Inventor biases create invention biases pp. 1260 & 1345

Soft-furred tree mice found to be echolocators p. 1305

Making LIGO even cooler p. 1333

Science

$15
18 JUNE 2021
SPECIAL ISSUE
sciencemag.org
AAAS

FALLBACK STRATEGIES

Planning for climate-induced relocation p. 1274

REVEALING THE NEXT BREAKTHROUGH IN IMMUNOLOGY

APPLY NOW FOR THE $30,000 GRAND PRIZE

APPLICATIONS DUE: OCTOBER 1, 2021

The Michelson Philanthropies & ***Science*** Prize for Immunology recognizes research that will have a significant impact on vaccine and immunotherapy development. Together with AAAS/***Science***, Michelson Philanthropies awards breakthrough research from the past three years. The applications are judged on innovation and potential impact.

With these prizes, scientists (35 and younger) have the opportunity to advance their research projects, are provided with a renowned platform in ***Science*** to publish their work and connect them with other innovators through the Michelson Prizes alumni network.

Young researchers from a wide range of disciplines are encouraged to apply.

APPLY TODAY:
www.sciencemag.org/michelson

MICHELSON PHILANTHROPIES & Science
PRIZE FOR IMMUNOLOGY

CONTENTS

18 JUNE 2021 • VOLUME 372 • ISSUE 6548

Children walk to their new home in Mertarvik, Alaska. Soon, their grandparents' village of Newtok will no longer be suitable for living, owing to permafrost melt and erosion.

PHOTO: VLAD SOKHIN/PANOS PICTURES/REDUX

SPECIAL SECTION

Climate-induced relocation

INTRODUCTION

POLICY FORUMS

REVIEW

ON THE COVER

An abandoned mosque sits outside a sea wall in Pluit, Jakarta, Indonesia. Confronted with both sea level rise and subsidence due to groundwater extraction in this city of more than 10 million people, the government has announced plans to relocate the national capital to the island of Borneo. Research has much to offer communities around the globe that are facing displacement from climate change—but community members must be involved, their voices heard. See page 1274. *Photo: © Kadir van Lohuizen/NOOR*

SEE ALSO EDITORIAL p. 1245
NEWS STORY p. 1254 EDITOR'S BLOG

NEWS

IN BRIEF

IN DEPTH

FEATURES

INSIGHTS

PERSPECTIVES

BOOKS *ET AL.*

LETTERS

RESEARCH

IN BRIEF

RESEARCH ARTICLES

REPORTS

DEPARTMENTS

SCIENCE (ISSN 0036-8075) is published weekly on Friday, except last week in December, by the American Association for the Advancement of Science, 1200 New York Avenue, NW, Washington, DC 20005. Periodicals mail postage (publication No. 484460) paid at Washington, DC, and additional mailing offices. Domestic individual membership, including subscription (12 months): $165 ($74 allocated to subscription). Domestic institutional subscription (51 issues): $2148; Foreign postage extra: Air assist delivery: $98. First class, airmail, student, and emeritus rates on request. Canadian rates with GST available upon request, GST #125488122. Publications Mail Agreement Number 1069624. **Printed in the U.S.A.**
Change of address: Allow 4 weeks, giving old and new addresses and 8-digit account number. **Postmaster:** Send change of address to AAAS, P.O. Box 96178, Washington, DC 20090–6178. **Single-copy sales:** $15 each plus shipping and handling available from backissues.sciencemag.org; bulk rate on request. **Authorization to reproduce** material for internal or personal use under circumstances not falling within the fair use provisions of the Copyright Act can be obtained through the Copyright Clearance Center (CCC), www.copyright.com. The identification code for *Science* is 0036-8075. *Science* is indexed in the *Reader's Guide to Periodical Literature* and in several specialized indexes.

PHOTO: QI LIU AND XIAO-FENG MA

EDITORIAL

Decolonize climate adaptation research

Climate-forced population displacement is among the greatest human rights issues of our time, presenting unprecedented challenges to communities and the governments responsible for protecting them. Sea level rise, heat, drought, and wildfires will cause people to move, losing homes and places they love, often with no ability to return. Indigenous Peoples have done the least to cause this crisis and face the loss of lands and connections to ancestral, cultural, and spiritual heritage. To ensure that their right to self-determination is protected and the horrific legacy of government-forced relocations is not repeated, communities must lead and define research on climate-forced displacement and managed retreat that involves them and the lands upon which they dwell and subsist. A focus on human rights, and decolonization of research to change institutional structures of knowledge production, can help communities define their future in a climate-altered world.

The government responsibility to protect people may require relocation against peoples' will. Determining which communities are most likely to encounter displacement requires sophisticated assessment of the vulnerability of a community's ecosystem, but also its social, economic, and political structures. Human rights principles, which include rights to food, to safe and sanitary housing, and to water, must be embedded in any relocation governance framework. The right to self-determination ensures that communities make the decision of whether, when, and how relocation will occur and that cultural and spiritual heritage is protected if relocation is the best strategy.

Human rights principles also ensure that racial and economic inequities, legacies of colonization and slavery, are addressed when responding to climate-forced displacement. Scholars continue colonization when Indigenous Tribes are not represented in, or consulted for permission to do, research on their communities and lands. Decolonization is the restoration of cultural practices, spirituality, and values that were taken away or abandoned through colonization and that are important for survival, well-being, and subsistence lifestyles. Decolonization advances and empowers Indigenous Peoples and stops perpetuating their subjugation and exploitation.

> "...communities must lead and define research..."

