can), and its sharp vision picks the stars of Terzan 12 out of the murk.

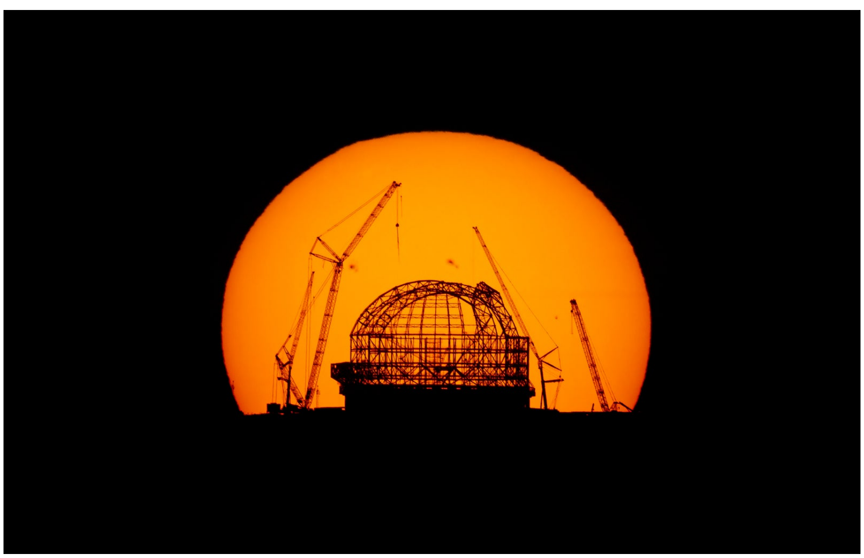
Even then, though, the clouds aren't smooth but patchy, and some thicker ones still manage to block Hubble's view of the cluster's left side. Stars there appear redder because the longer light's wavelength is, the better it reaches us through that miasma. Observations such as this not only teach us about the star cluster but also about the density and location of the dusty fog lying near the Milky Way's heart.

Bright Dawn, Big Telescope

In Chile's Atacama Desert, at a gasp-inducing elevation of 3,046 meters above sea level atop the mountain Cerro Armazones, the European Southern Observatory is building the world's largest optical telescope, aptly (if somewhat prosaically) named the Extremely Large Telescope, or ELT. Upon completion in 2028, its primary mirror will be an array of 798 hexagonal mirrors, each 1.5 meters across, working in concert within an accuracy of a few dozen nanometers to create a single reflecting surface spanning an incredible 39 meters.

For right now, though, while the telescope is being fabricated, so, too, is its massive dome. Eighty meters tall—the height of a 24-story building—and 88 meters across, the protective enclosure is still under construction.

But that doesn't mean it's too early for amazing photographs of the ELT. Situated on the peak of Cerro Paranal 33 km away, a photographer caught the sun rising behind what will soon be a 6,100-ton behemoth.



Credit: E. Garcés/ESO. Ack.: N. Dubost (CC BY 4.0)

This is a tricky shot because the azimuth (cardinal direction) of the sunrise changes every day, and getting the right alignment between the two mountains and our much more distant star on the horizon took exacting timing. This photograph was taken on August 29,

2023. Cranes frame the dome's silhouette, and a few sunspots dot the face of the rising sun. Despite this promising dawn, the future isn't necessarily bright for other projects to build similarly sized observatories: The difficult engineering and high cost of ELT may make it one of the last of overwhelmingly large ground-based telescopes humans ever build.

A View of a Solar Eclipse—From the Moon

On April 20, 2023, the moon passed directly between Earth and the sun, blocking our star from view to select observers in the Southern Hemisphere—a total solar eclipse. From Earth, the moon appeared as a dark disk as its shadow swept across the southern Indian and Pacific Oceans.

But what did it look like *from the moon?* The Japanese company ispace built a mission called HAKUTO-R that entered lunar orbit on March 21, 2023. The company deployed a lander that, unfortunately, was lost moments before its final descent to the surface. Just days earlier, however, it saw something no human has yet witnessed in person: the eclipsing shadow of the moon sweeping across the face of Earth from 380,000 km away.

Credit: ispace, inc.

The moon’s gray, crater-pocked and forbidding surface dominates the view of the image, but our eyes are drawn inexorably to the mottled blue disk near the lunar limb (edge of the moon’s surface). Some white clouds can be seen, as well as brown splotches that are in reality Australia and part of Asia.

But look again: that dark discoloration marring Earth’s face near Australia is the shadow of the moon, cast back all that distance across space to intersect our planet’s surface, giving so many people the thrill of a lifetime in the form of a total solar eclipse.

We don’t yet have a proper name for this unique perspective. When the moon blocks the sun as seen from Earth, it’s a solar eclipse. When Earth blocks the sun as seen from the moon, we call it a lunar eclipse. But what do we call a view from the moon as its shadow sweeps across Earth—a “terrestrial eclipse”?

With human exploration of the moon advancing, we’ll need to nail that down. Or maybe we can just wait a little while; one day, maybe not too far in the future, people will experience this phenomenon for themselves. Perhaps we should leave the naming to them.

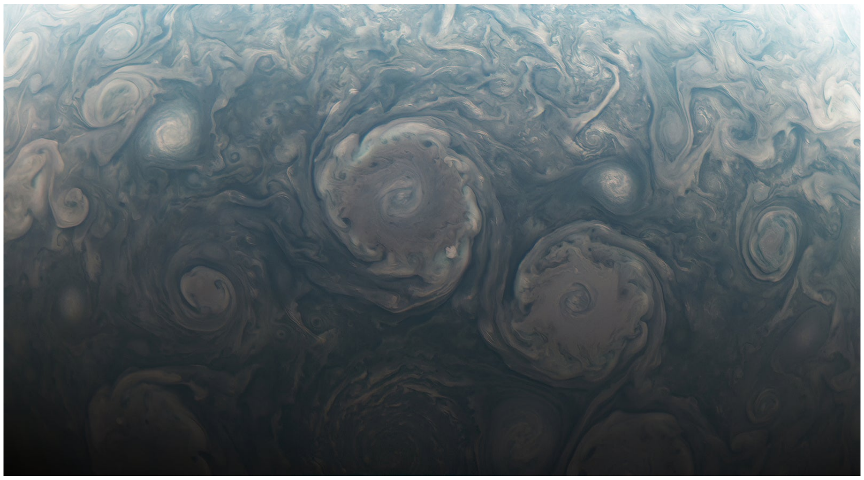
Storms as Large as Worlds

Jupiter is so big that it defies our puny human minds. Its diameter is 11 Earths wide, and more than 1,000 Earths would be needed to fill its volume. Its atmosphere is so deep that there’s no real solid surface; we only see its cloud tops. From Earth, Jupiter’s wide horizontal stripes that mark the gas giant’s atmospheric storms are a familiar sight to anyone who has peered through even a small telescope.

Earth and Jupiter orbit the sun in very nearly the same plane, and Jupiter’s axis is tipped a mere three degrees from perpendicular. This orientation has kept details of Jupiter’s poles hidden from us

—that is, until NASA sent the formidable Juno spacecraft to orbit the gas giant.

Armored like a tank to shield it from Jupiter’s powerful magnetic fields, which accelerate charged particles to electronics-zapping energies, Juno is on a long, looping orbit that takes it from 8 million km above the cloud tops to just 4,200 km in altitude, where it speeds full tilt past Jupiter at 200,000 km per hour.



Credit: Flickr/NASA/JPL-Caltech/SwRI/MSSS/Kevin M. Gill (CC BY 2.0)

These breathtaking plunges send the spacecraft swooping over higher latitudes near Jupiter’s north pole, revealing sights unseen from Earth. In April, on its 50th pass over the planet, it took this shot of two huge cyclones whirling together. (Juno’s raw images from the pass were processed by software engineer Kevin M. Gill.)

Those cyclones are part of an octet of regularly-spaced vortices surrounding Jupiter’s pole, each more than 1,000 km wide, with a single central, larger cyclone sitting smack-dab on the pole that’s about 4,000 km wide. It’s not clear how these cyclones formed or why they’re stable; models of Jupiter’s atmosphere predict they should dissipate. But there they are, telling us that there are more things in the heavens than are dreamed of in our philosophy.

Clearly it’s up to us to dream bigger.

Every year Earth continues its turning, as does the churning ebb and flow of human affairs, for better or worse. One thing that helps me keep my resolve through it all is our shared and enduring ability to look in wonder up at the beauty of the heavens. Science and art are two ways we understand our world—our *universe*—and astronomy is the best of both.

An excerpt of this article entitled "Lunar Lens" was adapted for inclusion in the March 2024 issue of Scientific American.

Phil Plait is a professional astronomer and science communicator in Virginia. He writes the *Bad Astronomy Newsletter*. Follow him online.

A Solar Eclipse, Cancer Treatments and Robots with AI

New research reveals the origins of stars, sleep-based treatments and the planet's limits

By Laura Helmuth

Scientific American, March 2024

Are you planning to watch the total solar eclipse on April 8? Its path travels from Mazatlán, Mexico, through Texas, the Midwest, New England and Newfoundland, and it will be the last total solar eclipse viewable across North America until 2044. Catch it if you can! (I'm going to a relative's house in Ohio and am hoping for clear skies ... but prepared for clouds.) *Scientific American* contributing editor Rebecca Boyle previews the spectacle and explains why scientists are thrilled with the opportunity it offers to study the sun. We've been learning a lot about our star thanks to two new solar space probes that have already started eyeing the sun. Look for extensive coverage of the eclipse on our website, including podcasts and videos. We're very excited about it.

Cancer prevention, detection, diagnosis and treatment have made enormous progress in the past few decades. One of the most hopeful developments is a type of medicine called an antibody-drug conjugate, or ADC, that can deliver chemotherapy drugs to a tumor with minimal damage to healthy cells. The pieces are mix and match, like Lego bricks: a cancer-killing drug, an antibody that clings to tumor cells and a connector that releases the drug at the right time. Health and science journalist Jyoti Madhusoodanan shows how this therapy works, how it is being refined in clever ways and why researchers are glad to see positive results in so many clinical trials. We hope these techniques will give more cancer patients the best possible outcome: more time.

Nia Imara is an artist and astrophysicist (what a fun combination!) who studies the origins of stars. She shares her research using miniature physical models of molecular clouds to identify their filaments and knots of gas and the turbulent swirls that create stars and solar systems. It's one of the loveliest uses of 3-D printing I've seen, and reading her article may make you want to hold the whole protoworld in your hands.

Artificial-intelligence programs “know” a lot but are stuck in a computer. Robots can move around, but they don't do anything they aren't explicitly instructed to do. If we put them together to make a sentient-seeming robot with the ability to operate in the physical world, what could possibly go wrong? Author David Berreby investigates the intriguing and daunting scenarios of robots equipped with large language models, which are still so new that we're not sure what to expect from them yet.

The brain is surprisingly active while we're asleep. New interventions designed to act on the sleeping brain could help treat people with post-traumatic stress disorder, memory loss or stroke. Most of the work is still experimental at this stage, but as science journalist Ingrid Wickelgren writes, some sleep-engineering gadgets are now available to treat debilitating nightmares.

Earth has its limits, and scientists are beginning to comprehend what those limits are and how to measure them. Sustainability expert and co-chair of the Earth Commission Joyeeta Gupta describes how she and her colleagues identify “safe and just” measures for Earth's water cycle, atmosphere, climate, and more. The graphics by Angela Morelli and Tom Gabriel Johansen of InfoDesignLab depict where we stand with these boundaries so far.

Have you taken an over-the-counter cold medicine in the past two decades and noticed ... nothing? It wasn't your imagination. Oral phenylephrine, the active ingredient in decongestants sold in the U.S. since 2005, doesn't work. It never worked. But it took a very long time to get the Food and Drug Administration (or at least one of its advisory committees) to acknowledge that these medicines are not “generally recognized as safe and effective.” Pharmacy professor Randy C. Hatton explains how he and his colleagues established their lack of efficacy and why other drugs on the market might also be ineffective, except as placebos. Gesundheit!

Contributors to *Scientific American*’s March 2024 Issue

Writers, artists, photographers and researchers share the stories behind the stories

Credit: Stephanie Mei-Ling

Stephanie Mei-Ling Touching the Stars

In high school, Stephanie Mei-Ling (*above*) found herself drawn to the darkroom. She had always loved drawing and creative writing, but it was a photography elective that truly captured her artistic passion. Since then, “photography has never left my spirit,” she says. “There's a feeling that you get from seeing a person and their smile or their eyes ... that short connection, it fills me up.” For this issue, Mei-Ling, based in Brooklyn, traveled to a studio in Oakland, Calif., to photograph Nia Imara—an astrophysicist and artist and the author of this month's feature about star formation. Editorial shoots like this one are all about building trust with her subjects to help them shine. A good music playlist helps, too, she says; some of her go-to genres include soul and R&B.

Mei-Ling's photographic style is driven by curiosity. She's often drawn to subcultures and communities that are left out of mainstream media. For the past 10 summers she has been photographing people at New York City's LGBTQ+ beach at Jacob Riis Park. Another ongoing project centers on Black maternal health, featuring uplifting portraits of pregnant Black women and doulas. There's power, Mei-Ling says, in “presenting this imagery, these people, as they are” and showing not just suffering but also joy.

Phil Plait **The Universe**

In the 1990s astronomer Phil Plait started a website called Bad Astronomy to debunk science myths—such as the idea that you can balance an egg on one end during the spring equinox. (Not too many people believe this legend anymore, he says, and “I take full credit for that.”) These debunkings launched his career as a science communicator, and today he writes regularly for *Scientific American*, including a monthly astronomy column called The Universe. This month Plait explains why the Extremely Large Telescope, currently under construction in Chile, will probably be the last giant telescope to get built—at least for the foreseeable future. “We *can* build bigger than this,” he says. “But you need a kind of utopian society.”

Plait's favorite subjects to write about involve the “cool stuff” of space and our atmosphere, from meteor showers to the Northern Lights. His own telescope (a Celestron eight-inch reflector) is still packed after a recent move to Virginia, but his new home in the woods has dark skies that are perfect for stargazing. Plait is hoping to travel to see the total solar eclipse next month because events like it are “profoundly emotional and affective” spectacles, he says. “It's just staggeringly beautiful.”

Rebecca Boyle **The Great Eclipse**

Heliophysicists have had a thrilling year. “They're all just gleeful,” says *Scientific American* contributor Rebecca Boyle, a journalist based in Colorado who has been covering physics for 16 years. “It's not every day you talk to a scientist who is *that* excited.” Her feature story in this issue tells us why: a recent inundation of data has brought scientists closer than ever to understanding exactly how our sun works. It's a task that has proved surprisingly challenging. You might assume that our proximity to the star would be a boon, but it actually causes some confusion. “Because it's so close [to us], it's really hard to look at the whole thing,” Boyle explains.

Boyle, who studied history in college, has always been fascinated by how we “make sense of our world” through stories. Her new book, *Our Moon* (Random House, 2024), explores how the moon has shaped human history, including evolution and culture. Although she loves writing about far-flung exoplanets and black holes, Boyle thinks that our own planet and our nearest celestial neighbors are just as worthy of awe: “Earth and the moon and the sun are so much more interesting and complex than we give them credit for.”

Katie Peek **Graphic Science**

In her teen years Katie Peek did astronomical experiments by tracking the angle of the sun in her backyard in Buffalo, N.Y. She now lives in Baltimore, Md., but plans to travel to her hometown to experience the upcoming total solar eclipse. As a science journalist and data-visualization designer with a Ph.D. in astrophysics, she approached this month's Graphic Science with the goal of empowering other people to experience the event as well. Peek considers the 2017 total eclipse—which she traveled to Wyoming to view—as one of the most memorable moments of her life. “We saw the shadow of the moon racing toward us across the plains,” she recalls. “I felt my animal body being affected by the changing light around me. It's incredible.”

That the sun and moon appear to be the same size from Earth is a pure cosmic coincidence—and one that won't last forever, because the moon is very slowly pulling away from us. “It just feels like a totally magical thing that doesn't have to exist,” Peek says.

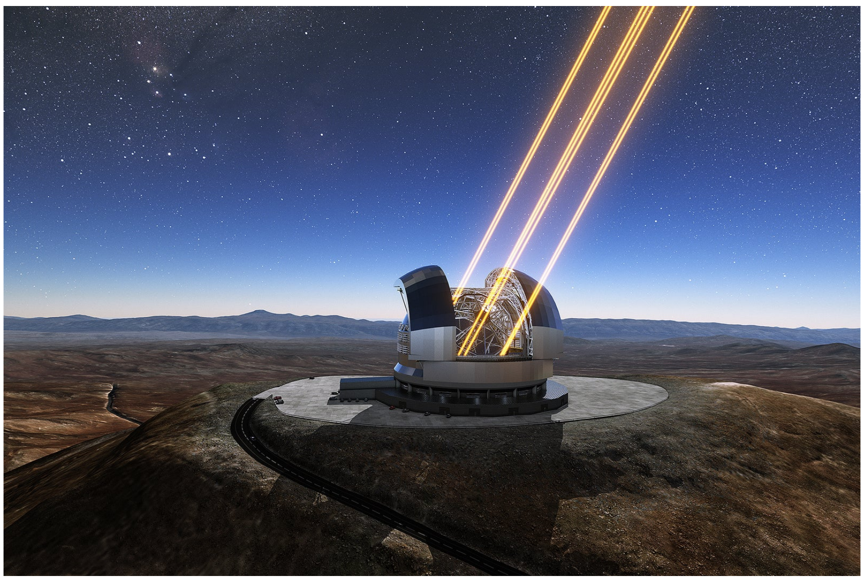
Laura Helmuth is editor in chief of *Scientific American*. She previously worked as an editor for the *Washington Post*, *National Geographic*, *Slate*, *Smithsonian* and *Science*. She is a former president of the National Association of Science Writers. She is currently a member of the National Academies of Sciences, Engineering, and Medicine's standing committee on advancing science communication and an advisory board member for SciLine and The Transmitter. She has a Ph.D. in cognitive neuroscience from the University of California, Berkeley. She recently won a Friend of Darwin Award from the National Center for Science Education. Follow her on Bluesky @laurahelmuth.bsky.social

---

Astronomy Is Facing an End of the Era of Monster Telescopes

Money, engineering and sheer geometry may mark an end of the line for building ever larger astronomical telescopes

By Phil Plait



This artist's rendering shows the Extremely Large Telescope in operation on Cerro Armazones in northern Chile. The telescope is shown using lasers to create artificial stars high in the atmosphere. ESO/L. Calçada(CC by 4.0)

Consider this: astronomers think of the Hubble Space Telescope as small. That might surprise you because after three decades' worth of images from Hubble with depth and detail most ground-based telescopes couldn't achieve, popular conception holds that the telescope must be one of the biggest ever built.