Indigenous-led research can help determine whether inclusion of human rights protections averts or minimizes severe consequences associated with government-mandated relocation. For example, in a letter to the US National Science Foundation expressing concerns with its Navigating the New Arctic program, four Alaska Native organizations explained the danger and damage to their communities when outside academics define food security, resilience, and adaptation, highlighting the importance of Indigenous scholarship and voices in research.*

Self-determination and decolonization mean that communities control the narrative about how the climate crisis affects them. Colonization continues when non-Indigenous scholars write narratives about "vanishing cultures." The Alaska Native Science Commission and Inuit Circumpolar Council provide a promising model, having protocols that ensure Indigenous communities lead research efforts, defining the questions and methodologies. Non-Indigenous scholars need to build relationships and trust with Tribes before submitting funding applications to understand how skills offered by academic researchers can benefit and complement skills and expertise of Indigenous knowledge holders.

Community-based environmental monitoring, and coproduction of knowledge, are important decolonizing tools that can facilitate empowerment and capacity building. Community-based monitoring is important to understand local ecosystem change, which is critical to implementing community-based adaptation strategies; global, regional, and national climate change assessments generally aggregate information above the level of resolution required for effective community policy.

We reflect here on our experience in the North American Arctic and Subarctic, but such issues arise in communities around the globe. Countries such as Kiribati and Maldives face inundation from sea level rise, possibly leaving residents stateless. Sea level rise and extreme weather threaten lives and livelihoods in coastal communities in Egypt, Panama, and elsewhere. Research must support and build the capacity of Indigenous Tribes and local communities so that they have tools to respond dynamically to support adaptation that protects their human rights.

–Robin Bronen and Patricia Cochran

Robin Bronen
is executive director of the Alaska Institute for Justice and a senior research scientist at the Institute of Arctic Biology, University of Alaska Fairbanks, Fairbanks, AK, USA.
robin.bronen@akijp.org

Patricia Cochran
is executive director of the Alaska Native Science Commission, Anchorage, AK, USA.
pcochran@aknsc.org

*M. Bahnke, V. Korthuis, A. Philemonoff, M. Johnson, Letter to "Navigating the New Arctic Program, National Science Foundation," 19 March 2020; https://kawerak.org/download/navigating-the-new-arctic-program-comment-letter/.

10.1126/science.abi9127

> "Should be on the 'must read' list of every world leader."
>
> **Climatologist Mark Maslin** of University College London, about the first joint report on links between controlling climate change and protecting species diversity from the Intergovernmental Panel on Climate Change and the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services.

IN BRIEF

Edited by **Jeffrey Brainard**

COVID-19

New vaccine, pledges boost global effort

Good news from a large efficacy trial of a protein-based vaccine and expanded pledges of donations by G7 nations have raised hopes that poorer countries will have more access to shots that could beat back the coronavirus pandemic. Novavax announced on 14 June that its COVID-19 vaccine, which delivers the spike protein of SARS-CoV-2, had 90% efficacy overall and 91% efficacy in people at high risk because of age, exposure, or medical conditions. Results from the trial of 25,000 people in the United States and Mexico also indicated the vaccine completely prevents moderate and severe disease. Novavax plans to file for emergency use authorizations in the United States and elsewhere by September. The small U.S. company and the Serum Institute of India last month agreed to sell 1.1 billion doses to a global vaccine program set up to help poorer nations and address the critical shortage of COVID-19 vaccines worldwide. Separately last week, the G7 group of nations agreed to provide poorer countries with an additional 870 million vaccine doses, with half delivered by the end of this year. But some public health experts want to see the donations happen faster.

Lab heads yanked from NIH grants

#METOO | Since 2018, the National Institutes of Health (NIH) has received more than 300 complaints of sexual and other harassment, and through April has approved institutions' requests to remove 75 principal investigators (PIs) from grants in response to validated allegations, the agency reported last week. About two-thirds of cases were sexual harassment complaints, whereas others involved bullying, racial discrimination, or other harassment. The accused scientists' institutions investigated and validated the claims in 26% of 268 cases closed. Overall, 61 PIs involved in validated harassment claims left their institutions. A larger number of NIH-funded scientists, 125, have been barred from serving as peer reviewers for NIH grants; some of them were also among the 75 removed from grants.

Biden acts on air, water, forests

ENVIRONMENT | President Joe Biden's administration last week moved to undo three of former President Donald Trump's signature environmental policies. The Environmental Protection Agency (EPA) announced it will reconsider limits on soot, air pollution composed of fine particles less than 2.5 microns wide. In 2020, the Trump administration declined to reduce allowable soot levels for industrial emissions, despite evidence of health risks and urging from some EPA scientific advisers. EPA also said last week it will toss Trump's rewrite of wetland protection rules, which lifted regulatory controls on many small marshes and ephemeral streams, and start to write updated ones. And the U.S. Department of Agriculture said it will restore protections to 3.6 million hectares of Alaska's Tongass National Forest that Trump had removed to permit logging and road construction. Each effort will likely take years to complete and could face court challenges.

Drug approval sparks resignations

DRUG DEVELOPMENT | Three members of a 10-person advisory committee resigned last week over the U.S. Food and Drug Administration's (FDA's) decision to approve the Alzheimer's drug aducanumab against their recommendation. The agency's verdict made aducanumab (Aduhelm) the first U.S. drug to target the underlying disease process of Alzheimer's—specifically, the brain buildup of amyloid protein—but the antibody showed modest cognitive benefits in one large clinical trial and no benefit in another (*Science*, 11 June, p. 1141). The resignation letter of panelist Aaron Kesselheim of Harvard Medical School called FDA's move "probably the worst drug approval decision in recent U.S. history" and warned that the agency had undermined patient care and public trust.