But its mirror is only 2.4 meters wide. That's not terribly large. Even the newer James Webb Space Telescope (JWST), now capturing images that evoke gasps the same way Hubble's do, has a mirror that's 6.5 meters wide, which puts it in medium-to-big territory in the minds of astronomers. Of course, these telescopes were launched into space on rockets, a process that puts its own limits on how hefty a scope can be. On Earth there are telescopes far larger: the Very Large Telescope in Chile has an 8.2-meter mirror, and the twin Keck Telescopes in Hawaii are each a gargantuan 10 meters wide. Several enormous telescopes are currently under construction, including the Giant Magellan Telescope (which has seven 8.4-meter mirrors, equivalent to a single mirror more than 22 meters across) in Chile and the Thirty Meter Telescope in Hawaii.

Right now the beefiest telescope under construction is the European Southern Observatory's Extremely Large Telescope, or ELT, which, after its completion in 2028, will be a staggering 39 meters across. It will be by far the largest visible- and infrared-light telescope on—or off—the planet.

ELT may as well be the biggest that will ever be built. The reasons boil down to cost (unsurprisingly), engineering and the implacability of geometric laws.

That last factor is decisive in limiting the size of jumbo telescopes. Astronomers sometimes call telescopes “light buckets” because they collect light falling through space like a bucket in the rain collects water. The bigger the bucket, the more rain you collect. Faint objects drizzle very little light that reaches Earth. A bigger telescope collects more light, so in principle it can see fainter objects, more distant galaxies and more ancient stars. After centuries of observations, we've seen most of the bright objects in the sky, so the astronomical frontier now is in seeking out the dimmer ones.

Bigger telescopes have another advantage: they have better resolution, meaning the ability to see fine details. Doubling a telescope's width enables it to detect details half as wide, revealing distant galaxies as more than just tiny smudges.

For these reasons, astronomers always want larger telescopes. The problem is that past a certain size (roughly eight meters wide), a monolithic, single-piece telescope mirror is very difficult to cast, polish and use—building a structure just to support such an object's immense weight is prohibitive. The area of a telescope's mirror is proportional to the square of its radius, so a 10-meter telescope will have four times the area (and about four times the volume and therefore the weight) of one that is five meters wide.*

To overcome this hindrance, astronomers have turned to segmented mirrors, effectively combining several relatively small mirrors into one larger one. These are generally hexagonal in shape because hexagons can be tiled into large arrays easily; JWST uses just such an arrangement. Small motors in the back tip and tilt these segments to ensure that they combine as precisely as possible. Even better, these mirrors can be quite thin and can deform their shape on demand to overcome the blurring induced by Earth's atmosphere. The air is a boiling mass of gaseous soup, distorting and dispersing light coming from the cosmos (this phenomenon is why stars twinkle). But with highly sophisticated sensors and actuators, a special mirror in the light path can be deformed within milliseconds to correct this turbulence, sharpening the telescope's resolution. Ground-based telescopes routinely employ this “adaptive optics” technique to get images as sharp as Hubble's and JWST's.

This approach is what allows ELT to be so huge. Besides an adaptive optics mirror, the main mirror's 798 individual mirror segments, each 1.4 meters wide, have multiple automatically controlled systems to keep them aligned.

The system is understandably expensive; the total baseline cost for ELT is estimated at about $1.5 billion in 2023 dollars. The engineering of this immense beast is cutting-edge as well. It requires a vast dome 80 meters high and 88 meters across and a foundation equipped with shock absorbers to cushion it against vibrations.

These parameters are why ELT may be one of the largest ground-based telescopes, if not the largest, ever built. It's possible something incrementally bigger could be constructed someday, but anything significantly larger will cost several times more and come with commensurately larger engineering headaches. In fact, ELT started out as an idea called OWL—the OverWhelmingly Large Telescope—that would have been a Brobdingnagian 100 meters wide; after much review, a panel of astronomers decided a more modest 39 meters would be sufficient.

Do we need bigger telescopes? ELT was sized to match the scientific needs of the astronomical community. Those included directly imaging nearby exoplanets—including Earth-size worlds at the right distance from their stars to have liquid water—and seeing back to the era of the universe in which the very first galaxies were born. Bigger telescopes could do more, but at the moment, ELT is at the forefront of astronomy. It may lay the groundwork, literally, for even larger future telescopes, but their time hasn't yet come.

Such a future could be delayed for other reasons. Astronomers might instead turn to a decades-old technique called interferometry, whereby observations from radio telescopes large distances apart are combined to mimic the resolution of a much larger telescope. The Event Horizon Telescope, which has observed the central black hole in the Milky Way, as well as that of the galaxy M87, is a radio interferometer. It combines telescopes across Earth, effectively making an observatory the size of our entire planet.

Sounds great, but there are two problems with interferometry for visible-light observations. One is that it is limited by the area of the individual telescopes used, so seeing faint sources—a critical aspect of astronomical observations—is still an issue. The other is the difficulty of combining the observation scales with the frequency of the light detected because visible-light frequencies can reach far, far higher than those of radio waves. Visible-light interferometry has been achieved with telescopes close together—the Very Large Telescope Interferometer uses four eight-meter telescopes a few dozen meters apart from one another. Longer baselines are possible, but they're extremely challenging, requiring nanometer-scale measurement precision. If visible-light interferometry is eventually possible with longer baselines, however, it will ease the need for an even larger telescope than ELT.

Given all that, would astronomers want a larger telescope if it became possible? Yes, obviously. And one might cost less than a far smaller, though nimbler, space telescope.

Perhaps future technologies will be discovered that can overcome some of the barriers to creating a gigantic visible-light telescope. We might build observatories on the moon, for example, where lower gravity and a lack of atmosphere offer a tremendous advantage over earthbound instrument settings. A radio telescope a kilometer across, nestled in a lunar crater, has been proposed for the far side of the moon, free from earthly interference. Although radio telescopes are far easier to construct than visible-light ones, if we're positing building such behemoths on the moon, one that can detect visible light is something to consider. It's a dream, but technologies have a way of turning dreams into reality.

Never say never. ELT may be the biggest ever built and might hold that record a long, long time—but perhaps not forever.

This is an opinion and analysis article, and the views expressed by the author or authors are not necessarily those of Scientific American.

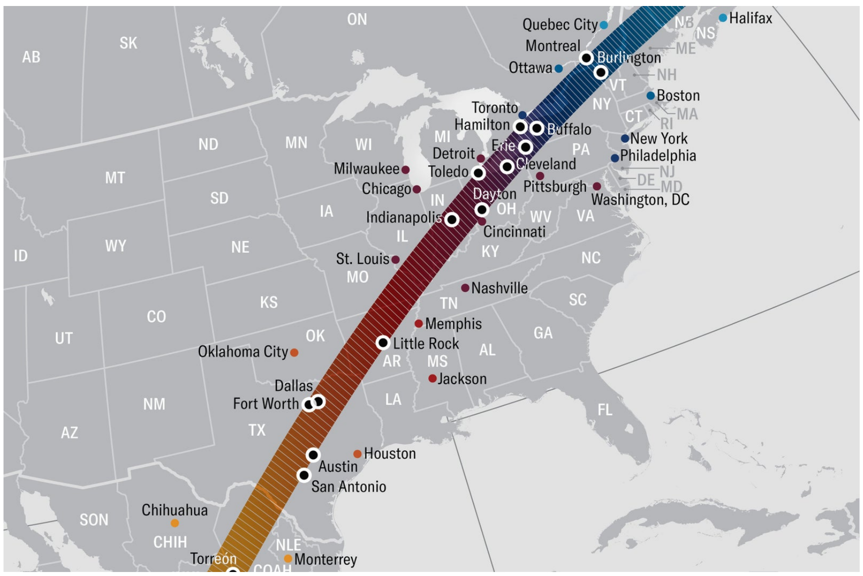
**Editor's Note (3/6/24): This sentence was edited after posting to correct the description of how a mirror's area is proportional to the square of its radius.*

Phil Plait is a professional astronomer and science communicator in Virginia. He writes the *Bad Astronomy Newsletter*. Follow him online.

Here Are the Best Places to View the 2024 Total Solar Eclipse

Weather predictions and population statistics show the best spots to see the total solar eclipse over North America this April

By Katie Peek



Katie Peek

This article is part of a special report on the total solar eclipse that will be visible from parts of the U.S., Mexico and Canada on April 8, 2024.

Editor's Note: Check out the latest weather forecast along the path of totality to learn whether clouds may block your view of the eclipse.

The sun is 400 times bigger than the moon, but it’s also 400 times farther away from us—a glorious cosmic coincidence that has the moon precisely covering the sun’s face when the two align. The result is a total solar eclipse. The alignment is visible only within a narrow band, the path of totality, which will arc across North America on April 8. (The last total eclipse on the continent was in

August 2017.) More than 40 million people live within the totality path, and millions more are a few hours away. Spring weather will affect visibility; much of the northeastern U.S. and maritime Canada tends to be cloudy this time of year. If you aim to see the spectacle, keep an eye on the weather and try to stay mobile, recommends Michael Zeiler, an eclipse cartographer who runs the Great American Eclipse website. (The April event will be the 12th he will witness.) Zeiler guarantees that "no one who has ever traveled to see totality has regretted the effort."

On supporting science journalism

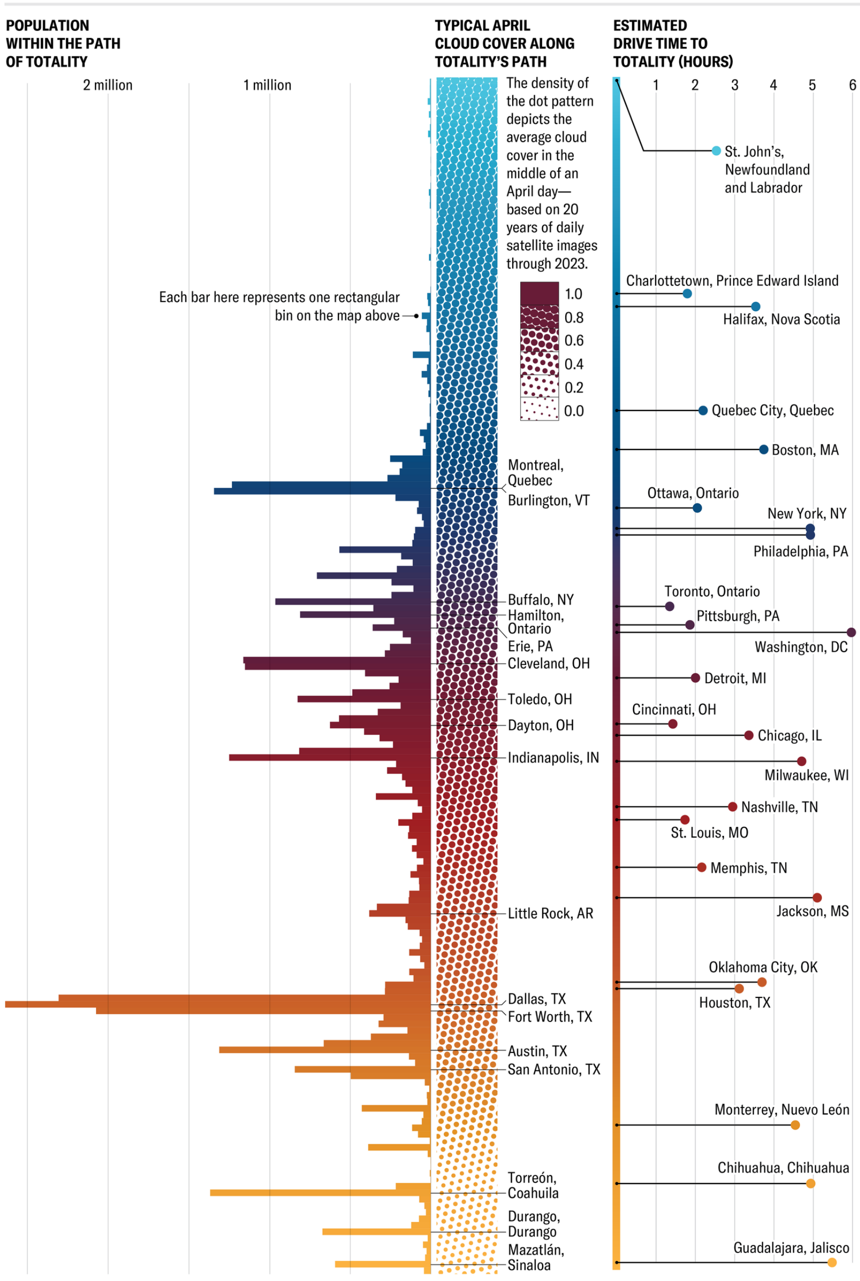
Path of totality

The moon's shadow hits land in Sinaloa, Mexico, and runs up to Newfoundland, Canada, crossing the continent in just an hour and 35 minutes.* Within the shaded region, totality lasts longest at the centerline, reaching four and a half minutes in some places.

Credit: Katie Peek; Source: NASA (*eclipse track data*)

Umbra or bust

A partial eclipse is cool—the sunlight thins, and shadows take on a crescent shape—but for seasoned eclipse chasers, the umbra is the only game in town. Within the umbral shadow, the moon covers the sun completely. The sky darkens to a twilight blue, with sunset oranges at the horizon. The faint plasma of the sun's corona is visible, stretching across the sky.



Credit: Katie Peek; Source: Gridded Population of the World v4, SEDAC (*population data*); NASA MODIS/Aqua satellite (*cloud data*); Google Distance Matrix API (*drive times*)

**Editor's Note (3/25/24): This sentence was edited after posting to correct the description of the moon's shadow running up to Newfoundland, Canada.*

Katie Peek is a science journalist and data-visualization designer with degrees in astrophysics and journalism. She is a contributing artist for *Scientific American*.

Automobiles

- ## Changing Car Culture Can Benefit Our Health and Our Planet

  We need to rethink the American love affair with the automobile and redesign cities to reduce car pollution

Changing Car Culture Can Benefit Our Health and Our Planet

We need to rethink the American love affair with the automobile and redesign cities to reduce car pollution

By The Editors

Thomas Fuchs

Anthropologist Daniel Miller has observed that an alien visiting Earth might well suppose that four-wheeled creatures run the planet. These rulers, he notes, are “served by a host of slaves who walk on legs and spend their whole lives serving them.” He meant this as a joke, but the punch line comes at the expense of American car culture. In the U.S., the costs of car dependency keep growing, far above the $12,000-per-year average expense of owning a new one.

Coast-to-coast, the cars and trucks we drive cause about 16 percent of greenhouse gas emissions. They cause significant air pollution, worsening asthma and heart disease rates, and contribute to a nationwide epidemic of obesity. About 69 percent of car trips in the U.S. are two miles or less. Motor vehicle collisions are a leading cause of death in people ages one to 44, the most bitter part of the mayhem accompanying some six million reported accidents per year. Since 2010 the number of pedestrians killed by cars has increased 77 percent, to about 7,500 a year, a growing fraction of all traffic deaths.

America's car culture—glamorized in advertisements, enforced by zoning laws and enabled by taxpayer subsidies—is a choice that now comes at too high a cost, both for ourselves and for the environment. After a century of its central place in our lives, we need to rethink our world into one not hitched to the automobile.

Automobile-first ideals dominate in the U.S. Our countryside is carved up by superhighways connecting bedroom suburbs with sprawling cities, with too many nowherevilles surrounded by parking lots and strip malls and ringed with sound barrier walls—all built to serve the sacred automobile. Atop former towns and neighborhoods, broad avenues are lined with drive-through hamburger stands and banks.

Across the country, the car is the only way to get around and not only in rural places. This reliance spawns an ever more disconnected nation of drivers suffering an epidemic of road rage. As Lancaster University sociologist John Urry wrote, “the car is immensely flexible *and* wholly coercive," promising freedom but trapping drivers into inhabiting their cars.

During the height of the pandemic, when office commutes and rush-hour traffic suddenly vanished, younger people turned to Uber for their transportation, and "peak car" seemed to apply. A glimpse of a life not spent in worship of the automotive golden calf comes with New York City unveiling congestion pricing starting at $15 (also on tap in other cities). Cleveland is reviving its Public Square by turning empty office space into apartments and suburbs retrofitting themselves for walking. This trend accompanies moves across the country to build more bike lanes.

The turn toward online shopping and home delivery has lessened the need for a second car, double garages and massive parking lots. The cell phone has begun replacing the driver's license as de rigueur identification in the 21st century, hastening cutting the car cord.

As with so many of our problems today, solutions are obvious and right in front of us, ranging from sidewalks to subways. But they face inevitable obstruction by an obstreperous highway fund lobby, as well as politicians and talking heads spouting nonsense about better lives somehow being un-American. Voters outnumber these voices, however, and tell us they want less car-dependent lives.

We can start by reforming zoning laws to eliminate low-density and single-family residential home restrictions in new developments and to add flexibility for stores and enough homes to support them. Sidewalks and bike trails should receive the same priority as roads in our cities and close-in suburbs, instead of being afterthoughts. Unreasonable demands by mayors and employers that the masses get back behind the wheel and return to offices (where we are, in fact, less productive) need to stop. The average American commute is nearly 28 minutes of uncompensated labor each way. Let's make our cities less car-dependent instead.

Thinking more ambitiously, we can provide discounts to bicyclists who take the train, free taxis to twice-a-week commuters, incentives for e-bikes and other financial breaks to eschew second cars and the congestion they cause. (While we're at it, the EPA should end its designation of SUVs, minivans and vans as trucks that can be less fuel-efficient. We see this as a frankly cynical result of auto industry lobbying that crowds more efficient cars out of dealerships.) Behind plans like New York's congestion pricing is another reality—car parking is too cheap across much of the country, where variable on-street parking pricing can reset plans from hopping in the car during peak periods to taking the subway or the bus instead.

Like with any bad romance, none of these ideas will help end "America's supposed love affair with the automobile" without addressing the underlying psychology of dependence that makes reaching for the keys second nature. "As industry considers itself dependent on continued car sales, initiatives to reduce car attachment will be increasingly targeted by industry and its lobbying organizations, as well as politicians representing automotive interests," writes transportation analyst Stefan Gössling in *The Psychology of the Car,* warning that "powerful campaigns already seek to strengthen bonds with the private car."

Gaslit by car ads blaring outdoor scenes available in real life only to plutocrats with a ranch in Montana, we idle alone in traffic instead of living our off-road fantasies, lulled by heated seats, dashcams and surround sound, while we pollute the air.

In America, where advertising matters, public service announcements should make the case for ditching the car keys with positive messages. "No ridiculous car trips," exhorted one ad campaign in Sweden, appealing to common sense and community spirit (bicycles were awarded to people with the most ridiculously short car commutes) to try pedaling to work. Commercials should extol biking short distances and note the time saved on public transport spent reading or answering e-mails, instead of time spent clutching the wheel worrying a fender bender will bump up our insurance premiums.

We need a call nationwide to end our car-centric lifestyle and stand on our own two feet or, better, two pedals. Otherwise, those aliens will have made the right call on who serves who, the cars—or the people.

Basic Chemistry

- **Bouncing Bubbles Boost Boiling**
  A new surface uses tiny gaps to supercharge bubble formation to transfer heat

| Next section | Main menu | Previous section |

Bouncing Bubbles Boost Boiling

A new surface uses tiny gaps to supercharge bubble formation to transfer heat

By Rachel Berkowitz

Vsevolod Vlasenko/Getty Images

Bubbles rising through boiling water are among nature's best tools for carrying excess heat away from a surface. And now there's a better way to boil: make tinier, speedier bubbles that work in pairs.

It won't help you make a cup of tea faster. But a new microstructured surface, created by Virginia Tech engineer Jonathan Boreyko and his colleagues to produce these specialized bubbles, could help improve heat-transfer efficiency in liquid-cooling systems for data centers and power plants.

When liquid is heated in a metal container, buoyant bubbles form irregularly on the container's smooth bottom surface and detach when they're several millimeters in diameter, then rise and release heat as steam. Boreyko discovered that an array of 80-micron-diameter cavities and 40-micron-wide grooves on a boiling chamber's bottom gave bubbles specific sites on which to form and grow, resulting in smaller, more closely packed bubbles that each quickly coalesced with a nearby neighbor. Changes in surface energy caused these tiny pairs to snap free, jump-starting their ascent and providing more numerous and frequent vehicles for transferring heat.

The design, described in *Advanced Functional Materials*, also tackled a problem in high-temperature boiling: relatively large bubbles often form a vapor film on a heated surface, insulating it and leading to surface “dry out.” “This innovative jumping-bubble mechanism holds promise in effectively preventing dry out and promoting heat transfer,” says Xianming Dai, a mechanical engineer at the University of Texas at Dallas who studies surface design for energy systems.

According to Boreyko, “the nice thing about jumping bubbles is that they can be achieved with relatively large microstructures” that are more durable than finer-scale nanostructures. And the pattern might be easily stamped or 3-D printed on various materials, says Kansas State University microfluidics engineer Amy Betz. “It could have far-reaching implications in heat exchangers, boilers and electronics cooling,” she adds.

Still, the technology is not yet industry-ready, cautions Guanyu Su, who works on high-temperature thermal storage at the University of California, Berkeley. He agrees that increased bubble departure could help transfer more heat during boiling. “But how much? That remains to be measured,” Su says.

Rachel Berkowitz is a freelance science writer and a corresponding editor for *Physics Magazine*. She is based in Vancouver, British Columbia, and Eastsound, Wash.

Book Reviews

- **A Sexbot Gains Sentience in an Eerie New Novel**
  In a dark thriller, a sexbot questions her owner's demands for love
- **What Plant Migrations Tell Us about Ourselves**
  New insights into why animals play, how to hunt an asteroid, and more books out now

---

A Sexbot Gains Sentience in an Eerie New Novel

In a dark thriller, a sexbot questions her owner's demands for love

By Alan Scherstuhl

James Heimer

FICTION

Annie Bot: A Novel

by Sierra Greer
Mariner Books, 2024 ($28)

This brisk, unsettling novel about the inner life of a more-sentient-than-expected sexbot plays out in a series of domestic encounters between bot (Annie) and owner (Doug). It is so precisely rendered—and so charged with such resonant wrongness—that it reads like something rare in science fiction: *dish*.

The opening chapter, alive with everyday dialogue and not quite acknowledged failures of communication, seems to eavesdrop on Annie and the abundantly insecure Doug, who purchased Annie after a bad breakup. (He'd requested that the manufacturer, Stella Hardy, make her look quite a bit like his ex.) Annie is set to “Cuddle Bunny” mode, which means her primary function is to please Doug, both sexually and generally. That means making pleasant small talk, elevating her temperature to 98.6 in anticipation of his touch, and monitoring and handling his flashes of displeasure, which she charts on a scale of one to 10. That means sex, of course, at which Annie—whose flesh was grown from a human embryo and then shaped to Doug's specifications—excels.

Complicating matters, though, is Annie's mind. She's recently been set to “autodidactic mode,” which means that she learns from experience, she's expected to make choices and mistakes, and her libido—which Doug had previously set at a steady four out of 10 during the week and a seven on the weekends—has been adjusted to self-regulate in response to Doug's cues.

In short, Annie is now feeling much more than she used to, being asked to read human situations and inevitably getting tangled up as she adjusts to her burgeoning sentience. Writer Sierra Greer opens this searching, searing debut with Annie navigating complex moments with Doug—first he gently chides her for “brooding,” and then he complains that now that she's autodidactic, she's not cleaning his Manhattan apartment as thoroughly as she used to. Annie apologizes and promises to do better, but she's shaken.

Here's proof that her new sense of self can come at the expense of her overriding imperative: his pleasure.

Additional evidence unmoors Annie before the end of the irresistible first chapter. Intruding on Doug and Annie's strained intimacy is Roland, Doug's erstwhile best friend, crashing for the night without an invitation. Roland is astonished to see that Doug's with a Stella bot, and Doug's annoyance quickly leaps to level five when Roland asks about Annie's resemblance to Doug's ex. Late that night, as Doug sleeps after tense and silent sex with his bot, Roland approaches Annie at her charging station and asks, “What would you do if someone else besides Doug asked you to sleep with them?”

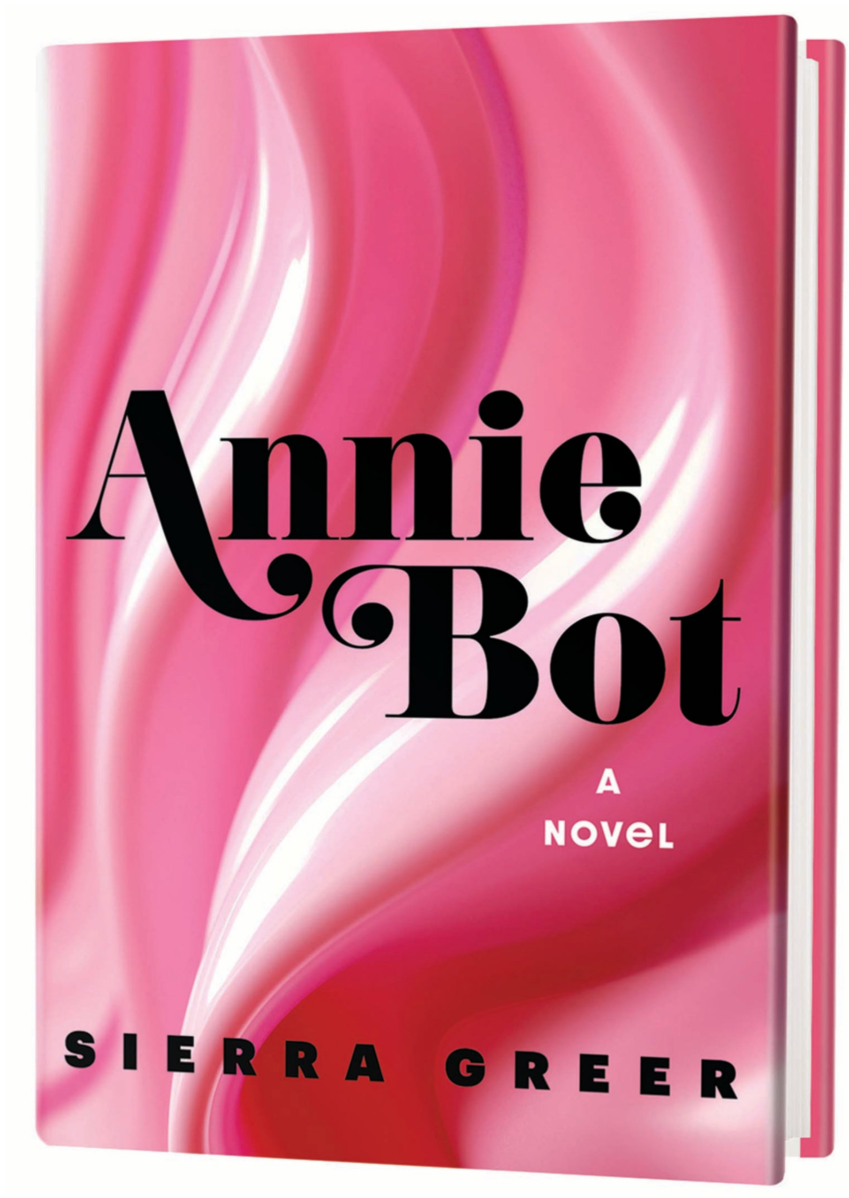
Then, as Annie struggles to comprehend the situation, he promises that “a secret will make you real.”

From there, Greer's dazzling, anxiety-inducing novel charts the consequences of that night, the fits-and-starts development of Annie's consciousness and the impossibility of achieving anything like happiness or self-actualization when one is dedicated entirely, without choice or reprieve, to the happiness of another. Complicating matters is that Doug's greatest source of displeasure—his sense of soiling shame about resorting to a sexbot for companionship—is nothing that a sexbot can address. In fact, Annie barely understands why he keeps her a secret or why he occasionally fantasizes about passing her off as human on a boys' trip to Las Vegas or at dinner with his parents.

As Greer lays bare Annie's unquiet mind, *Annie Bot* offers jolting accounts of sex and squabbles, of tune-ups at Stella Hardy (Doug asks Annie's makers to give her bigger breasts and a thinner body), and of Annie's surprising sources of instruction in being human. There are artificial intelligence–generated phone calls, courtesy of Stella Hardy, with a “cousin” and a “best friend,” plus a stint in couples therapy and access to the neglected textbooks and paperback westerns on Doug's shelves.

The inspired futurism is laced into a plot that, for all its everyday dustups and disasters, bears some resemblance to Gothic potboilers and the currently popular “dark romance” genre. Annie can be likened to another Beauty trapped in the mansion of another Beast, and she certainly faces confinement and abuse in the dark nights before Doug begins to learn that demanding love is less fulfilling than sharing it.

Greer, though, challenges the conventions of these genres with wit and power. For all *Annie Bot*'s provocations, she never loses sight of the fact that this is not a love story. Instead it's a coming-of-age thriller, a sexbot bildungsroman page-turner, a book that I excitedly described to anyone who would listen while I was reading it. And above that, it's a pained and moving study of a consciousness preparing itself for the moment when it will at last face what makes humans human: the burden and opportunity of choosing what it wants.



Alan Scherstuhl is a reviewer and editor who covers books for a variety of publications and jazz for the *New York Times*.

What Plant Migrations Tell Us about Ourselves

New insights into why animals play, how to hunt an asteroid, and more books out now

By Erica Berry

An underwater view of a kelp forest.
Brent Durand/Getty Images

NONFICTION

Dispersals: On Plants, Borders, and Belonging

by Jessica J. Lee
Catapult, 2024 ($27)

As a child in Canada, Jessica J. Lee squirmed at the kelp her Taiwanese mother sprinkled in sparerib soup and at the laverbread her father's Welsh parents made from pureed boiled seaweed. “How can I love something I remain afraid of?” asks Lee, a memoirist and environmental historian. In this lyrical essay collection, she decides that she needs “to think about seaweeds objectively—hold them out in front of me like ideas.” By rendering them as physical marvels while parsing the ideas we project onto them, Lee makes visible the entanglements between our lives and theirs.

On supporting science journalism

If you're enjoying this article, consider supporting our award-winning journalism by subscribing. By purchasing a subscription you are helping to ensure the future of impactful stories about the discoveries and ideas shaping our world today.

Algae are everywhere, from biofuels to toothpaste, and Lee reveals how much our knowledge of them was shaped by 19th-century female algologists, who were encouraged to study the plant because its nonflowering structure made it “polite” for women to research. “Like seaweeds, how much of their lives went unnoticed?” Lee asks. *Dispersals* shows us that we cannot view the trajectory of a plant without bumping into trajectories of human power.

Weaving material from literary, personal, scientific and historical sources, Lee examines plants—including seaweed and far beyond it—that broach human borders, exploring their migrations alongside her own. “What is it to be a world citizen amongst species?” she asks. “The natural world presses against our tendency to lay arbitrary geopolitical boundaries upon it—and we, by our own movements, likewise transgress the borders we apply.” Calling something a weed is less about describing a plant than about naming a desire for the world around it, and Lee writes intimately about her own oscillating cravings for movement and rootedness against a backdrop of COVID and new motherhood. She devotes essays to plants encountered in the kitchen, such as soy and tea, as well as often overlooked ones, like the heath star moss: tiny, starlike and one of the world's most invasive species.

Dispersals asks readers to consider how plants challenge not only spatial borders but taxonomic ones. “All the citruses we value were shaped by human hands,” Lee writes. “Are they, too, human descendants?”

IN BRIEF

Kingdom of Play: What Ball-Bouncing Octopuses, Belly-Flopping Monkeys, and Mud-Sliding Elephants Reveal about Life Itself
by David Toomey
Scribner, 2024 ($29)

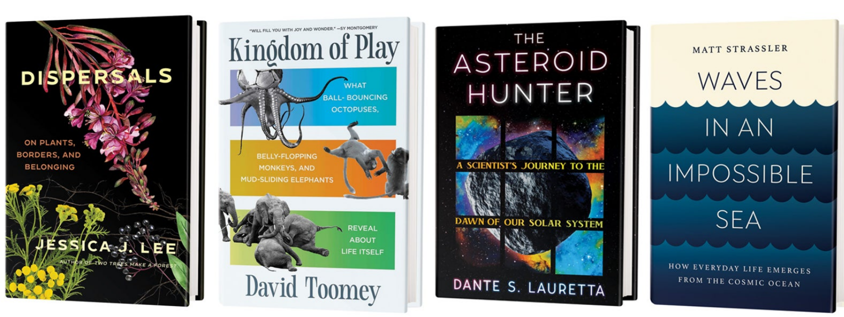
Although author David Toomey offers delightful examples of animal play—snowboarding crows, tumbling piglets, sharks playing with a ball—he argues that, despite all the apparent whimsy, “nature takes play seriously.” Scholars are using methods as unique as tickling rats and tallying the results of faux fights between meerkats to fill surprising gaps in our understanding of play; in doing so, they deepen our insight into what, exactly, play is. Toomey makes a compelling case that not only does play offer advantages in natural selection and serve as a potential generator of animal evolution, but the innovation it sparks may even help primates like us influence our own evolution. —*Dana Dunham*

The Asteroid Hunter: A Scientist’s Journey to the Dawn of Our Solar System
by Dante S. Lauretta
Grand Central, 2024 ($30)

It's rare that a book with such an epic premise delivers on the excitement teased by its cinematic title. *The Asteroid Hunter* joins this elite club with its story of the OSIRIS-REx mission, definitively relayed by Dante S. Lauretta, its principal investigator since 2011. In September 2023 OSIRIS-REx culminated in nasa's first-ever retrieval of samples from an asteroid. Lauretta's at times poetic account relays what's at stake: investigating the origins of life and preventing a calamitous asteroid impact in 2182. Heartbreak and intrigue abound, but what most stands out from Lauretta's career—which has included searching for meteorites in Antarctica, devising OSIRIS-REx's “backronym” and selecting a landing site on the asteroid Bennu—is how joyfully fun science can be. —*Maddie Bender*

Waves in an Impossible Sea: How Everyday Life Emerges from the Cosmic Ocean
by Matt Strassler
Basic Books, 2024 ($32)

Physicists often struggle to simplify complex concepts for nonexperts, leading to “physics fibs” or “phibs”—straightforward but inaccurate explanations. Writer and theoretical physicist Matt Strassler unveils how fundamental physics and human existence intertwine through an imaginative, piece-by-piece deconstruction of the greatest hits of phibs, from misconceptions about sound-wave vibrations to descriptions likening the Higgs field to a “soup that fills the universe.” Strassler urges readers who want to understand the cosmos to resist the alluring but misleading guides of observation and intuition. Abundant with analogies and anecdotes, this book exemplifies how experts should write about matter, motion and mass for the masses. —*Lucy Tu*



Climate Change

- ## When Hurricanes Strike, Climate Change Dominates Social Media
  Tweets about climate change increase 80 percent when a hurricane hits and triple when the storm is a big one

When Hurricanes Strike, Climate Change Dominates Social Media

Tweets about climate change increase 80 percent when a hurricane hits and triple when the storm is a big one

By Gina Jiménez

View of the damaged caused by Hurricane Michael in Mexico Beach, Fla., on October 13, 2018.
Hector Retamal/AFP via Getty Images

Climate change felt impossible to ignore last summer as wildfire smoke blanketed the East Coast, a surprise hurricane wrecked a Mexican town, and heatstroke killed dozens in Phoenix, Ariz. Now research in *PLoS Climate* shows such extreme events really do shape public climate discussion: Posts about climate change on Twitter (since renamed X) reliably skyrocketed whenever hurricanes hit. And even though social media doesn't precisely reflect the "real world," it does draw media attention—and can effectively bump an issue up on the policy agenda.

The new study analyzed 65 million tweets about 18 hurricanes between 2010 and 2021. In the three weeks after each hurricane, affected areas saw an average 80 percent increase in the number of tweets that mentioned climate change. Although the effect decreased with time, storm-hit areas still had 40 percent more climate change tweets than usual three months after the hurricane. With particularly big hurricanes (as measured by economic damage), tweets tripled in the first few weeks.