Europe to probe Venus volcanoes

PLANETARY SCIENCE | The European Space Agency last week announced it will send an orbiter to Venus in 2031 to search for volcanoes that may help feed the furnacelike temperatures on Earth's cloud-wrapped twin. The €610 million EnVision will rely on infrared spectrometers and radar to peer through the murky atmosphere and try to find hot spots on the surface and lava flows that could show current or recent volcanism. The orbiter's design and mission will complement those of two Venus probes that NASA plans to launch by the end of this decade. Compared with Mars, Venus has had fewer visits from robotic spacecraft. But increased interest in climate change and Earth-like exoplanets has prompted researchers to ask why Venus is now a greenhouse inferno with a sulfuric acid atmosphere, after starting out similarly to Earth.

Rigor sought in animal studies

BIOMEDICINE | The U.S. National Institutes of Health (NIH) should shore up the rigor, transparency, and clinical usefulness of

The magnet segment (bottom) is the first of six to be shipped to the ITER fusion reactor.

PHOTO: ITER

ENERGY

U.S. magnet heads to fusion reactor in Europe

A powerful magnet that will form part of the central solenoid of the ITER fusion reactor begins the journey from its manufacturer in San Diego to the reactor site in southern France this week. ITER aims to show that a fusion reaction of hydrogen nuclei could be a viable source of energy. The 18-meter-tall solenoid, which will stand in the central hole of the massive, doughnut-shaped reactor, is designed to help heat the ionized hydrogen fuel and drive it to a current of 15 million amps that flows around the doughnut, creating a magnetic field that helps confine the superheated gas. The solenoid, which incorporates 27 kilometers of superconducting niobium-tin wire, was manufactured by General Atomics and is a major U.S. contribution to the international project. Construction is scheduled to be completed in 2025.

animal studies it funds, an NIH working group said last week. The group's recommendations—aimed at studies using vertebrates and cephalopods—include adding a page to the current 12-page grant proposal to include details such as the number of animals to be used and plans for data analysis. Funded researchers should include similar details in their manuscripts so the work can be reproduced, the group said in its 11 June report. And NIH should explore requiring investigators to preregister their plans by posting them in a database before experiments begin, it said. More rigorous experiments may drive up costs by requiring more animals, but they could also reduce the need to redo small studies that fail to produce statistically robust results, says molecular biologist Barbara Wold of the California Institute of Technology, who co-chaired the group.

IN OTHER NEWS

EUROPE OPENS GRANTS The European Commission said this week it would permit some non-EU countries to compete for research grants in quantum computing and space science funded by Horizon Europe, a new €95.5 billion program. After seeking to bar Israel, the United Kingdom, and other non-EU countries that pay to access Horizon Europe from the strategically sensitive research areas, the European Commission compromised by opening them up to these countries—provided they give as-yet-undefined "assurances," which could include restrictions on the transfer of intellectual property.

THIRD DOSES A third dose of COVID-19 vaccine stimulates the production of protective antibodies in organ transplant patients who previously had none or low levels, researchers reported this week. Eight out of 24 such patients generated the antibodies when they had none after a two-dose regimen. And six people with low antibodies after two doses had high levels after a third, the researchers report in the *Annals of Internal Medicine*.

WHALE TESTING A first-of-its-kind study of captive baleen whales' hearing started last week in Norway despite criticism from other scientists that restraining the animals may harm them. The research aims to measure how well juvenile minke whales can hear human-generated noise—findings that could inform restrictions on offshore oil exploration and other activities that generate loud underwater sounds blamed for impairing some cetaceans. Critics wrote Norway's prime minister to urge the study's cancellation. The researchers say their plan minimizes risks to the animals.

#METOO LOCKOUT Harvard University last week stripped prominent anthropological archaeologist Gary Urton of emeritus status and banned him from all its events and buildings on campus because of findings of "persistent" sexual harassment by him. Urton, 72, was anthropology department chair from 2012 to 2018 before he was put on administrative leave in June 2020. He retired that year, before an investigation into his behavior ended. Students said Urton had sexually harassed them or abused his authority to coerce them into affairs. In an email response to *Science*, he wrote, "I do not feel [the sanctions] are fair or just."

VOLCANOLOGY

Accusations of colonial science fly after eruption

Congolese volcano watchers say European data monopoly hampered Nyiragongo forecasts

By **Roland Pease**

On 22 May, Mount Nyiragongo, perhaps the most dangerous volcano in the world, erupted in a show of fire. Lava swept toward the city of Goma in the Democratic Republic of the Congo (DRC), pushing thousands from their homes and killing dozens. Although the volcano has since settled down, a new flashpoint has erupted at the geophysical observatory that monitors it.

In a 2 June open letter addressed to the DRC's president, staff at the Goma Volcano Observatory (GVO) have condemned what they say is corruption by the observatory's Congolese leadership. They also accuse European partners of a "neocolonial" attitude and of depriving them of timely data that might have allowed them to provide early warnings of eruptions.

Signed by union leader Zirirane Bijandwa Innocent on behalf of dozens of staff researchers and technicians, the letter alleges that GVO leaders squandered money from international donors, failed to pay staff for months, and even had some researchers arrested for complaining about the situation. It also charges that the Royal Museum for Central Africa (MRAC) in Belgium and the European Centre for Geodynamics and Seismology (ECGS) in Luxembourg, long-term partners with GVO, wield too much influence over its leadership. The letter says the observatory "was taken hostage … by a small group of scientific neo-colonialists" who shut out local experts and focused on their own volcanology research at the expense of developing local capacity to monitor geohazards.