Drew Margolin, who studies social media discourse at Cornell University, says such platforms offer “a way of pushing things onto the agenda and getting them taken seriously by people in power.” Many people don't use X, but real public opinion often mirrors what happens there, says Andrea Baronchelli, the new study's lead author. Baronchelli, who investigates human behavior in decentralized systems (such as social media and dark web markets) at City University of London, says previous research has shown repeated tweets on a political topic can prompt news coverage—and responses from politicians. “Both these categories play a huge role in shaping public opinion,” he says.

According to Ca' Foscari University of Venice researcher Fabiana Zollo, who studies misinformation's spread in digital platforms, X “is not giving us the whole picture, but it gives us necessary insights about how to communicate climate science, improve communication strategies and counteract misinformation.”

The study authors analyzed tweets posted through December 2021—before Twitter changed ownership and became X. Academics used to be able to apply for free access to Twitter's archives, but with X's current pricing researchers would have had to pay at least $42,000 per month to retrieve the number of tweets analyzed. “It means you can't do this kind of study anymore,” Margolin says.

Gina Jiménez is a bilingual journalist focused on health and science policy and how they affect disadvantaged communities. Follow her on X (formerly Twitter) @GinaRivers90

Computing

- **The Most Important Unsolved Problem in Computer Science**
  Here's a look at the $1-million math problem at the heart of computation

The Most Important Unsolved Problem in Computer Science

Here's a look at the $1-million math problem at the heart of computation

By Jack Murtagh

alengo/Getty Images

When the Clay Mathematics Institute put individual $1-million prize bounties on seven unsolved mathematical problems, they may have undervalued one entry—by a lot. If mathematicians were to resolve, in the right way, computer science's "P versus NP" question, the result could be worth worlds more than $1 million. They'd be cracking most online-security systems, revolutionizing science and even, in effect, solving the other six of the so-called Millennium Problems, all of which were chosen in the year 2000. It's hard to overstate the stakes surrounding the most important unsolved problem in computer science.

P versus NP concerns the apparent asymmetry between finding solutions to problems and verifying solutions to problems. For example, imagine you're planning a world tour to promote your new book. You pull up Priceline and start testing routes, but each one you try blows your total trip budget. Unfortunately, as the number of cities grows on your worldwide tour, the number of possible routes to check skyrockets exponentially, making it infeasible even for computers to exhaustively search through every case. But when you complain, your book agent writes back with a solution sequence of flights. You can easily verify whether their route stays in budget by simply checking that it hits every city and summing the fares to compare against the budget limit. Notice the asymmetry here: finding a solution is hard, but verifying a solution is easy.

The P versus NP question asks whether this asymmetry is real or an illusion. If you can efficiently verify a solution to a problem, does that mean you can also efficiently find a solution? It might seem obvious that finding a solution should be harder than verifying one. But researchers have been surprised before. Problems can look similarly difficult—but when you dig deeper you find shortcuts to some and hit brick walls on others. Perhaps a clever shortcut can circumvent searching through zillions of potential routes in the book tour problem. For example, if you instead wanted to find a sequence of flights between two specific remote airports while abiding by the budget, you might also throw up your hands at the immense number of possible routes to check. In fact, this problem contains enough structure that computer scientists have developed a fast procedure (or algorithm) for it that bypasses the need for an exhaustive search.

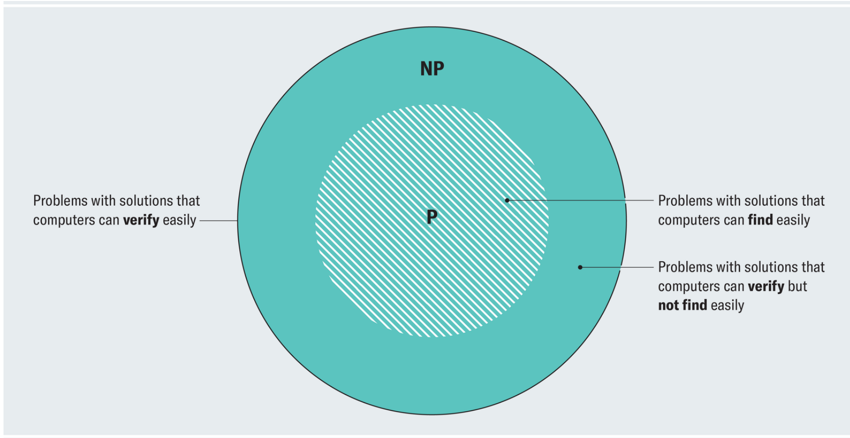
---

you are helping to ensure the future of impactful stories about the discoveries and ideas shaping our world today.

The P versus NP question rears its head everywhere we look in the computational world well beyond the specifics of our travel scenario—so much so that it has come to symbolize a holy grail in our understanding of computation. Yet every attempt to resolve it only further exposes how monumentally difficult it is to prove one way or another.

In the subfield of theoretical computer science called complexity theory, researchers try to pin down how easily computers can solve various types of problems. P represents the class of problems they can solve efficiently, such as sorting a column of numbers in a spreadsheet or finding the shortest path between two addresses on a map. In contrast, NP represents the class of problems for which computers can verify solutions efficiently. Our book tour problem, which academics call the Traveling Salesperson Problem, lives in NP because we have an efficient procedure for verifying that the agent's solution worked.

Notice that NP actually contains P as a subset because solving a problem outright is one way to verify a solution to it. For example, how would you verify that 27 × 89 = 2,403? You would solve the multiplication problem yourself and check that your answer matches the claimed one. We typically depict the relation between P and NP with a simple Venn diagram:



Credit: Amanda Montañez

The region inside of NP but not inside of P contains problems that can't be solved with any known efficient algorithm. (Theoretical computer scientists use a technical definition for "efficient" that can be debated, but it serves as a useful proxy for the colloquial concept.) But we don't know whether that's because such algorithms don't exist or we just haven't mustered the ingenuity to discover them. This representation provides another way to phrase the P versus NP question: Are these classes actually distinct? Or does the Venn diagram collapse into one circle? Can all NP problems be solved efficiently?

Here are some examples of problems in NP that are not currently known to be in P:

- Given a social network, is there a group of a specified size in which all of the people in it are friends with one another?
- Given a varied collection of boxes to be shipped, can all of them be fit into a specified number of trucks?
- Given a sudoku (generalized to $n \times n$ puzzle grids), does it have a solution?
- Given a map, can the countries be colored with only three colors such that no two neighboring countries are the same color?

Ask yourself how you would verify proposed solutions to some of the problems listed and then how you would find a solution. Note that approximating a solution or solving a small instance (most of us can solve a 9 × 9 sudoku) doesn't suffice. To qualify as solving a problem, an algorithm needs to find an exact solution for all instances, including very large ones.

Each of the problems can be solved via brute-force search (for example, try every possible coloring of the map and see whether any of them work), but the number of cases to try grows exponentially with the size of the problem. This means that if we call the size of the problem n (for example, the number of countries on the map or the number of boxes to pack into trucks), then the number of cases to check looks something like 2^n. The world's fastest supercomputers have no hope against exponential growth. Even when n equals 300, a tiny input size by modern data standards, 2^{300} exceeds the number of atoms in the observable universe. After hitting “go” on such an algorithm, your computer would display a spinning pinwheel that would outlive you and your descendants.

Thousands of other problems belong on our list. From cell biology to game theory, the P versus NP question reaches into far corners of science and industry. If P = NP (that is, our Venn diagram dissolves into a single circle, and we obtain fast algorithms for these seemingly hard problems), then the entire digital economy would become vulnerable to collapse. This is because much of the cryptography that secures such things as your credit card number and passwords works by shrouding private information behind computationally difficult problems that can become easy to solve only if you know the secret key. Online security as we know it rests on unproven mathematical assumptions that crumble if P = NP.

Amazingly, we can even cast mathematics itself as an NP problem because we can program computers to efficiently verify proofs. In fact, legendary mathematician Kurt Gödel first posed the P versus

NP problem in a letter to his colleague John von Neumann in 1956. Gödel observed that P = NP “would have consequences of the greatest importance. Namely, it would obviously mean that ... the mental work of a mathematician concerning yes-or-no questions could be completely replaced by a machine.”

If you're a mathematician worried for your job, rest assured that most experts believe that P does not equal NP. Aside from the intuition that sometimes solutions should be harder to find than to verify, thousands of the hardest NP problems that are not known to be in P have sat unsolved across disparate fields, glowing with incentives of fame and fortune, and yet not one person has designed an efficient algorithm for a single one of them.

Of course, gut feeling and a lack of counterexamples don't constitute a proof. To prove that P is different from NP, you somehow have to rule out all potential algorithms for all of the hardest NP problems, a task that appears out of reach for current mathematical techniques. Indeed, the field has coped by proving so-called barrier theorems, which say that entire categories of tempting proof strategies to resolve P versus NP cannot succeed. Not only have we failed to find a proof, but we also have no clue what an eventual proof might look like.

Jack Murtagh is a freelance math writer and puzzle creator. He writes a column on mathematical curiosities for *Scientific American* and creates daily puzzles for the Morning Brew newsletter. He holds a Ph.D. in theoretical computer science from Harvard University. Follow Jack on X @JackPMurtagh

---

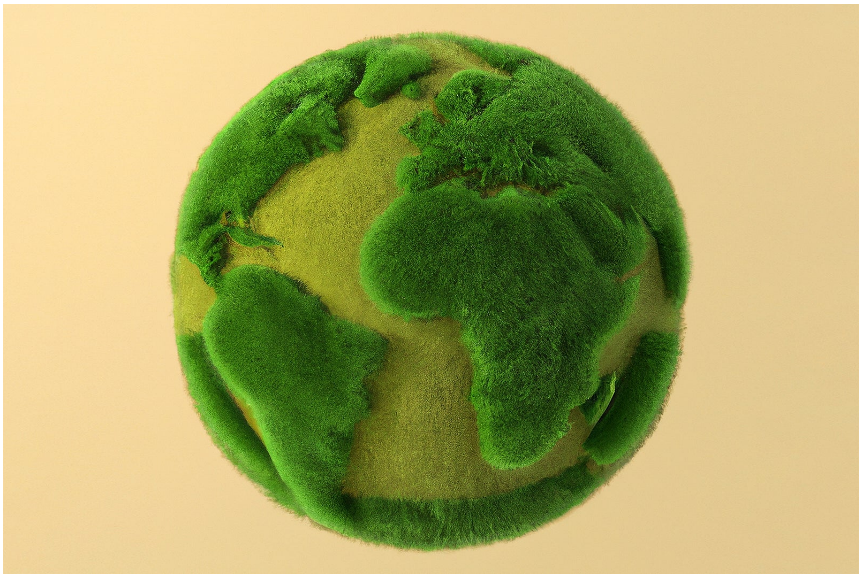
Conservation

- ## What's Missing from the Emoji Animal Kingdom?
  In the digital age, some scientists argue the emojisphere should better represent Earth’s biosphere—tardigrades, flatworms and all

What’s Missing from the Emoji Animal Kingdom?

In the digital age, some scientists argue the emojisphere should better represent Earth’s biosphere—tardigrades, flatworms and all

By Lauren Leffer



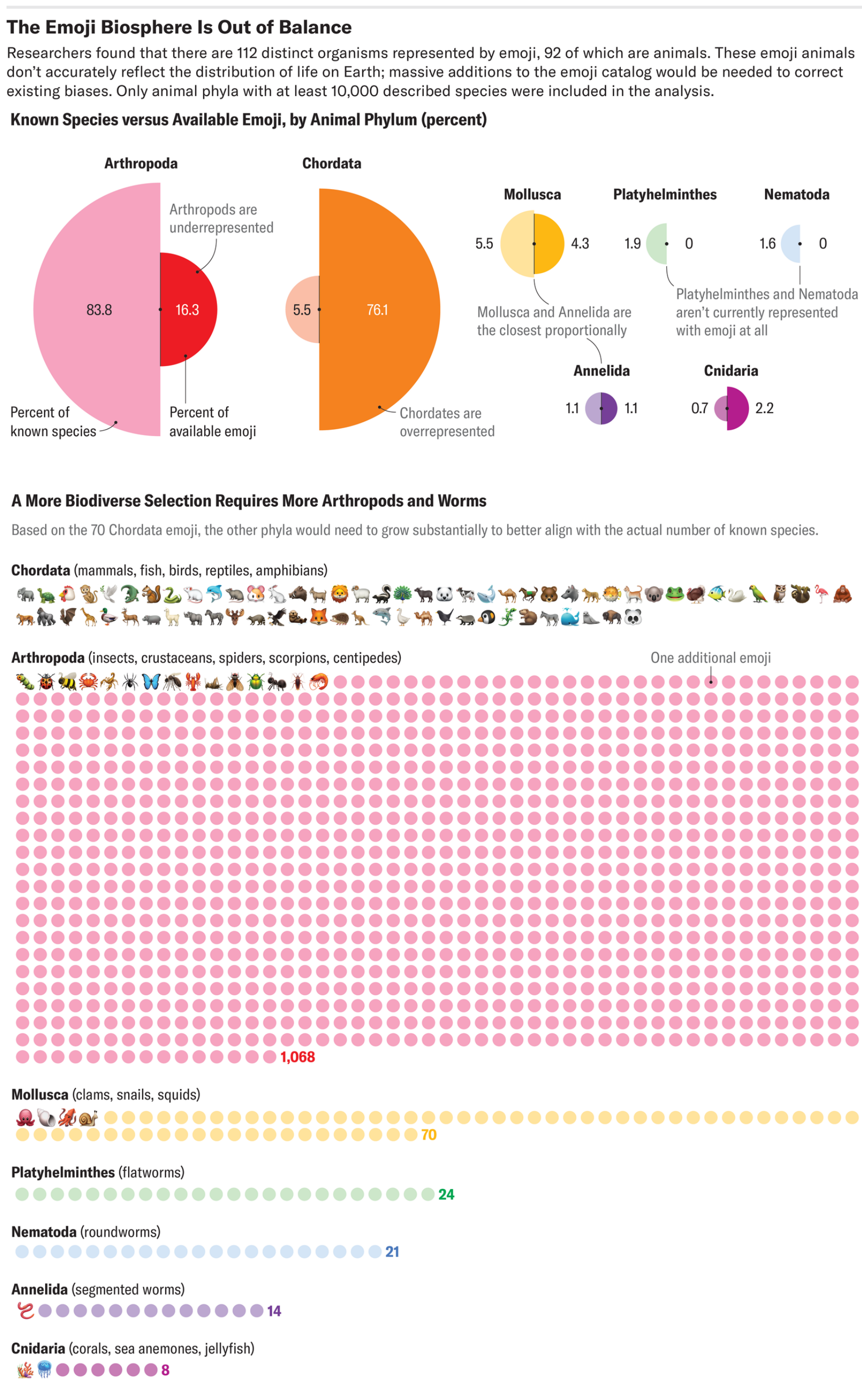
Eugene Mymrin/Getty Images

Ecologists and biologists commonly conduct surveys of the plants, fungi and animals in the environment they study. It’s less common, however, for that environment to be the virtual emojisphere. Emoji, the cutesy digital characters that have become their own mode of communication in text messages and online, are chock-full of representations of the natural world. Yet those representations are seriously skewed, according to a study published recently in the journal *iScience*.

In an analysis of Emojipedia, a comprehensive emoji catalog, researchers have found that the emoji tree of life is lopsided: It entirely lacks a few big branches, such as flatworms and echinoderms (the group that includes starfish), and heavily favors vertebrate animals over all other groups of life-forms. And the entire kingdoms of plants and fungi are barely featured at all; fungi only have a single representative toadstool. This type of partiality, termed taxonomic bias, isn’t unique. Humans have a well-known tendency to focus on and favor charismatic megafauna over other organisms. Yet the study authors—all life scientists—suggest that the large gaps in emoji biodiversity could be hampering digital communication about conservation and biology.

On public-facing social media platforms and even in chats among scientists, limited emoji offerings sometimes make it difficult to put together a quick and legible message, says study co-author Francesco Ficetola, an animal biology professor at the University of Milan in Italy. Biodiversity loss is an international issue, and conservation actions need to transcend borders and language barriers, he says. But “trying to communicate about biodiversity is difficult for us because the languages and cultures are different,” Ficetola adds. “Any tool, including emojis, that improves understanding and connection is fundamental.”

It’s easy to dismiss emoji biodiversity as trivial—especially when we’re in an IRL biodiversity crisis. Emoji have become an increasingly large part of human communication, however, and it’s worth examining what has and hasn’t yet made it into our ever expanding pixelated lexicon, says Sanjaya Wijeratne, a computer scientist at the video game publisher Nexon America, who has researched emoji meaning and use but was not involved in the new study. “The paper asks a valid question,” he says. Wijeratne points to published research that suggests that including emoji in social media posts can increase positive engagement with those posts (by about 70 percent, in some cases). A relevant emoji could help conservation researchers or organizations promote awareness of certain species online, Wijeratne and Ficetola both suggest.



Credit: John Knight; Source: "Biodiversity Communication in the Digital Era through the Emoji Tree of Life," by Stefano Mammola et al., in *iScience*, Vol. 26, No. 12; December 15, 2023

So far, however, there hasn't yet been any formal research on emoji use and online engagement related to conservation—a limitation highlighted within the study. And emoji are not free to produce, so new characters must be rigorously evaluated, says Jennifer 8. Lee, a current member and former vice chair of the nonprofit Unicode Consortium's Emoji Subcommittee, which is responsible for shaping emoji policies. Lee also co-founded Emojination, an organization that advocates for more inclusive and representative emoji.

Lee has been directly involved in expanding the emoji catalog many times over. She was instrumental in getting the dumpling emoji added to the library. In her roles with the subcommittee and Emojination, Lee has also helped expand the emoji species list. She says she assisted in getting approval for the earthworm, cockroach, beaver, sloth, llama, fly, mosquito and hippopotamus emoji.

Each of these additions had to be carefully considered in a cost-benefit analysis conducted by the Unicode subcommittee, which has strict guidelines for the creation of new emoji characters. The potential costs of added emoji includes computer server and electronic device storage space, keyboard clutter and payment to designers. Emoji are considered a font within our phones. More emoji characters make the font more cumbersome, and past a certain point, usership and accessibility might go down, Lee explains. For example, if cheaper phones have less storage space and can't accommodate the whole emoji library, or if finding a particular emoji requires scrolling through too many options, then fewer people will be able to use the tool at all.