Deadly flows
On 22 May, vents on Mount Nyiragongo's flank spilled lava that destroyed thousands of homes.

In a statement to the DRC Parliament on 9 June, the science minister denied the charges of misappropriation and embezzlement, although in January GVO's previous director-general was replaced after similar allegations. GVO's scientific director declined to discuss the allegations, saying he needed "time to discuss with colleagues." And in a statement to *Science* written on behalf of the European partners, ECGS Scientific Director Adrien Oth said they were "very surprised" by the charges of colonial science, "which we consider to be very unfair and unfounded."

Difficulties at GVO came to a head in October 2020 when the World Bank decided not to renew financial support that had been in place since 2015 (*Science*, 16 October 2020, p. 270). Without confirming allegations of corruption, the bank cited "weaknesses in implementing such a grant" in justifying its decision.

The cuts left the observatory unable to afford even an internet connection. That deprived GVO of real-time data from a network of seismometers and GPS stations deployed across the region by MRAC and ECGS since 2012. These devices can detect the small tremors and movements of Earth's surface

CREDITS: (MAP) EUROPEAN COMMISSION DIRECTORATE-GENERAL FOR EUROPEAN CIVIL PROTECTION AND HUMANITARIAN AID OPERATIONS. ADAPTED BY N. DESAI/*SCIENCE*; (PHOTO) JUSTIN KATUMWA/AFP/GETTY IMAGES

Lava destroyed homes on the outskirts of Goma, Democratic Republic of the Congo.

that can precede eruptions, as magma rises inside a volcano. The sensors send their data directly to ECGS, where they are automatically processed, before being returned to GVO.

"We spent about 6 months without receiving the data," says a GVO staff member, speaking on condition of anonymity for fear of reprisals. In February, ECGS arranged a web repository for the data, although without an internet connection GVO could not access it. Instead, the GVO source says, the team in Goma would buy one-off mobile data packages to keep a stream of day-old readings coming in. In the end, GVO's internet access was restored in April by the Volcano Disaster Assistance Program (VDAP), a U.S. relief program.

The complaint exposes the fragility of the relationship between observatory staff and their European partners—and the frustrations that result from the data bypassing GVO. A U.S. government volcanologist who works with VDAP but did not have permission to speak publicly says VDAP's policy, in contrast MRAC and ECGS, "is to donate equipment to the country, ensuring that the resulting data belongs only to that country." GVO staff say they are sometimes treated as "field boys," good for data collection, but no more, says Jonathan Esole, a DRC-born mathematician at Northeastern University who has visited GVO and advocates on behalf of its staff. The U.S. government volcanologist confirmed the impression, saying the Belgians appear to believe "that the GVO and DRC are their 'turf.'"

In his statement, Oth strongly rejects these assertions, pointing to the collaboration's long-term commitment to GVO. It has paid for three GVO staff members—including the current director-general—to do Ph.D. work in Belgium and for seven staffers to earn master's degrees. Oth says the partners have also conducted several training sessions in Goma for staff members. As far as the data are concerned, Oth says it was fortunate that they passed through Europe first, because the arrangement preserved continuity despite the internet interruption at GVO.

The GVO letter also implies that Europeans downplayed warning signs in the weeks before the eruption. In presentations at GVO on 26 April and 10 May after they regained access to the data, staff seismologists highlighted tremor activity that might indicate magma rising through cracks, according to the letter and to *Science*'s source at the observatory. They urged GVO leadership to send teams out to make field observations, but nothing happened. The complainants allege that GVO leaders deferred to advice from their European partners.

In the statement, Oth denies that he or his colleagues knew anything of these discussions. ECGS's automatic analysis did not flag anything concerning at the time, he says. An international group of seismologists convened after the eruption "confirmed the absence of obvious precursors," Oth says. In any case, he adds, "Our contribution cannot and must not be considered as a substitution of the daily work of GVO's seismology department."

At the request of the DRC science minister, European scientists have arranged to go to Goma to meet with GVO leaders. In the statement, Oth says the minister confirmed his "full confidence" in the partners and "reiterated the importance of this partnership." As for the staff concerns, Oth says, "As is generally known, the GVO suffers from structural problems, but these are political problems of lack of resources and of adequate management to be resolved by Congolese authorities."

Esole acknowledges the institution's dysfunction and its need for better leadership. But he also reserves criticism for MRAC and ECGS and what he sees as their unequal relationship with GVO. They often refer to their work as "aid," he says. "But who is helping who? Because there are no volcanoes in Belgium or Luxembourg. So to do their study, they need to go to another country. I believe GVO is helping them big time, and it is not a fair trade." ■

Roland Pease is a journalist in London.

MARINE ECOLOGY

Report traces surge in ocean plastic studies

But many studies describe the problem rather than pointing to solutions

By **Tania Rabesandratana**

Plastic winds up everywhere—from the top of Mount Everest to remote Antarctica. Every year, millions of tons of discarded plastic also wash into the ocean.

Research about ocean plastic is swelling, too, from 46 papers in 2011 to 853 in 2019, according to a UNESCO science report published last week. The report found that ocean plastic research grew much faster than any of the other 55 development-related topics it tracked. "It has really skyrocketed in recent years," says Utrecht University's Erik Van Sebille, who uses plastic particles to trace ocean currents.

But gaps remain in the research. Journals get many papers about "the presence of plastic on beaches, on the seabed, or in animals, but not [many] about sources or solutions," says Ángel Borja, a marine ecologist at the AZTI research center in Spain.