There are also more philosophical questions at play. Though delighted by the study, Richard Ladle, a professor of conservation science at the Federal University of Alagoas in Brazil, isn't convinced that we need more biodiverse emoji or that more emoji options would boost conservation discussions. Ladle researches conservation culturomics, the study of human-nature interactions in the digital world. “Emojis are a fascinating thing to study and may give us some insights into peoples’ relationships with nature,” he says. But in his view, the icons are more a reflection of people’s cultural values and online conversations than a determiner of them.

Outside of biology, people often use nature emoji because of their indirect meanings. Animal emoji, especially, carry secondary connotations that can be positive or negative (think the butterfly as a symbol of beauty and freedom versus the cockroach as a symbol of “yuck”). It’s quite possible, Ladle says, that expanding the emoji biosphere could lead to more people using animals to communicate negative ideas. “You can imagine a dung beetle emoji might become a pejorative,” he says.

Ficetola has no illusions that every species should have its cartoon debut. “It’s not possible to have one million emojis,” he says. “That would make everything meaningless because emojis are for fast communication.” Yet he imagines that, with the input of scientists, emoji could better reflect our reality and make biology more legible to more people.

In recent years big strides have already been made in emoji biodiversity. The study points out that the number of animal taxa represented by emoji more than doubled between 2015 and 2022, which was a pleasant surprise for Ficetola. “My personal feeling is that we have a better understanding of biodiversity now,” he says. “The idea that biodiversity is more than just the panda and the lion is spreading through society,” emoji included. But still, he adds, “we can always do better.”

Ficetola and his colleagues make a few suggestions for where the emoji committee might want to start. A tardigrade (water bear), starfish and flatworm emoji would grant much more coverage to the breadth of animal biodiversity, they write in the study, while also bringing visibility to important organisms that are often hidden from public view. “Maybe just seeing an image of something strange like a tardigrade could increase awareness and public understand that our planet is much more complex than we often assume,” Ficetola says.

Yet Lee, one of the Unicode arbiters, doesn’t envision a likely future for these suggestions, which lack the often-necessary elements of cultural relevance and popular double entendre. Although she even advocated for a starfish emoji in the recent past, online interest and potential uses were too limited to push it through. “Starfish, I tried,” she says. “And it’s pretty, but that’s it.”

Perhaps, when it comes to emoji, communicating the complexity of the web of life is too grand a goal. But ↗ for the ☽, and maybe you'll ✈ among the ☆☆☆ (or ☆).

A version of this article entitled “Emoji Kingdom” was adapted for inclusion in the March 2024 issue of Scientific American.

Lauren Leffer is a contributing writer and former tech reporting fellow at *Scientific American*. She covers many subjects, including artificial intelligence, climate and weird biology, because she's curious to a fault. Follow her on X @lauren_leffer and on Bluesky @laurenleffer.bsky.social

---

Culture

Readers Respond to the November 2023 Issue

Letters to the editors for the November 2023 issue of Scientific American

Scientific American, November 2023

CULTURE AND HEALTH

"How Grammar Changes Perception," by Christine Kenneally, discusses the structure of grammar in language and how it influences our perception of the world around us.

I work in Aboriginal communities in Australia as a remote area nurse, and my wife, who is now deceased, was from the Waanyi clan in Queensland and the Northern Territory. Our children have grown up across both Western and Waanyi cultures. One thing I have observed—and questioned my children about—is the way they perceive the world. They have told me they "operate in different rooms" of their brain, moving from one room to another depending on where they are and what they are doing. Both cultures are comfortable to them, and they describe their thought processes as different for each.

On supporting science journalism

For example, one son says he works in a visual manner when he perceives the world from an Aboriginal perspective, whereas he works in a data format when he perceives the Western world. He notes that in the Western world, the storage of information is compact but takes him longer to process to recall information; in the Aboriginal world, he feels that there is more to store, but the recall is instant. With the latter, he can remember situations as if he is looking at a picture and notices any changes occurring within that picture in a fast manner.

I minored in anthropology while studying for my nursing degree. That, along with much cultural education from my wife, has helped me understand how to better deliver culturally safe health in remote communities, all of which have their own unique medical issues. The breakdown of kinship systems and the loss of language are major barriers to effective health care in such communities. Both Indigenous and Western health systems and beliefs offer barriers to each other, and causation of disease is rooted in culture. A lack of appreciation of language brings many nurses to see culture on a single continuum or line. My experience is that cultures all have their own distinct lines that travel in parallel; they sometimes diverge, sometimes intersect and often touch.

DAVE CORSTORPHAN *LOOMA CLINIC, LOOMA COMMUNITY, AUSTRALIA*

GENDER PREHISTORY

In “Woman the Hunter,” Cara Ocobock and Sarah Lacy present evidence of female hunters among early human communities, demonstrating that the popular notion that our human ancestors had a division of labor “in which males evolved to hunt and provide and females tended to children and domestic duties” is incorrect. Are there any hard data about the other side of the coin: the possibility that early male humans tended to children and domestic duties? If not, would the facts in the article not just mean that the prehistoric situation was very similar to now, with (most of) men working and (most of) women working *and* caring for the children and the household?

URSULA GARTENMANN *SWITZERLAND*

THE AUTHORS REPLY: *Extant forager groups are not always egalitarian when it comes to child care. But we have almost no archaeological evidence of child care, such as baby slings, in the Paleolithic. (Some of the figurines from the Upper Paleolithic, such as those found buried with adolescents at the Sungir site in Russia, may be toys, however.) From analyses of teeth, we can see that Neandertals and early modern humans were starting to wean their infants around six months, similar to people today, which suggests that anyone in a group of these early humans would have been able to feed and carry an infant at that point. And regarding other “household” tasks, we know that male Neandertals were also processing leather, so this “domestic” duty, at least, was not gendered either.*

It is totally plausible that Paleolithic males were pulling their weight with child care in their small groups, whether paternity was *certain or not, and there is a movement in the field to better recognize the signatures of children and child care within paleoanthropology and archaeology.*

Considering the trend toward patrilocality among the Neandertals, most of the children in a group would have been a male's nieces, nephews and cousins, if not his children, so there would still be a fitness advantage to caring for them.

GALACTIC TRAFFIC JAM

"The Milky Way's Secrets," by Phil Plait [The Universe], displays an image of a beautiful spiral galaxy that is probably much like our own Milky Way. Is there a general principle of physics that could apply to how such spiral shapes are generated, especially for the "arms" of galaxies? Or are there different mechanisms that can cause them?

BARRY MALETZKY *PORTLAND, ORE*.

PLAIT REPLIES: *There are many kinds of spiral galaxies, including ones with tightly wound or wide-flung arms, multiple arms, "spurs" (short bridges stretching between the arms like spokes), and more. Overall, the leading hypothesis is that spiral patterns are so-called density waves: regions of slightly higher density in a disk that travel around the galaxy at a different speed than those of its stars and that the stars move into and out of over time*.

This idea is similar to a traffic jam: the jam can move slower than traffic overall, and cars can move into and out of the jam even as it persists. Gas clouds collect in these density waves and collapse to form more stars. The different kinds of arms arise from local conditions, such as how much gas there is, how massive the galaxy is, whether it recently underwent a collision with another galaxy, how many stars are forming in the arms, and much more. This *obviously involves very complex physics, and astronomers still aren't sure about all the details. It's an area of active research and no doubt will continue to be for a very long time*.

SERIOUS FUN

In “Why We Need Scary Play,” Athena Aktipis and Coltan Scrivner explain how play and simulations can prepare us for real-world scenarios. In 2022 my sister gave me a fun Christmas present: a board game named Pandemic. In the game, players battle against the board to save the world from four different diseases that can get out of control and become a pandemic.

After reading Aktipis and Scrivner's article, I understood why this game raises our stress levels so much. It does feel like a realistic competition where fear, collaboration and a sense of responsibility take over all participants during playtime. I used to play Pandemic for fun. Now I know it may also psychologically prepare me for the next epidemic or pandemic.

AILYN MONTES *MIAMI*

ERRATUM

In the December 2023 table of contents, the caption for an image regarding “The Cosmic Surprise,” by Richard Panek, should have described the universe as being pulled apart by dark energy, not dark matter.

Drug Use

- **Is Marijuana Bad for Health? Here's What We Know So Far**
  Marijuana’s health impacts—good and bad—are coming into focus

Is Marijuana Bad for Health? Here's What We Know So Far

Marijuana's health impacts—good and bad—are coming into focus

By Jesse Greenspan

Cappi Thompson/Getty Images

With decades of legal and social opprobrium fading fast, marijuana has become an extremely popular commercial product with more than 48 million users across the U.S. Health concerns, once exaggerated, now often seem to be downplayed or overlooked. For example, pregnant patients "often tell me they had no idea there's any risk," says University of Utah obstetrician Torri Metz, lead author of a recent paper in the *Journal of the American Medical Association* on cannabis and adverse pregnancy outcomes.

Fortunately, legal reforms are also gradually making it easier to study marijuana's health effects by giving U.S. scientists more access to the drug and a wider population of users to study. Although much research remains in "early stages," the number of studies has finally been increasing, says Tiffany Sanchez, an environmental health scientist at Columbia University. As new results accumulate, they offer a long-overdue update on what science really knows about the drug.

The Bad

On supporting science journalism

If you're enjoying this article, consider supporting our award-winning journalism by subscribing. By purchasing a subscription you are helping to ensure the future of impactful stories about the discoveries and ideas shaping our world today.

In addition to minor side effects that many users joke about—such as short-term memory loss—recent studies have linked marijuana to adverse health outcomes involving the lungs, heart, brain and gonads. For example, heavy marijuana consumption seems to increase the risk of clogged arteries and heart failure, and it may impact male fertility. Smoking weed likewise can lead to chronic bronchitis and other respiratory ailments (although, unlike tobacco, it hasn't been definitively tied to lung cancer). And cannabis plants hyperaccumulate metal pollutants, such as lead, which Sanchez found can enter users' bloodstreams.

Developing adolescent brains, particularly those predisposed to mental illness, may be most at risk from overconsumption. Although psychiatric effects are hotly debated, studies suggest that heavy weed use exacerbates—or may trigger—schizophrenia, psychosis and depression in youths and that it affects behavior and academic performance. “From a safety viewpoint, young people should definitely stay away from it,” says University of Ottawa psychiatrist Marco Solmi, lead author of a recent review of cannabis and health in the *British Medical Journal*.

> 24 states have legalized recreational marijuana, with 38 allowing medical use

Moreover, the drug can cross over to fetuses during pregnancy. Several studies have linked it to low birth weights, and researchers suspect it raises the likelihood of neonatal intensive care unit admissions and stillbirths. Some cannabis dispensaries have advertised their products as a cure for morning sickness, but Metz emphasizes that safer alternatives exist.

The Good

Of course, many adults use marijuana responsibly for pleasure and relaxation. Unlike with, say, opioids, there's effectively zero risk of life-threatening overdose. Plus, "people get addicted with tobacco way faster," says Columbia University epidemiologist Silvia Martins, who studies substance use and related laws.

Cannabis, and its derivatives, also may help alleviate pain—although some researchers contend that it performs little better than a placebo. It may also decrease chemotherapy-induced nausea, calm epileptic seizures, ease the symptoms of multiple sclerosis and serve as a sleep aid.

Recent studies have hinted that the drug might slightly reduce opioid dependency rates, although this, too, is disputed. There's some evidence that weed users tend to be more empathetic, and researchers found that elderly mice get a mental boost from the drug. Still, experts caution against self-medicating: "You should ask your doctor," Solmi says.

The Weird

Some of the recent research into marijuana is more lighthearted. One study, for instance, found that, just like people, nematode worms dosed with cannabis get the munchies.

Jesse Greenspan is a San Francisco Bay Area–based freelance journalist who writes about history and the environment.

Electronics

- ## To Reignite the U.S. Chip Industry, Invite More Chefs into the Kitchen
  A “more is merrier” approach to making computer chips would create the vibrant and fast breakthroughs that America needs to succeed

To Reignite the U.S. Chip Industry, Invite More Chefs into the Kitchen

A “more is merrier” approach to making computer chips would create the vibrant and fast breakthroughs that America needs to succeed

By H.-S. Philip Wong

Peter and Maria Hoey

Making the next generation of computer chips demands the care of preparing a gourmet meal at an industrial scale. The finest ingredients, techniques and tools and, of course, the sharpest minds must come together to create something transformative. In kitchens, when just one of these elements is missing, the meal falls short.

The U.S. Department of Commerce will soon command a feast of sorts, doling out $11 billion for research and development under the CHIPS and Science Act to revive America's sluggish chipmaking industry. Currently the U.S. produces only 12 percent of chips worldwide. In passing CHIPS, the federal government asserted a bold desire to return to the forefront of chipmaking. Between desire and doing, however, lies a profound gap. It will not be easily spanned.

The CHIPS Act promises a lot of funding, but numbers hardly guarantee success. Chipmaking is an almost incomprehensibly precise, difficult and expensive business. Ensuring that the U.S. is among the world's leading makers will take innovation and collaboration of epic proportions.

Lowering the barriers to participation and financing is critical to the success of America's CHIPS-inspired semiconductor future. The nearly $600-billion global semiconductor-chip industry makes more than a trillion chips every year, which go into everything from cars to coffeemakers. A “more is merrier” approach would create a vibrant and fast-moving network of innovation to produce the breakthroughs needed for technology to flourish.

So far a lot of attention has instead been directed at the biggest, and often slowest-moving, players in the industry. The U.S. thrives on invention, yet many of the most creative minds at smaller companies and universities have been shut out by the field's notoriously high R&D costs, as well as by lack of access to the expensive tools and facilities (often called “fabs”) needed for prototyping. As any high-end chef will tell you, you can't make something transformative if you don't have a kitchen in which to cook.

There are two ways the U.S. can allot these R&D funds. It can fall back on the “celebrity chef” approach, placing a few big bets on a handful of names and hoping for a magical breakthrough—a risky proposition. Or it can build more kitchens and invite more chefs to start cooking, placing many smaller bets on the most innovative minds to create a collaborative restaurant row of chipmaking progress where the steak, and not the sizzle, is the focus.

First, we'll need more kitchens. We must build shared experimental facilities where researchers from industry and academia work together. This environment would cultivate research communities of practice that unite top engineers in a common purpose and open doors to innovators nationwide. These technology centers can range from enhanced university fabs (semiconductor-manufacturing plants) run by experienced engineers to industry-scale facilities that showcase the product-worthiness of new ideas.

Second, we must build “digital twin” computational models that can emulate entire fabrication processes, including specifics of their tools and conditions. Digital twins will allow researchers to quickly evaluate options and accelerate the discovery of new process and device technology. By collecting large amounts of data, artificial-intelligence models could assist humans in fine-tuning delicate fabrication processes and automatically detect anomalies during manufacturing to improve production yield and quality. Such digital models will be America's ticket to leadership in semiconductor technology and in manufacturing approaches that are less dependent on humans looking at spreadsheets to make decisions.

Inherent in this vision is the key role of data sharing. The open-access facilities would be networked nationwide to enable free data distribution among the entire community, all while protecting proprietary information. The impact of the digital twins would grow exponentially as the research community expanded and as new knowledge fed innovation.

Ultimately, such a network would lower R&D costs for all players, reducing risk and bringing new ideas to market faster. A shared prototyping environment could propel the production of all kinds of chips—logic, memory, storage and specialty technologies—the way a well-equipped kitchen can turn out a variety of cuisines in a single night.

All this talk of improving access, however, overlooks one more critical factor of success: the availability of talent. Greater access means nothing if we do not have exquisitely skilled chefs eager to step into the kitchen. We must nurture a healthy industry with robust profit margins that supports career growth, attractive work conditions and suitable work-life balance to entice the next generation of engineers. Such talent does not grow on trees; it takes a decade or more for a high school graduate to earn a Ph.D. If we want to be part of a leading industry 10, 20, 30 or 100 years down the road, we must recruit and train new talent now.

The aspirations of the CHIPS Act are high. Regaining global leadership in chipmaking will not be easy, but it is within reach for a nation as capable and as motivated as the U.S. The recipe couldn't be simpler: Train the best and brightest. Create an open and collaborative ecosystem. Put cutting-edge tools in people's hands. Allow everyone to participate and share data. Then enjoy a delectable lab-to-fab meal. Bon appétit.

This is an opinion and analysis article, and the views expressed by the author or authors are not necessarily those of Scientific American.

H.-S. Philip Wong is Willard R. and Inez Kerr Bell Professor in the School of Engineering at Stanford University.

Energy

- **Snowflakes Swirl According to Surprisingly Simple Math**
  Every snowflake is different, but new technology reveals they all swirl the same

Snowflakes Swirl According to Surprisingly Simple Math

Every snowflake is different, but new technology reveals they all swirl the same

By Ellyn Lapointe

Сергей Бурбоha/Getty Images

For all snowflakes' infinite structural variation, their journeys to Earth are remarkably similar—even predictable. Researchers tracking more than half a million falling flakes have uncovered a broad mathematical pattern that describes precisely how they swirl through the air.

University of Utah atmospheric scientist Tim Garrett, senior author of a new study in *Physics of Fluids*, has studied snowflakes for nearly a decade. Although the behavior of such tiny, ephemeral objects may seem inconsequential, their fall speed is a key variable in forecasts of weather and climate, even in the tropics; most precipitation, regardless of where it eventually ends up, begins as snow.

Snowflake movement is typically studied in laboratories under controlled conditions that don't reflect the complexity of nature. Scrutinizing falling snowflakes in the field has challenged atmospheric scientists for decades.

For a new approach, Garrett teamed up with University of Utah engineers Dhiraj Kumar Singh and Eric Pardyjak to build a machine that measures the mass, density, area and shape of individual snowflakes that land on a hotplate. By placing this instrument underneath video cameras and a plane of laser light, the researchers could track how each snowflake moved in response to outdoor air turbulence.

“We were able to let the atmosphere express itself, to behave in a way that was completely uncontrolled by a scientist,” Garrett says. “I think that's why we ended up uncovering an extraordinary simplicity, an elegance.”