The ecological effects of plastic pollution are another hot research topic. Plastic itself is inert, but often contains toxic additives such as flame retardants, pigments, or chemicals to make plastic more flexible and durable. "These additives are what we're worried about," says Carmen Morales, an ecotoxicologist at the University of Cádiz's Marine Litter Lab, who in a study published last week found that takeout food and drink packaging is the most pervasive source of ocean plastic.

Researchers are also concerned about animals that eat plankton-size particles without deriving any nutrition. Nanoplastic particles, small enough to penetrate tissues, may be the most harmful of all. Yet the overall ecotoxicological effects of plastic are poorly understood. Regardless, plastic pollution is a pressing problem, says Bart Koelmans, an aquatic ecologist at Wageningen University. "It's an eyesore to have all this plastic on beaches," he says. "For many people, that is enough to be concerned." ■

RESEARCH INTEGRITY

Biologist blows whistle on prominent co-author

When inquiries went nowhere, Ken Thompson decided to publicly disavow his first paper

By **Martin Enserink**

A scientist's first academic paper is usually a career milestone as well as a source of pride. For evolutionary biologist Ken Thompson of the University of British Columbia, Vancouver, it's neither. Instead, it has become a case study in the frustrations facing a would-be whistleblower.

On 10 May, Thompson published a "technical comment" via Dropbox that reads like a frontal attack on his own first paper, published in 2014 in *Biodiversity and Conservation*. Thompson was the first author of the study, which pitted traditional techniques for identifying plant species against DNA barcoding, which uses short genetic sequences to differentiate species. But his post identified what he said were serious issues in data from his co-author and concluded: "Until and unless these matters are resolved, I feel that I can no longer stand by the results of the study."

Thompson wrote the 2014 paper as an undergraduate student at the University of Guelph (UG), but the data came from botanist Steven Newmaster, a prominent lab leader there. Among the issues Thompson raised: Records showed the DNA data were not posted online in 2014, as the paper said, and when they were posted—in 2020, after Thompson began to raise concerns—they did not support all of the paper's conclusions. They also showed a striking and, in Thompson's view, implausible resemblance to a different molecular data set collected for a separate study by a researcher at the university's Centre for Biodiversity Genomics (CBG).

"I am not accusing anybody of anything untoward," Thompson wrote in a guest post on the blog Eco-Evo Evo-Eco, the same day he published his disavowal of the paper. But the post did appear to incriminate Newmaster, the paper's only other author, who did not respond to multiple interview requests from *Science*.

Email correspondence that Thompson shared with *Science* shows he tried for more than a year to persuade UG to investigate the paper, with little success. A university spokesperson emailed *Science* that "allegations of research misconduct are taken very seriously" at UG, but wrote that "details and outcomes of specific allegations are confidential." Last month, after *Biodiversity and Conservation* editors also declined to investigate the paper, Thompson decided to go public. "I don't want to deal with this alone anymore," he wrote in his post, describing the process as "incredibly isolating."

Thompson soon received support from Paul Hebert, founder and director of CBG and a DNA barcoding pioneer. On Eco-Evo Evo-Eco, Hebert took his own university to task for not properly investigating the paper. "For me to watch a young academic who brought to light serious concerns ... I could not stand back," Hebert tells *Science*. "If I have concerns I have to speak up."

Mounting a challenge to his first paper was "incredibly isolating," Ken Thompson wrote. "I don't want to deal with this alone anymore."

Thompson's 2014 paper set out to determine the best way to identify and survey plant species: the old-fashioned way, which relies on morphology; or by extracting DNA from leaves and using barcoding. Data from 337 plots in northeastern Ontario showed barcoding was clearly better. The method identified more species per plot and more species overall, at a lower cost per species.

Interviewed by email, Thompson wrote that he long feared something was wrong with the paper, but was afraid that airing his doubts could harm his career. On the other hand, if others raised questions about the paper, "I would have to either lie and say I had no idea or tell the truth that I did not say something even though I had substantiated suspicions," he wrote. A scandal around social spider researcher Jonathan Pruitt in early 2020 (*Science*, 7 February 2020, p. 613) inspired him to take action.

The emails between Thompson and UG officials indicate the university did look into the matter. But on 10 September 2020, UG Associate Vice-President Karina McInnis wrote Thompson that an "initial inquiry" indicated a full investigation was not required and the case was closed. McInnis rebuffed several pleas from Thompson to revisit the case.

Thompson suspects UG may have been reluctant to investigate because Newmaster is an important asset. His research has generated more than $7 million in funding, according to UG's website; Newmaster also founded the NHP Alliance, which aims to improve authentication processes for natural health products such as herbal supplements and receives major industry funding. (The UG spokesperson says misconduct procedures "are the same regardless of the identity of the respondent and complainant.")

After Thompson began to press for an investigation, Newmaster's lab uploaded thousands of genetic records supposedly associated with the paper to GenBank. They only deepened Thompson's suspicions. The data did not support the identification of all of the species supposedly identified in the study, he says, because DNA barcoding lacks the specificity to distinguish some of them. For example, the paper claimed to have distinguished seven willow species, but the data allowed only the identification of the genus, *Salix*, and not individual willow species, Thompson wrote. Even more troubling, most of the genetic data posted in 2020 are almost identical to data for two different papers published by CBG researchers in 2012 and 2017, he wrote.

Hebert agrees barcoding alone can't identify every plant species, as the study claimed. "That's never been seen in any plant study," he says. "It's difficult to resolve closely allied plant species; this is a well-known problem in certain genera." He also confirms that "more than 6000 sequences" deposited into GenBank by a member of

PHOTO: MACKENZIE URQUHART-CRONISH

Newmaster's lab in 2020 "are duplicates or lightly edited duplicates of sequences" that were previously posted.