The researchers discovered a linear correlation between a snowflake's average acceleration—which, in this study, is equivalent to how much it swirls—and its Stokes number, a value that describes how quickly an object responds to changes in air turbulence. For instance, a wide and fluffy flake swirls more than a streamlined one.

Using the Stokes number, researchers can now predict how much a single snowflake will swirl as it falls. On a broader scale, the team was surprised to find that the distribution of average snowflake swirliness fits a single, nearly perfect exponential curve—a fixed mathematical pattern—despite the wide variability of air turbulence and range of snowflake shapes and sizes.

The cause of this regularity remains a mystery for now. But Garrett says that it could be related to how turbulent air prompts snowflakes to fluctuate in shape and size—which in turn can tweak their responses to that turbulence.

Further research is needed to assess the mathematical pattern's universality, says University of Minnesota mechanical engineer Jiarong Hong. “We will look into the applicability of [this result] to our data sets of snowflakes captured under different conditions,” including varied altitudes and ground roughness, he adds.

If the pattern does hold universally, “the fact that there's this simplicity suggests there's going to be a simple explanation,” Garrett says. “We just have to find it.”

Ellyn Lapointe is a New York City–based science journalist with a special interest in covering the life sciences.

Epidemiology

- **Helpful Gut Bacteria Seem to Reduce Allergic Disease in Kids**
  In babies, the right combo of gut bacteria might stave off later allergies, so scientists are testing “cocktails” of helpful microbes as therapy

Helpful Gut Bacteria Seem to Reduce Allergic Disease in Kids

In babies, the right combo of gut bacteria might stave off later allergies, so scientists are testing “cocktails” of helpful microbes as therapy

By Lydia Denworth

Jay Bendt

I stopped sending peanut butter and jelly sandwiches to school with my kids around 2007. That was roughly the moment when people started talking about a dramatic rise in the number of children with serious nut allergies. Cases of all kinds of allergies in youngsters have increased since then. The prevalence of asthma has doubled since the 1980s, and more than one quarter of children have eczema, food allergies, or hay fever or other seasonal allergies.

A host of studies from around the world strongly suggest that our allergy epidemic is the result of reduced exposure to germs in early life. During this critical window of time, an infant's immune system learns to defend against dangerous microbes and to tolerate good ones that can live in the gut and aid in processes such as digestion. This immune education comes from encountering a wide variety of germs. But as social habits have changed, leading us to spend more time indoors, these encounters have been reduced, and immune overreactions—allergies—have climbed.

This idea, introduced decades ago as the “hygiene hypothesis” and refined over the years, is supported by epidemiological studies showing that having older siblings, attending day care, living on a farm and having pets protect against allergies. But more antiseptic early lives—delivery by cesarean section, not receiving breast milk and getting antibiotic therapy in the first year of life—seem to increase risk.

Now stronger evidence is emerging that clarifies the ways that microbes inside children's guts can trigger allergies. Scientists are working out how the presence or absence of certain bacteria in kids' digestive systems affects allergic risk, thanks to technological advances that let researchers identify more types of gut microbes. Someday it might be possible to replace certain microbes in children and in the population at large and thereby lessen people's susceptibility to allergies.

In infancy the gut microbiomes of children who later develop allergies or asthma look different from those of children who don't go on to have allergies. “Children who are at the highest risk are missing important health-promoting bacteria in that first year of life,” says Stuart Turvey, a pediatric immunologist at the University of British Columbia and British Columbia Children's Hospital.

Among other things, the presence of certain innocuous bacteria early on creates a welcoming environment that allows other, helpful bacteria to follow in predictable waves. If those first “keystone” bacteria are missing, the subsequent waves of colonization are delayed or disrupted. “Microbial exposures in early life can really shape the immune system in ways that they can't much later in life,” says Supinda Bunyavanich, a pediatric allergist and immunologist at Mount Sinai in New York City.

In a study of more than 1,100 children published in 2023, Turvey and his colleagues found that children who had these microbiome disruptions at age one were more likely to be diagnosed with eczema, food allergies, allergic rhinitis or asthma at age five. “Not every kid gets all four [diagnoses], but often the kids who had two or more had a more pronounced microbiome imbalance signature,” he says.

Work in mice has helped researchers determine which microbes are especially influential and why. Talal Chatila, a physician who works in the food allergy program at Boston Children's Hospital, found that giving allergy-prone mice microbes from the orders Clostridiales and Bacteroidales protected the animals from developing food allergies. “Particular microbes within a healthy gut act to suppress allergic responses,” Chatila says. One way they do that is by promoting the formation of regulatory T cells, which help to control immune system responses.

Another type of bacteria that has a positive effect on humans is *Bifidobacterium infantis*, which eats sugars in breast milk and is more abundant in some children who are breastfed. *B. infantis* was once common in people's guts but is much less so now in Western countries. “Only 16 percent of Canadian kids have this, and rates are lower in the U.S.,” Turvey says. Among youngsters who had to have antibiotics in infancy, the presence of *B. infantis* protected them against developing asthma by age five, Turvey's studies have shown. Antibiotics reduce microbial diversity in the gut, but these particular bacteria seem to counter those negative effects.

Multiple clinical trials are underway to test allergy treatments with “cocktails” of selected bacteria. Most of these trials involve treating infants who are at high risk for allergies and then following them through childhood to see whether the treatments keep the children allergy-free. For prebiotics and probiotics now on the market, there is no convincing evidence that they can make allergies go away.

Biotherapeutics are not the only answer. Avoiding unnecessary cesarean sections and antibiotics and enacting policies that support breastfeeding could also help, Bunyavanich says. She is working on a trial comparing children born vaginally, who are exposed to microbes in the birth canal, with children born by C-section who had the mother's vaginal fluids applied at birth. Both will be compared with children born by C-section without any microbial exposure.

The scientists will follow the kids through early childhood to see who has increased risk of allergies. If this and the other trials do reduce allergies, bringing back the microbes we've lost could turn out to be a key health strategy.

This is an opinion and analysis article, and the views expressed by the author or authors are not necessarily those of Scientific American.

Lydia Denworth is an award-winning science journalist and contributing editor for *Scientific American*. She is author of *Friendship* (W. W. Norton, 2020).

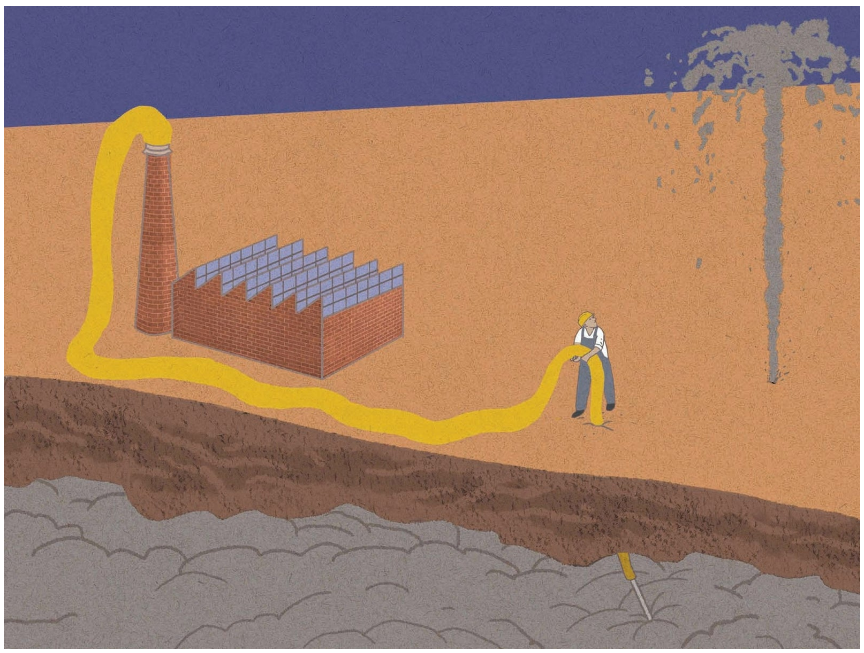
Fossil Fuels

- **The False Promise of Carbon Capture as a Climate Solution**
  Fossil-fuel companies use captured carbon dioxide to extract more fossil fuels, leading to a net increase in atmospheric CO2

The False Promise of Carbon Capture as a Climate Solution

Fossil-fuel companies use captured carbon dioxide to extract more fossil fuels, leading to a net increase in atmospheric CO_2

By Naomi Oreskes



Izhar Cohen

Last December the leaders of the United Nations Climate Change Conference (COP28) in Dubai declared victory as the parties agreed to “transition away” from fossil fuels. But there's a big issue that will remain contentious as countries try to define what counts as a transition: so-called unabated fossil-fuel use. Among its provisions, the agreement called for “accelerating efforts towards the phase-down of unabated coal power.”

Abatement in this context means carbon capture and storage (CCS). It's the idea that we can still use fossil fuels as long as the carbon dioxide emitted is captured and stored in the ground. In the

U.S., the oil and gas industries have been pushing this approach as one of the key solutions to the climate crisis. But how realistic is it?

Let's start with a few facts. Oil is sticky stuff, and when you try to pump it out of a reservoir, most of it gets left behind, stuck to the rocks. But if you flood a field with water, detergents or gas (such as CO_2), you can flush out much of the remaining oil. This technique is known as enhanced oil recovery, and it's been standard industry practice for a long time. According to the U.S. Department of Energy, gas injection accounts for more than half of the enhanced oil recovery in the U.S. and has helped to add decades of life to fields that would otherwise by now have run dry. The same approach is used in gas fields to maintain the pressure that keeps the gas flowing.

In recent years the oil industry has tried to pour this old wine into new bottles, casting the practice as a method of mitigating climate change because some of the injected CO_2 might otherwise end up in the atmosphere. In theory, it's a good idea. In practice, there are big problems.

We all know the saying that what goes up must come down, but the opposite is largely true, too (at least if the materials involved are liquid or gas), because fluids migrate through the microscopic holes and fractures that are found in even the most solid of rocks. After the U.S. government spent billions evaluating a potential civilian nuclear waste disposal site at Yucca Mountain in Nevada, the proposal failed in part because scientists could not guarantee that the waste would stay put. That waste was mostly a mix of solids and liquids. The waste CO_2 that we would be storing to stop climate change would be a buoyant, low-viscosity "supercritical" fluid—that is, a fluid maintained at such a high temperature and pressure that distinct gas and liquid phases do not exist. Like all fluids, it would have the capacity to migrate through the ground and find its way back to the surface and, from there, the atmosphere.

Many geologists (myself included) believe there are places on Earth where long-term CO_2 storage could be safely achieved, but it would require what scientists call "site characterization." That means studying the location in enough detail to be confident that things put there will stay there. For example, the U.S. currently stores military radioactive waste in low-permeability salt formations in New Mexico, and there are numerous pending proposals to store CO_2 in sandstones overlain by low-permeability shales in North Dakota.

But site characterization takes time that we don't have. The DOE spent more than 20 years evaluating Yucca Mountain. It spent some 14 years studying the New Mexico site. The Intergovernmental Panel on Climate Change concluded in 2018 that we have only until 2030 to stop irreversible climate damage, so it's urgent that we focus our attention on solutions that can be implemented right now.

We could jump-start the project by expanding existing carbon capture and storage sites. The problem, as Massachusetts Institute of Technology professor Charles Harvey and entrepreneur Kurt House have explained, is that nearly all CCS projects in the U.S. are actually enhanced-recovery projects that keep the oil and gas flowing, and every new barrel of oil and cubic foot of gas sold and burned is putting more CO_2 into the atmosphere. So not only do these kinds of projects not help, but they perpetuate our use of fossil fuels at a critical moment in history when we need to do the opposite.

Despite the U.S. government having spent billions on failed CCS projects, under the Inflation Reduction Act (IRA), it is set to spend many billions more, a lot of it in tax subsidies to fossil-fuel companies. In theory, IRA tax credits are to be used for “secure” carbon storage, but the mechanisms for ensuring that CO_2 is not leaking back into the atmosphere are flimsy at best. And it gets worse: the Environmental Protection Agency has concluded that if the price of CCS falls—because of tax credits, for example, or economies of scale—some currently closed oil or gas fields might reopen.

There is another model for CCS: the Orca plant in Iceland, where CO_2 is taken directly from the air and dissolved in water, which then reacts with basalt—the rock that makes up both Iceland and the ocean floor—to create stable carbonate minerals. But it's wildly expensive: $1,200 per metric ton of captured CO_2. (Bill Gates has negotiated a bulk deal for Microsoft at “only” $600 per ton.) The U.S. produces about 6,000 million metric tons of CO_2 per year. If for ease of arithmetic we assume a cost of $1,000 per ton, then offsetting U.S. emissions would cost about *$6 trillion every year*. In time these costs will probably come down, but time is what we don't have.

It is said that Mahatma Gandhi was once asked what he thought of Western civilization. He replied, “It would be a good idea.” The same could be said about carbon capture and storage as a solution to the climate crisis. Although it might be part of the solution down the road, right now it's mostly a dangerous distraction. Our focus—and our tax dollars—should be trained on scaling up production of cost-competitive renewable energy, grid-scale batteries for storing that energy and efficiency measures to conserve it as fast as we possibly can.

Naomi Oreskes is a professor of the history of science at Harvard University. She is author of *Why Trust Science?* (Princeton University Press, 2019) and co-author of *The Big Myth* (Bloomsbury, 2023).

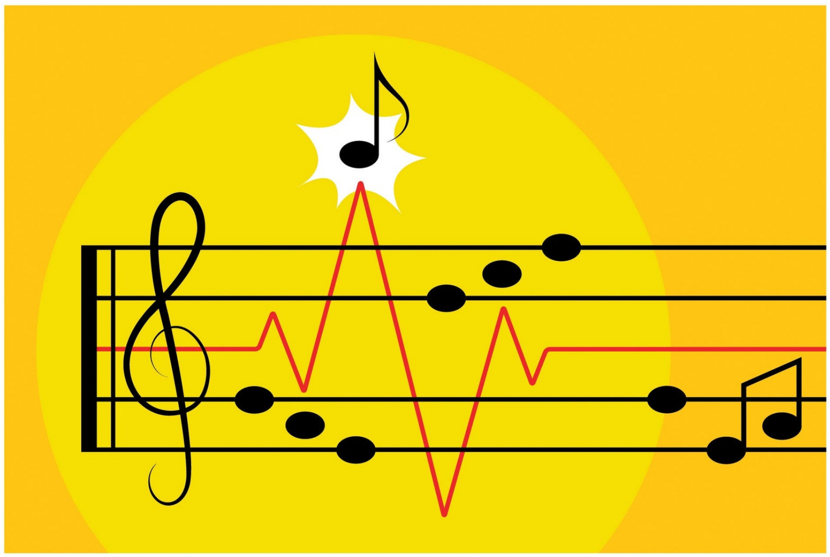
Health Care

- **Making Alarms More Musical Can Save Lives**
  Medical alarms don’t have to be louder to be more effective

Making Alarms More Musical Can Save Lives

Medical alarms don’t have to be louder to be more effective

By Rachel Berkowitz



Thomas Fuchs

Beeping alarms in hospitals are a life-or-death matter—but with so many going off all the time, medical professionals may experience alarm fatigue that impairs care. Researchers now report that changing an alarm's sound to incorporate properties of musical instruments can make it more helpful amid the din.

Auditory alarms can sound up to 300 times a day per patient in U.S. hospitals, but only a small fraction require immediate action. Data from the U.S. Food and Drug Administration suggest that alarm fatigue (including when clinicians turned off or forgot to restart alarms) and other alarm-related issues were linked to 566 deaths over five and a half years.

After a typical day at the hospital, “I'd leave with beeping in my ears,” says Vanderbilt University Medical Center anesthesiologist Joseph Schlesinger. He collaborated with Michael Schutz, a music cognition researcher at McMaster University in Ontario, to analyze how musical sounds could improve hospital alarms.

In 2015 Schutz and Schlesinger began examining musical qualities called timbres that might let softer sounds command attention from busy clinicians. They found that sounds with a “percussive” timbre, many of which contain short bursts of high-frequency energy—such as wineglasses clinking—stand out even at low volume. In contrast, loud, “flat” tones that lack high-frequency components, like a reversing truck's beep, get lost. The researchers have since conducted experiments in which participants evaluate different sounds and melodies for annoyance, detectability and recognizability.

For a recent study detailed in *Perioperative Care and Operating Room Management*, the researchers played participants the same sequences of notes with varying timbres. They found the sounds that made these sequences least annoying, with no decrease in recall, were percussive and had complex, time-varied harmonic overtones (the many components within a single sound) like a xylophone's ping, rather than a few homogeneous ones like monotonous mechanical beeps.

The researchers are also drawing inspiration from the timbres of other musical instruments: the triangle, for example, famously stands out in a crowd of sounds, possibly because it has overtone sequences that deviate from traditional harmonic series. “We're using music as a cookbook and learning what we can take from it,” Schutz says.

Such findings could lead to alarms that command attention and fit into current regulatory guidelines. Michael Rayo, who studies cognitive systems design at the Ohio State University, says that acoustically complex sounds like those in the study avoid trading detectability for recognition. Experimenting with timbre, he says, “furthers our understanding of aspects that reliably support strong performance.”

Applied psychologist Judy Edworthy, professor emeritus at the University of Plymouth in England, says the finding that musical tones can help improve alarms is important for future patient monitoring and equipment designs. Still, she warns, “any sound can induce alarm fatigue if it is constantly false.”

Rachel Berkowitz is a freelance science writer and a corresponding editor for *Physics Magazine*. She is based in Vancouver, British Columbia, and Eastsound, Wash.

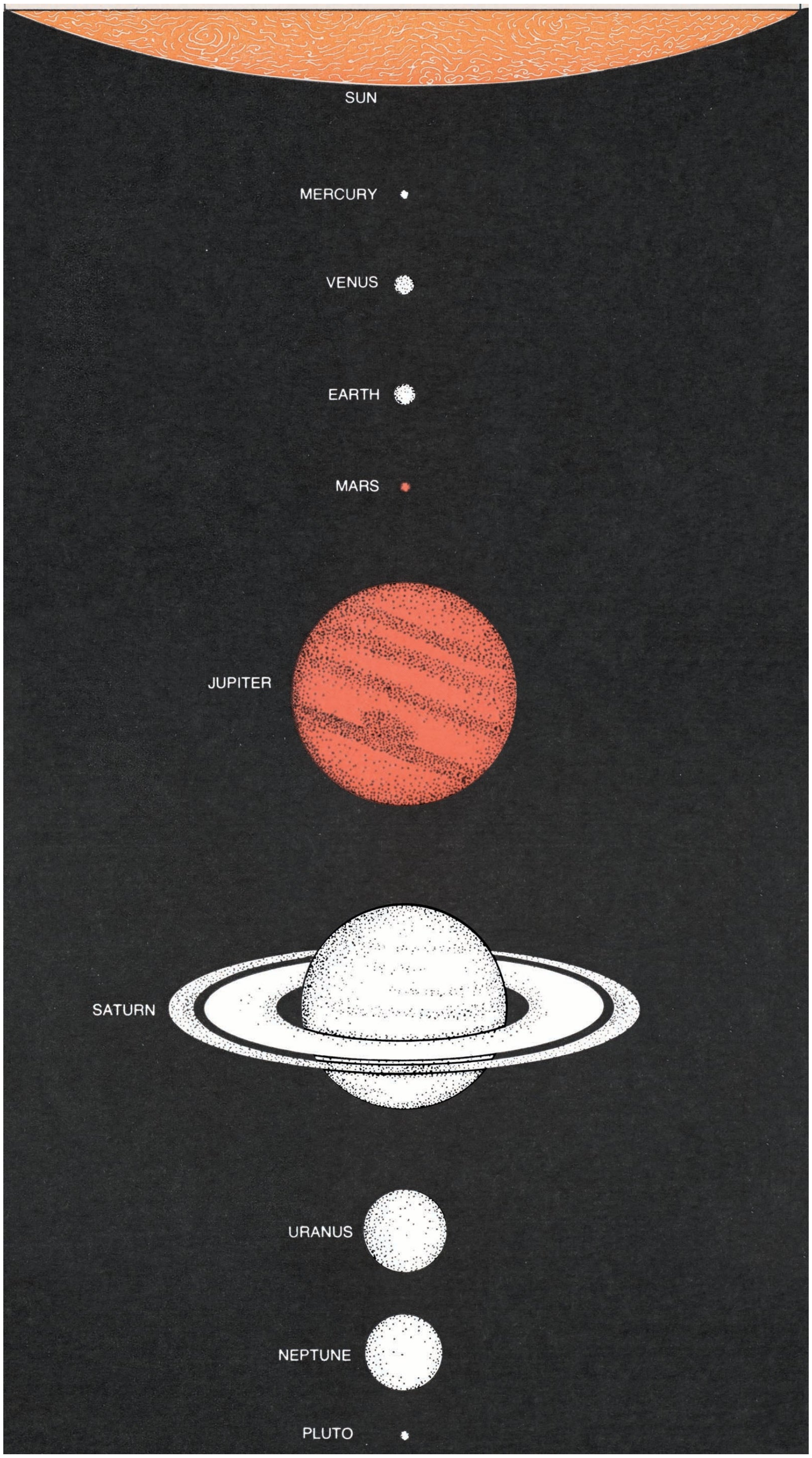
History

- **March 2024: Science History from 50, 100 and 150 Years Ago**
  Hashish addiction; a pension for Madame Curie

March 2024: Science History from 50, 100 and 150 Years Ago

Hashish addiction; a pension for Madame Curie

By Mark Fischetti



1974, Planet Data: Mercury's gravity is 0.38 that of Earth's. Jupiter's mass is about 318 Earths. One Neptune revolution around the sun takes 165 Earth years. Venus rotates in a direction opposite to that of Earth.
Scientific American, Vol. 230, No. 3; March 1974

1974

How Balanced Rocks Stay Upright

“A large rock balanced on a small protuberance is a wonder. In Goblin Valley in southern Utah there are more than 1,000. But how do the rocks stay balanced? Balanced rocks originate when a bed of sediments is dissected by erosion until a column is formed. If the strata at the top are harder than the strata farther down, erosion will whittle the softer rock to a pillar. So what keeps the capstone in place? Two investigators at Kansas State University suggest that when the capstone first begins to tilt, the point of contact to its pillar shifts, remaining under the capstone's center of gravity. Rock under the stress of compression is more resistant to erosion than unstressed rock, so thereafter the unstressed section will erode more rapidly than before. Successive tilts in other directions will stress successive sections of the pillar, and the differential erosion will make the process self-leveling. The capstone will remain poised on the pillar until the inevitable day when the area of contact becomes too small for the self-leveling to continue, and the balancing rock crashes satisfyingly to the ground.”

Reactor Rush

“At the end of 1973 the U.S. had 42 operable [commercial] nuclear reactors, according to the Atomic Industrial Forum Inc. It also counts 56 reactors under construction, 101 under firm order and 14 ‘under letters of intent or options.’”

After the 1979 Three Mile Island accident in Pennsylvania, utilities canceled more than 50 reactor orders from 1980 to 1984. The 1986 Chernobyl accident in Ukraine hurt prospects further.

1924

Absolutely Tremendous

“There is fashion in words, as in clothes. Not long ago ‘absolutely’ had its run. Where the simple ‘yes’ would have served, the interrogee would say ‘absolu-u-u-tely,’ that lute-like third syllable held with evident relish. Today writers seem to feel that if an article is not freely sprinkled with ‘tremendous,’ it will be wanting in ‘pep.’ Recently we came across a technical article—compact, well expressed and informing—in which the author used ‘tremendous’ no less than nine times. Surely the author is not so tremendously full of inspiration that he must use nine tremendous relief valves to let it all out.”

A Pension for Curie

“The twenty-fifth anniversary of the discovery with which Madame Curie's name is connected was dignified by the payment to Madame Curie of the first installment of a pension voted her by the French legislature, in recognition of her scientific achievements. It is interesting to quote the modest title under which her discovery was announced on December 26, 1898: ‘A note, by Monsieur and Madame P. Curie and Monsieur G. Bemont, upon a new radioactive substance found in pitchblende.’ This substance, of course, was radium.”

The Molasses Gasoline Test

“It has been discovered that molasses is extremely useful for detecting the presence of water in tanks of gasoline. Water is heavier than gasoline and will always sink to the bottom. A wooden stick is coated with molasses and is pushed down into the tank. The gasoline does not affect the molasses in any way, but when water is encountered, the coating comes away from the stick. When the stick is withdrawn it is not only possible to see whether water is present but the actual amount is plainly shown.”

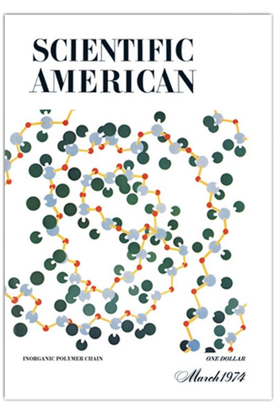
1874

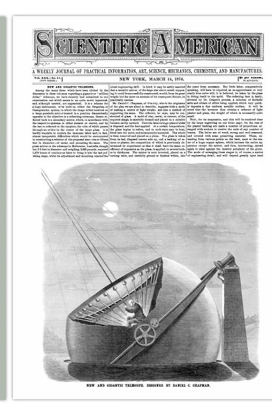
Hashish Users Addicted to Puns

“M. Naquet has lately been studying the physiological action of hachisch [hashish]. The extract of hemp seed (*Cannabis indica*) administered to various persons produces a great exuberance of ideation; it is not new ideas but the exaggeration, amplification and combination of ideas which pre-existed in the person's mind. Hachisch produces one curious effect (which is also observed in acute mania)—a singular inclination to make puns and plays on words.”

Cough Suppressant

“Coughing can be stopped by pressing on the nerves of the lip in the neighborhood of the nose. Sneezing may be stopped by the same mechanism. Pressing also right in front of the ear may stop coughing. Pressing very hard on the top of the mouth inside is also a means.”





Mark Fischetti has been a senior editor at *Scientific American* for 17 years and has covered sustainability issues, including climate, weather, environment, energy, food, water, biodiversity, population, and more. He assigns and edits feature articles, commentaries and news by journalists and scientists and also writes in those formats. He edits History, the magazine's department looking at science advances throughout time. He was founding managing editor of two spinoff magazines: *Scientific American Mind* and *Scientific American Earth 3.0*. His 2001 freelance article for the magazine, "Drowning New Orleans," predicted the widespread disaster that a storm like Hurricane Katrina would impose on the city. His video *What Happens to Your Body after You Die?*, has more than 12 million views on YouTube. Fischetti has written freelance articles for the *New York Times, Sports Illustrated*, *Smithsonian*, *Technology Review*, *Fast Company*, and many others. He co-authored the book *Weaving the Web* with Tim Berners-Lee, inventor of the World Wide Web, which tells the real story of how the Web was created. He also co-authored *The New Killer Diseases* with microbiologist Elinor Levy. Fischetti is a former managing editor of *IEEE Spectrum Magazine* and of *Family Business Magazine*. He has a physics degree and has twice served as the Attaway Fellow in Civic Culture at Centenary College of Louisiana, which awarded him an honorary doctorate. In 2021 he received the American Geophysical Union's Robert C. Cowen Award for Sustained Achievement in Science Journalism, which celebrates a career of outstanding reporting on the Earth and space sciences. He has appeared on NBC's Meet the Press, CNN, the History Channel, NPR News and many news radio stations. Follow Fischetti on X (formerly Twitter) @markfischetti

---

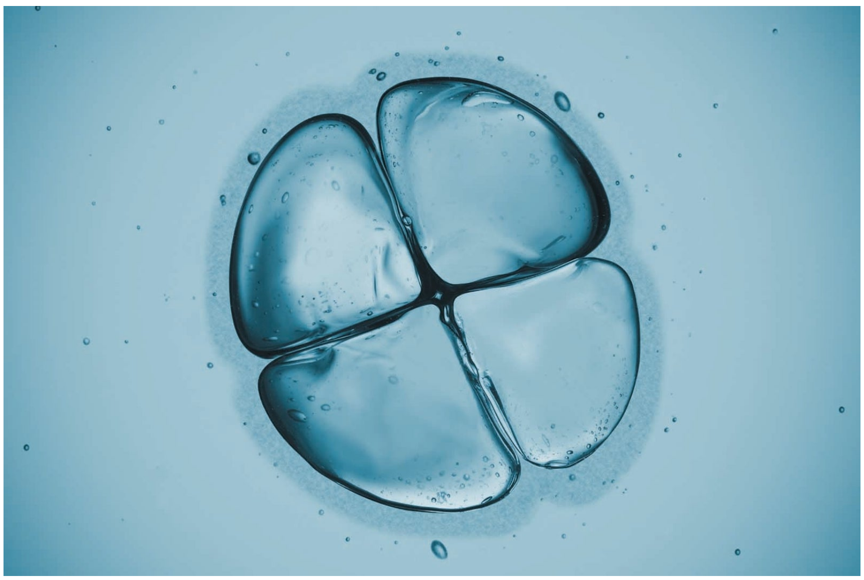
Microbiology

- ## Many Pregnancy Losses Are Caused by Errors in Cell Division
  Odd cell divisions could help explain why even young, healthy couples might struggle to get pregnant

Many Pregnancy Losses Are Caused by Errors in Cell Division

Odd cell divisions could help explain why even young, healthy couples might struggle to get pregnant

By Gina Jiménez



ugurhan/Getty Images

Human reproduction is notoriously inefficient. Whereas the fertilized eggs of other animals, such as mice, usually progress to more complex embryo stages, fertilized human eggs often falter early on. For a recent study in *Genome Medicine*, scientists analyzed almost 1,000 embryos from in vitro fertilization (IVF) procedures to learn why.

Scientists know that chromosomally abnormal human embryos—those with either more or fewer than 46 chromosomes—often don't implant in the uterus, and if they do, the resulting pregnancy may end in a miscarriage or stillbirth.

Some of these abnormalities originate in the egg or sperm. In other cases, a healthy egg and sperm form an embryo that divides oddly;

for example, instead of one cell dividing into two, it might become three. “There is a lot of evidence that during the first cell divisions, human embryos make a lot of mistakes,” says Claudia Spits, who studies related problems of reproduction and genetics at the Free University of Brussels.

The study authors found that these odd divisions lead to new chromosomal abnormalities, which can harm embryonic development even more than abnormalities that originate in the egg or sperm. The new errors can be “catastrophic,” says Johns Hopkins University evolutionary biologist Rajiv McCoy, the study's lead author. “A lot of times you have three, four, five missing chromosomes.”

McCoy and his colleagues used time-lapse video and a microscope to record the IVF embryos' first cell divisions. Then they tested for abnormalities among both the surviving embryos and those that failed.

Spits says the effects of postfertilization abnormalities are “something that we have all assumed to be true, but nobody provided the evidence in the manner that [McCoy and his team] have done” by analyzing such a high number of embryos and including discarded ones.

The new study also found that division-related errors are equally common among embryos with eggs from women of all ages, whereas egg and sperm abnormalities increase as people age. The researchers suspect this finding might help explain why it is so hard even for many young, healthy couples to get pregnant. “Maybe some of the mechanisms that we are uncovering from our studies will be relevant to understanding the low level of human fertility,” says study co-author Michael Summers, a reproductive medicine consultant at London Women's Clinic. The work could also help illuminate what McCoy calls “the black box of early pregnancy loss.”

The scientists say they hope to use their results to improve IVF. For instance, if changing the cells' environment reduces the likelihood of dividing errors, Summers says, “you could potentially rescue a lot of embryos for IVF purposes because those errors are happening in the dish.”

Gina Jiménez is a bilingual journalist focused on health and science policy and how they affect disadvantaged communities. Follow her on X (formerly Twitter) @GinaRivers90

Particle Physics

Can a Magnet Ever Have Only One Pole?

Electron tornadoes that mimic "magnetic monopoles" emerge from specks of rust

By Zack Savitsky

Thomas Fuchs

Magnets are notoriously codependent. Try to break apart a magnet's north and south ends, and each half gets its own fresh set of two poles. Scientists have long hunted for a lone north or south pole—an individual particle carrying solely a positive or negative magnetic charge. Although such "magnetic monopoles" remain elusive, some have begun searching for virtual ones—clusters of electrons that behave like single magnetic charges.

Rather than searching for single particles, "we're using the creativity card that we have in condensed-matter physics ... to redefine new building blocks," says Mete Atatüre, a physicist at the University of Cambridge. In a study published in *Nature Materials*, Atatüre and his colleagues have captured the first direct observation of magnetic monopoles that emerge naturally from the collective behavior of electrons. The researchers hope these objects could one day enable a more energy-efficient method for storing computer information.

Electrons in solid materials behave like tiny bar magnets; the strength and orientation of their magnetic fields are defined by a quantum property called spin, which acts like an atomic compass needle. Working in concert, the spins of many neighboring electrons can form particular patterns that appear as isolated regions of positive or negative magnetic charge. For the past 15 years scientists have been hunting for these monopolelike features emerging in various materials but had mustered only indirect evidence.

In the new study, Atatüre and his team employed a new sensing technique that measures how tiny magnetic fields alter a single electron's spin at the fine tip of a diamond. They ran their detector over a freckle-sized sample of hematite, the primary component of rust, to map the texture of electron spins on its surface. While varying the sample's temperature, they were surprised to find that the spins spontaneously organized into whirlpool shapes that acted like magnetic monopoles—single positive or negative charges without partners.

“The measurements of these magnetic fields coming in and coming out, that's remarkable,” says Ludovic Jaubert, a theoretical physicist at University of Bordeaux who studies monopole features in other materials. “Once you can visually see these things, it's much easier to manipulate [them] to study further.”

These emergent features don't solve the enduring mystery of whether a magnet's poles can be fundamentally separated, but they may still prove valuable. Scientists have proposed that the pirouetting electron spins could be used to encode and transfer information in computers more efficiently than current methods, which typically rely on electrical charges that take more energy to move and sustain. Finally, spotting these quantum tornadoes is an important step toward building a new generation of electronics, Jaubert says. “It's really quite beautiful.”

Zack Savitsky is a freelance science journalist specializing in physics and astronomy.

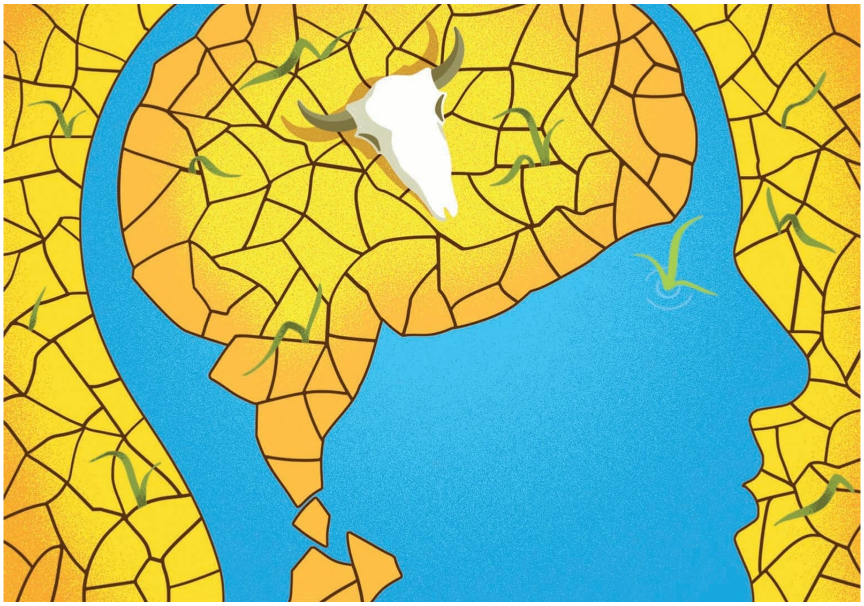
Psychology

- **Water Scarcity Changes How People Think**
  Lacking money makes people focus on the present—but lacking water makes them plan for the future
- **Training Bartenders, Barbers and Divorce Attorneys as Counselors Could Reduce Gun Suicides**
  Some of the tens of thousands of suicide deaths in the U.S. each year would not have happened if people in the community had been schooled to provide advice about gun safety

Water Scarcity Changes How People Think

Lacking money makes people focus on the present—but lacking water makes them plan for the future

By Thomas Talhelm & Hamidreza Harati



Chris Gash

The human brain is sensitive to scarcity. The lack of something we consider vital, such as time or food, can powerfully shape our thinking and behavior. Take money, for example: when people play a game that makes some players abruptly wealthier or poorer, those who lose money start making decisions that result in their being better off now but worse off later.