Hebert says the 2014 paper also appears to include a false claim: that the barcoding of its plant samples was done at the Canadian Centre for DNA Barcoding, where he is scientific director. The samples aren't in a database "that tracks all samples that come to the lab for analysis," he says. UG's investigation was "incomplete," Hebert concludes. "I admire Ken's determination to provoke evaluation of his concerns."

Thompson failed to persuade *Biodiversity and Conservation* to investigate the paper. "The journal is not in a position to question [UG's] investigation," a representative of the publishing company, Springer Nature, wrote to Thompson on 7 May, 3 days before he went public. "Institutions are usually in a better position to investigate such concerns," the journal's editor-in-chief, U.K. mycologist David Hawksworth, wrote in an email to *Science*. However, Hawksworth added, "We are now looking into the matter further and will take editorial action, if we conclude it is necessary."

It's not uncommon for research journals to defer to institutional misconduct investigations, in line with recommendations from the Committee on Publication Ethics (COPE). But that practice can be immensely frustrating to would-be whistleblowers, especially when universities dismiss allegations against their own scientists.

"I feel very sorry for [Thompson], who seems to have been very honorable and done the right thing," says Liz Wager, a former COPE chair and co-author of a new report on how institutions and journals can better collaborate to resolve misconduct allegations. "But equally, the journal is stuck." Unlike universities, journals can't compel authors to cooperate with an investigation, she says. Journals often lack the required expertise, and editors may worry about the legal ramifications of retracting a paper against an author's wishes. "They don't want to get involved in really expensive legal cases, which could sink a small journal."

Thompson says he is hugely relieved to have Hebert on his side, adding, "I almost fainted" when he expressed his support. But he's disappointed by the response from both the university and the journal. "The institutions meant to uphold the process of science have utterly failed," he says.

Thompson is now preparing a new, detailed document, he says, with scientists at UG and elsewhere, in a final attempt to get the university to launch a formal investigation. "I am optimistic," he says, that "they will take this seriously." ■

Ancient DNA from a Siberian cave reveals the first known father-daughter pair of Neanderthals.

IMAGE: TOM BJÖRKLUND

PALEOANTHROPOLOGY

Genomes offer rare glimpse of Neanderthal family groups

Low genetic diversity in males from two Siberian caves suggests females moved to their mates' families

By **Ann Gibbons**

More than 49,000 years ago, a family of Neanderthals set up camp in a cave high in Siberia's Altai Mountains, overlooking a river valley where bison, red deer, and wild horses roamed. In the cave's main gallery, a teenage girl lost a tooth, perhaps while gnawing on bison that her father or his kin had hunted in the sweeping grasslands.

Now, researchers have analyzed the genomes of this father and daughter and 12 of their relatives, many of whom sheltered in the same cave over less than 100 years. The new genomes almost double the number of Neanderthal genomes known and offer a glimpse of the Neanderthal population at the eastern end of their range, at a time when they were headed toward extinction.

The genomes also offer the first real clues to the social structure of a group of Neanderthals. In addition to identifying the first father-daughter pair, the genetic evidence suggests these males stayed in their family groups as adults, like men in many modern human societies, says geneticist Laurits Skov of the Max Planck Institute for Evolutionary Anthropology. He presented the work in a virtual talk at the ninth International Symposium on Biomolecular Archaeology earlier this month.

"It's really remarkable that they managed to get genomes from seven males at one site," says paleogeneticist Cosimo Posth at Tübingen University. "For this group in this cave, it is indeed suggestive that they lived in small groups of closely related males."

Over the past decade, geneticists have sequenced the genomes of 19 Neanderthals. But that DNA mostly came from females who were distantly related and lived at sites across Europe and Asia anywhere between 400,000 and 50,000 years ago.

Computational biologist Benjamin Peter and paleogeneticist Svante Pääbo at Max Planck led the new study with a team including Skov, a postdoc. They extracted Neanderthal DNA from teeth, bone fragments, and a jawbone dug up during ongoing excavations at Chagyrskaya and Okladnikov caves by archaeologists at the Russian Academy of Sciences in Novosibirsk. Optically stimulated luminescence dates of the sediments around the teeth and bones suggest the Neanderthals lived between 49,000 and 59,000 years ago. Both caves are close—50 to 130 kilometers—to the famous Denisova Cave, which was inhabited by both Neanderthals and their close cousins, the Denisovans, off and on between 270,000 to 50,000 years ago.

The researchers analyzed DNA from more than 700,000 sites across the genomes from seven males and five females from Chagyrskaya, and from a male and a female from Okladnikov. They found family ties: The nuclear DNA from one Chagyrskaya bone fragment linked the father to the tooth shed by his teenage daughter. Some individuals shared two types of maternally inherited mitochondrial DNA (mtDNA). Those genomes hadn't yet differentiated from each other, which happens in a few generations, so the individuals must have lived during the same century.