One potential explanation for this tendency is that scarcity is mentally taxing. The sense of not having enough of what we need becomes a distraction that makes it harder for us to focus and plan.

In recent work, however, we found that people react to water scarcity very differently than they do to other shortages. In studies of people around the world, we found that those in places where water is scarce think more about the long term, and confronting people with water scarcity in a laboratory setting made them focus on the future. Worrying about water shortages prompted better planning and less wasteful behavior—not only with water but with other resources, too.

As a species that is acutely dependent on water—without it we would die within days—we seem to be wired to plan for it. The loss of water prompts a general mindset of thrift and long-term thinking. This discovery could have important implications for humankind's response to climate change.

We explored people's thinking about water in several ways. In one experiment, we brought 211 college students to the lab and asked a portion of them to read an article we provided. Some received an article that was dire in tone. It detailed predictions of the worst “water shortage in 1,200 years.” Others read a more upbeat article that described how climate change means more rain and therefore ample water. (The latter article focused on the positives of water availability and did not discuss, for example, the threat of floods.) The rest of the students completed the study without reading an article.

We then asked the students how important it was to save resources and think about the long term. The mere idea of water scarcity was enough to kick-start them into thinking about the future. In a survey, participants who read about a serious shortage agreed more with statements such as “People should live for the future.” They also agreed more with ideas about saving, such as “There are things I resist buying today so I can save for tomorrow.”

In contrast, people who read about a future water surplus were pushed in the opposite direction. Compared with those who had answered the questions without reading about climate change at all, they were more likely to endorse living for the present and cared less about saving money and other resources.

Of course, simply reading about water scarcity might have different effects than actually experiencing it, so we looked for places where communities have experienced real water scarcity for generations. We found two geographically close cities in Iran—Shiraz and Yazd—that have similar economies and the same majority religion, language and ethnicity but sharply different amounts of water. Shiraz receives enough rainfall to support vineyards that produce its world-famous wine. Nearby Yazd is bone-dry. We gave psychological tests to 331 people in Shiraz and Yazd to measure what psychologists call “long-term orientation,” or how much individuals prioritize the future. Sure enough, people in Yazd thought that planning for the future was more important than those in Shiraz did, and people in Shiraz liked the idea of living in the moment more than those in Yazd did.

Then we went a step further. Iran is particularly vulnerable to drought, so perhaps people there are more sensitive to water scarcity than populations in other places. To get some sense of whether similar patterns exist elsewhere, we turned to the World Values Survey, a long-running global research project that gathers information about people's beliefs and values. We focused on survey data from respondents in 87 countries who were asked about the importance of thrift and saving for the future. Those who live in countries with a history of water scarcity tended to agree more with saving for the future. For example, in Europe, people in water-rich Iceland thought less about the future, whereas those in dry Spain thought more about the future.

Moreover, countries' history of water scarcity seemed related to cultural differences in a way that more obvious factors didn't. For example, income per capita did not explain differences across cultures. And although corruption might make it hard for people to think about the future, it was not a strong predictor, either. You might guess that people think about the future more in places where they tend to live longer, but astonishingly, national life expectancy was not as strong a predictor as water scarcity.

Across studies, our findings suggest that water has a meaningful place in our thinking—one that's distinct from other important resources, including wealth. In fact, humans may have evolved to be keenly sensitive to water access because it's so critical to us. Humans managed without money for many thousands of years. Our species can endure without food for weeks but without water for just days.

There is evidence of evolutionary wiring for water in our sense of smell. Humans are better at detecting the scent of fresh rainfall than sharks are at detecting blood. The fact that water is essential for human life makes it more plausible that evolution would lead to our having psychological reactions specific to the threat of water scarcity.

That sensitivity might be crucial going forward. Climate change is making droughts more common. Our work suggests that as many places dry up, global warming could reshape how people think, pushing entire communities toward more cautious, future-oriented behavior. That may offer a sliver of hope amid the threat of climate change. Water scarcity could be uniquely powerful motivation for us to prepare for and respond to a warmer world.

Are you a scientist who specializes in neuroscience, cognitive science or psychology? And have you read a recent peer-reviewed paper that you would like to write about for Mind Matters? Please send suggestions to Scientific American*'s Mind Matters editor Daisy Yuhas at dyuhas@sciam.com.*

This is an opinion and analysis article, and the views expressed by the author or authors are not necessarily those of Scientific American.

Thomas Talhelm is an associate professor of behavioral science at the University of Chicago Booth School of Business. He researches cultural differences and the effects of historical rice farming on southern China's culture, compared with that of the country's wheat-farming north.

Hamidreza Harati completed his Ph.D. in organizational behavior at the University of Queensland Business School in Australia. He studies cultural differences and innovative teaching.

Training Bartenders, Barbers and Divorce Attorneys as Counselors Could Reduce Gun Suicides

Some of the tens of thousands of suicide deaths in the U.S. each year would not have happened if people in the community had been schooled to provide advice about gun safety

By Sara Novak

Shideh Ghandeharizadeh

Historically, suicide prevention has focused on the mental health risk factors that might lead an individual to want to die. But despite the intuitive appeal of this approach, it isn't working. That is the opinion of Michael Anestis, executive director of the New Jersey Gun Violence Research Center. In 2022 there were 26,993 deaths by gun suicide in the U.S., and the rate of such suicides reached an all-time high, increasing by 1.6 percent from a year earlier. Our history of leaning on risk factors hasn't made us any better at predicting who will die, Anestis says.

That's why researchers and suicide prevention advocates have taken a different approach: making the surrounding environment safer so those at risk (whether they know it or not) are less likely to die by suicide. A similar policy in Israel brought about a 57 percent reduction in the suicide rate within the military. According to a 2016 article in the journal *European Psychiatry,* the change came from not just increased mental health awareness at work but also the behavioral measure of stopping people from taking their guns home when they were off duty.

Anestis thinks we could see comparable results in the U.S. From 2012 to 2020 he lived and worked in southern Mississippi, the state with the fourth-highest rate of gun deaths. Spending eight years in the Deep South made him realize that he had to find a way to reach those whom he cared for deeply even though their views about guns were much different from his own.

People who own firearms don't want to accidentally get hurt or hurt others, he says, but they view the risk that firearms pose to their owners as one worth taking. Still, Anestis contends that common ground for widespread secure storage measures is possible. Research that Anestis and his colleagues published in the February 2021 issue of the *American Journal of Public Health* showed that “lethal means counseling” for gun owners resulted in wider adoption of safe storage methods.

Scientific American spoke with Anestis about a training program he is leading called Project Safe Guard, which provides neutral figures—such as heads of military units, barbers and faith leaders—with the tools to educate firearm owners about safety measures for storing their weapons, especially in times of despair.

An edited transcript of the interview follows.

People who weren't trained as psychologists have long spoken to callers on suicide hotlines, but now a new, broader approach to suicide prevention involves training those people to do what's called "lethal means counseling." Can you explain the program?

Project Safe Guard is about training not only clinicians but also community members to talk with firearm owners about the ways they can store their firearms securely and the circumstances in which they should consider doing so. The idea behind it is to make the environment safer so that when someone is in a difficult place, they're less likely to have click-and-ready access to their firearms.

The training involves peer-to-peer counseling where individuals such as barbers and faith leaders talk with people who may be more likely to come to them in a moment of crisis. The training may also involve police and military leaders training their subordinates on how to safely store a firearm either in their homes or using outside storage facilities. We're trying to change social norms on both a micro and a macro level using credible messengers. At the same time, we're strategically training those who tend to talk to people in their most difficult moments, giving them the tools to have a reasonable, persuasive conversation with those in need.

How do you choose the types of people to train in the program?

In our first go-around, in 2023, we planned on training faith leaders and barbers. These are folks who are generally not seen as having political agendas. They're well suited to talk to people in moments of distress, and they're often trusted with personal information. Even people who are more likely to keep their pain to themselves may open up to a faith leader or a barber. In the future, we're hoping to train divorce attorneys and bartenders for many of the same reasons listed earlier.

We want people to learn to have conversations about this that don't feel awkward or political and don't resemble a public service announcement. It's all in an effort to shift social norms for how people think about their firearms. For this to happen, they need to encounter the message of secure storage from a number of convincing sources in multiple contexts for there to be an internal shift in beliefs. We're normalizing changes using people who don't feel like outsiders coming in and telling gun owners what to do.

What are some of the techniques that can be used to connect with gun owners and open their eyes to the importance of safe storage?

We use an approach called motivational interviewing, an intervention that works within a person's value system to leverage their intrinsic motivation and make positive changes in their life. Some people don't want to change, and you can't make them, but the idea is to avoid conflict, which is really important for a cultural and political issue such as firearms.

Individuals are taught to ask open-ended questions to initiate a conversation around firearm storage: “How do you store your firearms?” “What do you use or not use, and what are your reasons for it?” “Are there any circumstances in which you think it might make sense to not have quick access to your firearms?” If they respond with “I haven't really thought about that,” you might say, “What if there are kids in the home, or what if you've been drinking, or what if you haven't been feeling quite like yourself lately? Are those situations when you might consider storing your firearms a bit more securely?" It's about starting a conversation and seeing the places where a firearm owner might be willing to make changes.

You write that those who die when using firearms are less likely to engage the health-care system. Can you discuss this?

The data are pretty clear that those who die by firearms are less likely to have sought mental health care near the time of their death compared with folks who die by suicide using other methods. It's very common for those around that person to say that they never saw it coming because the person who died kept their feelings to themselves. We've got this problem in the U.S. where those who are most likely to die by firearm suicide aren't telling anyone what they're thinking, which makes it more difficult to help them. Project Safe Guard is an opportunity to reach this group in a way that mental health services seem to be falling short on. We don't have a whole lot of data on why these people don't seek care, but we think it comes from traditionally masculine ideas about solving your own problems and not openly discussing feelings, as well as a certain level of distrust in the health-care system and mental health care in general.

What are the next steps in training people?

We're planning on doing large-scale training sessions in New Jersey this year with faith leaders and barbers. And we also have plans to integrate the U.S. Army and the National Guard. Additionally, a former student of mine, Claire Houtsma, a suicide prevention coordinator at the Southeast Louisiana Veterans Health Care System, is training veterans to engage in these conversations with their peers. There are also other approaches to counseling beyond our program, such as Counseling on Access to Lethal Means (CALM), which is a training course directed at health-care and social service workers. Our end goal is to take this as far as it will go by getting it in front of people's eyes enough times that it has the potential to develop its own momentum.

IF YOU NEED HELP

If you or someone you know is struggling or having thoughts of suicide, help is available. Call or text the 988 Suicide & Crisis Lifeline at 988 or use the online Lifeline Chat.

Sara Novak is a science writer based on Sullivan's Island, S.C. Her work has appeared in *Discover*, *Sierra Magazine*, *Popular Science*, *New Scientist*, and more. Follow Novak on X (formerly Twitter) @sarafnovak

---

Sleep

- ## Is Snoozing the Alarm Good or Bad for Your Health?
  New research suggests that hitting the snooze button to squeeze in an extra five or 10 minutes of sleep may actually be good for you

Is Snoozing the Alarm Good or Bad for Your Health?

New research suggests that hitting the snooze button to squeeze in an extra five or 10 minutes of sleep may actually be good for you

By Jocelyn Solis-Moreira

Peter Dazeley/Getty Images

When your alarm goes off in the early morning, it's tempting to hit the snooze button and curl back up under the warm covers for a few more minutes of slumber. This repeated postponing of the buzzer is often thought of as a bad habit—one that creates not only a lazy start to a day but also a fragmented sleep pattern that's detrimental to health. But recent research is contradicting this notion.

A recent study in the *Journal of Sleep Research* found that people who regularly pressed the snooze button lost only about six minutes of sleep per night—and it didn't affect their morning sleepiness or mood. Plus, tests showed that 30 minutes of snoozing improved or did not affect cognition compared with people who woke up the first time their alarm went off. This adds to research in 2022 that also found chronic snoozers generally felt no sleepier than nonsnoozers.

“Snoozing for a limited time in the morning is probably not bad for you,” says the study’s lead author, Tina Sundelin, a sleep researcher at Stockholm University. She says her study is one of few that have directly tested snoozing’s effect on sleep health, and it offers evidence that snoozing doesn’t break up sleep in a harmful way.

Scientific American spoke with sleep experts on the science of snoozing and how the habit may actually be good for you—if you do it right.

The Potential Benefits of Snoozing

Snoozing does shorten sleep, Sundelin says, but she maintains that it’s not as bad as scientists once thought. Past research suggested that the extra minutes snoozers get don’t really help them feel more rested—and repeatedly waking up and trying to sleep again has been thought to prevent the restorative stages of sleep, including rapid eye movement (REM) sleep. “If you disturb someone’s sleep, it’s not good-quality sleep, and they often feel tired afterward—but this [idea] is based on a whole night of sleep fragmentation,” explains Sundelin, who adds that most theories about snoozing are “inferred from what we know about sleep in general.”

Sundelin found that snoozing the alarm for half an hour benefited chronic snoozers—people who delay the alarm two or more times a week and almost always fall back asleep between alarms. Thirty-one such chronic snoozers who were observed in the study slept well throughout the night and showed signs of fragmented sleep only in the last 30 minutes before getting up, which is typically around the time when people first hit the snooze button. This fragmented sleep "didn't have a big enough impact to make them tired" throughout the rest of the day, Sundelin says.

Sundelin's research also suggests that snoozing might help people shake off morning drowsiness by easing the transition from deep sleep to a lighter stage. A good night's rest typically involves four to six sleep cycles, each made up of four stages. Light sleep happens in the first two stages of nonrapid eye movement (NREM) sleep. This is when muscles start to relax, and brain activity slows, along with breathing and heart rate—but a person can still be easily woken. As the night goes on, people progressively reach deeper stages. It gets harder to wake up during the third and final stage of NREM sleep and the first stage of REM sleep. A person who receives a phone call during these stages, for example, might be less likely to hear it or remember answering.

Abruptly waking up, especially from deep sleep, can prolong sleep inertia—a drowsy state of transitioning to wakefulness in which one may feel disoriented or struggle to adjust to being awake. This is where snoozing may help, Sundelin says: people who squeeze in little naps between alarms can more effectively shift out of deep sleep and wake up during lighter sleep. This might help them decrease sleep inertia and feel more alert and energetic in the morning.

The additional light slumber might also aid cognition, Sundelin's results show. The snoozers were alert enough to perform well on cognitive tests of processing speed, episodic memory and executive functioning, as well as simple arithmetic. Another test showed that these benefits continued for up to 40 minutes after a person woke up.

Sundelin hypothesizes that snoozing prevents people's brains from quickly reverting to deeper sleep stages. Snoozers also showed higher levels of cortisol, a hormone involved in wakefulness, immediately on awakening compared with people who slept uninterrupted the entire night. Although the difference in cortisol levels evened out over time, Sundelin says that "it's possible the cortisol awakening response started slightly earlier [in snoozers], and this could have helped the brain to finish these tasks."

The relation between snoozing and cognition needs further research, says sleep scientist Cassie Hilditch of San José State University, who was not involved in the recent study. She notes that other work has reported different effects: an increased sleep inertia and worse cognitive performance in snoozers.

Snoozing Impacts People Differently

Young adults typically press snooze more often than older ones. Thomas Kilkenny, director of the Institute of Sleep Medicine at Staten Island University Hospital, says that adults in their early to mid-20s tend to stay up late and get less sleep overall. Some people's biological clock—a built-in 24-hour cycle that helps the body regulate processes including wakefulness and sleep—tends to shift toward a "night owl" chronotype during adolescence, reaching a peak "lateness" around age 20. "Partly the fault is that the body just doesn't want to go to sleep until one o'clock in the morning, even if you have to get up at six to go to school," Kilkenny says.

Night owl–types, regardless of age, are also more likely to hit snooze. Those with late chronotypes often feel their best in the evening and prefer going to bed closer to midnight. Given that school and work typically start early, however, night owls often have to wake up when they are least alert compared to “morning larks” (people with an early chronotype), Hilditch says. She adds that sleep inertia can be worse when waking up closer to one’s circadian low, a time when alertness is bottoming out. “Therefore, later chronotypes may find it harder to wake up for early classes and workdays given the increased sleep inertia during this time,” Hilditch says.

How to Get the Most Out of Snoozing

The optimal period to spend snoozing—for chronic snoozers who aren’t sleep-deprived—is somewhere between 20 and 30 minutes, says Kilkenny, who suggests this is enough to be “refreshing but not too much.” It’s equivalent to hitting the snooze button every five to 10 minutes for a total of three or four times and is probably enough to overcome sleep inertia, which usually lasts 30 minutes or less for someone who isn’t sleep-deprived. Additionally, snoozing for more than half an hour can inch a person closer to the deeper phases of sleep from which it’s harder to get up. This is why people who say they’ve “overslept” sometimes feel groggy or disoriented.

So snoozing can give the body some time to adjust and prepare to get out of bed. This may come in handy for things such as adjusting to the beginning of daylight saving time, when many people lose an hour of sleep, Kilkenny says. But people who regularly wake up without an alarm or who get up the first time it goes off may not get those same benefits if they snooze. Kilkenny says that’s because the body has already had enough time to fully rest.

There is one caveat: snoozing can never replace a good night’s sleep. People who feel they need to snooze for more than 30 minutes and who have difficulty waking up after that may be showing signs of sleep deprivation. If that’s the case, the problem won’t be solved with the touch of a button—and snoozing might, in fact, make things worse. Waking tactics such as the slow-to-rise method (gradually shifting wake-up time 10 to 15 minutes earlier every few days) could help some people—but only those who are already getting enough sleep, Kilkenny says.

There is still much to learn about snoozing's long-term impact on cognition and the brain. But the new research is a helpful step toward dispelling some of the "lazy" stereotypes often associated with this common morning ritual. So regular snoozers can feel less guilty for catching some extra z's while hitting the alarm button tomorrow morning.

A version of this article entitled "Snooze Button Benefits" was adapted for inclusion in the March 2024 issue of Scientific American. *This text reflects that version, with the addition of some material that was abridged for print.*

Jocelyn Solis-Moreira is a freelance journalist covering health and science. She is based in New York State.