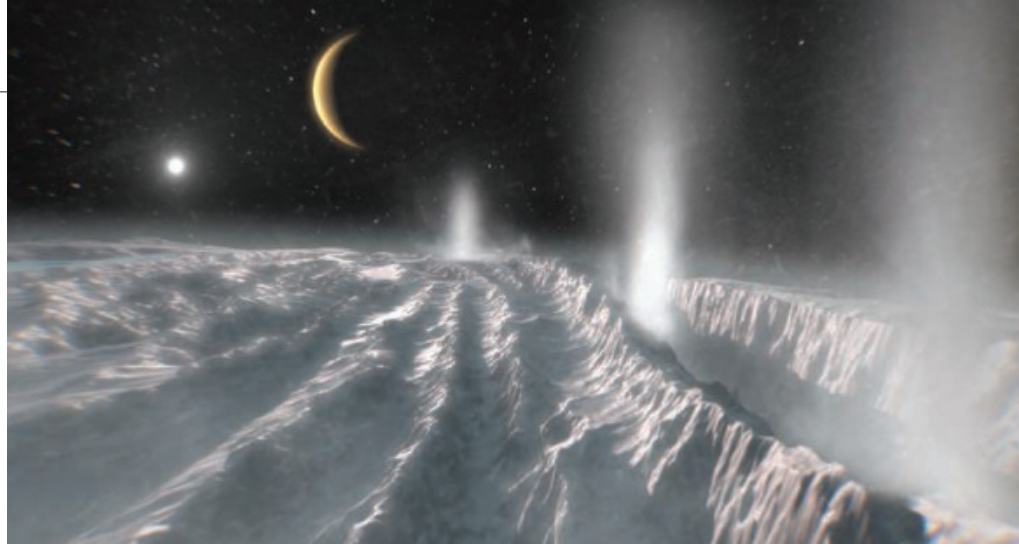
The DNA painted a bigger picture of Neanderthal society. Several Chagyrskaya males carried long chunks of identical nuclear DNA from the same recent ancestor. Their Y chromosomes were also similar and came from a modern human ancestor, like those of the only three other male Neanderthal genomes known. The nuclear DNA also showed they were more closely related to later Neanderthals in Spain than to earlier ones at neighboring Denisova, suggesting migration.

The similarities among the males suggest they belonged to a population of only hundreds of men who were fathering children—about the same number of breeding males as seen in endangered mountain gorillas today. "If you were to think of this Neanderthal population like [populations today], they would be an endangered population," Skov says.

In contrast to the Y chromosome and nuclear DNA, the mtDNA of both males and females was relatively diverse, implying that more female ancestors contributed to the population than males. That could be a founder effect, in which the initial group included fewer fertile males than females. Or it could reflect the nature of Neanderthal society, says paleogeneticist Qiaomei Fu of the Chinese Academy of Sciences, who heard the talk. Either "fewer men than women contributed to the next generation, or women moved more frequently between groups," she says.

To Skov, the evidence suggests the latter. He says modeling studies show it's unlikely that a small group of migrants expanding from Europe into Siberia would include mostly females and few males. Instead, he thinks these Neanderthals lived in very small groups of 30 to 110 breeding adults, and that young females left their birth families to live with their mates' families. Most modern human cultures are also patrilocal, underscoring another way that Neanderthals and modern humans were similar.

Posth cautions that 14 genomes can't reveal the social lives of all Neanderthals. But he sees ominous signs in the males' low diversity. The end was fast approaching for our closest cousins: In just 5000 to 10,000 years, they would be gone. ■



Saturn's water-spewing moon Enceladus (artist's conception) is a possible target for a future European lander.

IMAGE: EUROPEAN SPACE AGENCY/SCIENCE OFFICE

SPACE SCIENCE

Decades ahead, Europe picks goals for big space missions

Themes for billion-euro Voyage 2050 missions include icy moons, exoplanets, and the early universe

By **Daniel Clery**

Big space missions take patience and long planning. And all start with a goal. The European Space Agency (ESA) took that first step last week when it revealed the science themes it wants to pursue in billion-euro missions to be launched between 2035 and 2050. They include taking a close look at ocean-bearing moons around Jupiter and Saturn, dissecting the atmospheres of temperate exoplanets, and using new tools to study the formation of the universe's first stars, galaxies, and black holes.

The choices reflect areas most likely to generate "breakthroughs in science on that timescale," says Linda Tacconi of the Max Planck Institute for Extraterrestrial Physics, who chaired the selection panel.

ESA refreshes its slate of science missions every decade or so. The current program, called Cosmic Vision, has three flagship missions that will launch by 2034: a spacecraft to study Jupiter's icy moons, an x-ray telescope, and a gravitational wave detector.

Competition for the next round, dubbed Voyage 2050, kicked off in 2019 with almost 100 suggested missions or themes from teams of researchers (*Science*, 25 October 2019, p. 410). Dozens of researchers were recruited into committees to winnow down the suggestions into three categories, which ESA's Science Programme Committee approved last week. Although the themes do not specify particular missions, some clearly nod at concepts that researchers have been working on for years.

The first calls for looking for life in the Solar System with a visit to one of the moons of Jupiter or Saturn thought to have water oceans hidden beneath their frozen shells, including Europa, Enceladus, Titan, and Ganymede. Such oceans, warm and laden with minerals essential to life, "hold important clues to the emergence of life on Earth," says Athena Coustenis of the Paris Observatory in Meudon. NASA and ESA both have upcoming missions to swoop past some of these moons, but ESA wants to get closer with a lander or drone. Technology development needs to start now, Coustenis says. "Today, we don't know how to land and drill down into the oceans, how to shield landers from radiation, how to communicate," she says.

A second theme calls for new probes of the early universe. One possibility is a mission that would build on Planck, which mapped the relic microwave radiation from the big bang to learn about the expansion of the universe and how initial clumps of matter seeded the formation of galaxies. Advances in microwave spectrographs could allow the mission to test the standard model of cosmology and look for evidence of primordial black holes or proposed dark matter particles like axions.

Another possibility is a successor to the Laser Interferometer Space Antenna (LISA),

a fleet of three spacecraft flying in formation, due for launch in 2034. LISA aims to detect gravitational waves with long wavelengths, such as those from the merger of massive black holes—events that detectors on Earth cannot catch. A new mission could probe even longer wavelengths, revealing how quasars—superbright galactic cores in the distant universe—form and whether the supermassive black holes at the centers of galaxies grow by swallowing stars and gas or by merging with other giants.

The third theme calls for studying either the Milky Way's stars or planets. If ESA chooses exoplanets, one option would be scrutinizing the midinfrared glow from temperate exoplanets for atmospheric gases that could hold signatures of life. One mission concept in contention is the Large Interferometer for Exoplanets (LIFE), a flotilla of four orbiting telescopes that would combine their light in a fifth craft not only to sharpen their view, but also to cancel out a star's glare, revealing the faint infrared glow of surrounding planets. "This is what we really need to understand exoplanets," says Sascha Quanz of ETH Zurich, who heads the LIFE team. Tacconi says the exoplanet theme was "very, very compelling," but acknowledges that many questions hang over such a costly mission, including whether there are enough exoplanets within range to get a good sample.

If ESA opts for Milky Way stars instead, one option could be a follow-on to ESA's Gaia telescope, which is mapping the positions and motions of 2 billion stars to understand how the Milky Way evolved. The new mission would shift to near-infrared light, mapping five times as many stars with 10 times the accuracy. "It tells you how the galaxy is moving," says David Hobbs of Lund University, who wants to use the stars as tracers for the halo of invisible dark matter that helps hold the galaxy together.

ESA plans to build the ocean moon mission first, in part because after it launches the Jupiter Icy Moons Explorer next year, teams with relevant expertise will be free to work on a new mission. Fabio Favata, ESA's head of strategy, planning, and coordination, says around the end of the year ESA will form a committee of experts to pick the target moon and decide what is technologically possible within the time and cost constraints. Feasibility studies from contractors will follow, and after as many as 5 years, teams will be invited to propose missions for the slot. A year or two later, the process will start again for the next theme. Building such missions means "finding a compromise between ambition and boundary conditions," Favata says. "It takes years of work to finally converge." ■

U.S. SCIENCE POLICY

DNA researchers question Senate bill's security provisions

Measure aimed at stopping China from misusing human genome data could harm research efforts, groups argue

By **Jocelyn Kaiser**

A provision buried in a 2400-page bill approved last week by the U.S. Senate to help the United States compete with China is drawing fire from human genome researchers. It would require the National Institutes of Health (NIH) to develop new security protocols aimed at preventing the misuse of U.S.-funded genomic data by China and other nations.

The provision is not based on substantiated security risks, and "could slow biomedical advances and impose unintended burdens," the American Society of Human Genetics (ASHG) warned last week in a letter to lawmakers. The Association of American Medical Colleges cautioned in a statement that "any additional protections or restrictions … should be commensurate with the actual risk."

> **"Any additional protections or restrictions … should be commensurate with the actual risk."**
> Association of American Medical Colleges

Research advocates are applauding many provisions of the huge Senate bill, the United States Innovation and Competition Act (S. 1260), which calls for increasing federal research spending and creating a technology directorate at the National Science Foundation (*Science*, 21 May, p. 777). But they're less enthusiastic about a provision reflecting concerns that China is amassing DNA data on U.S. citizens. Some fear China could use such data to give its drug industry an economic advantage, devise bioweapons tailored to Americans' genetics, or even blackmail people by threatening to release private information. To reduce those risks, Senator Marco Rubio (R–FL) and other lawmakers have been pushing NIH to take additional steps to strengthen security protections for genome data, and more intensively scrutinize foreign scientists who request access to that those data.

The bill would require federal officials to "ensure" that research backed by NIH and other agencies that involves "the sequencing of human genomic information, and collection, analysis, or storage of identifiable, sensitive information … is conducted in a manner that appropriately considers national security risks." NIH must work with intelligence agencies to issue, within 1 year, "a comprehensive framework" for managing risks, such as requiring more training for NIH-funded investigators and peer reviewers and including security experts on data access panels.

In the past, NIH has argued that existing security measures are adequate. Researchers already strip identifying information from genome data, and NIH reviews, and sometimes rejects, scientists' requests for access. But in 2019, the Office of Inspector General (OIG) of the Department of Health and Human Services, NIH's parent agency, suggested NIH do more, for example by adding controls on foreign scientists who use U.S. genome data.

In a response to OIG, NIH questioned the severity of the threat. It noted security worries were largely based on "a single Congressional testimony," from FBI agent Edward You, who has long warned of the risks of sharing genomic research data. Fears of economic harm were "theoretical," NIH said, noting that many experts argue that sharing data promotes innovation. And it scoffed at the "improbability" of weaponizing human genetics data. Research would "come to a halt," NIH said, if it had to write craft policies "to counter every theoretical risk." It said more training, for example, would not stop someone with "nefarious intent."

ASHG makes similar points in its 10 June letter opposing the Senate provision to Representatives Frank Pallone (D–NJ) and Cathy McMorris Rodgers (R–WA). They are senior members of the House of Representatives Committee on Energy and Commerce, which will likely have a say in any final legislation that might emerge later this year from negotiations between the House and Senate. ASHG wants Congress to remove the Senate provision from any final bill, and instead first have a panel of scientific and security experts determine "what national threats exist, if any," then work together on new policies. ■

MARSHES ON THE MOVE

Coastal ecologist Matt Kirwan is optimistic that coastal wetlands can outrun rising seas. But some question that rosy picture

By **Gabriel Popkin**, *at the Blackwater National Wildlife Refuge, Maryland; Photography by* **Michael Snyder**

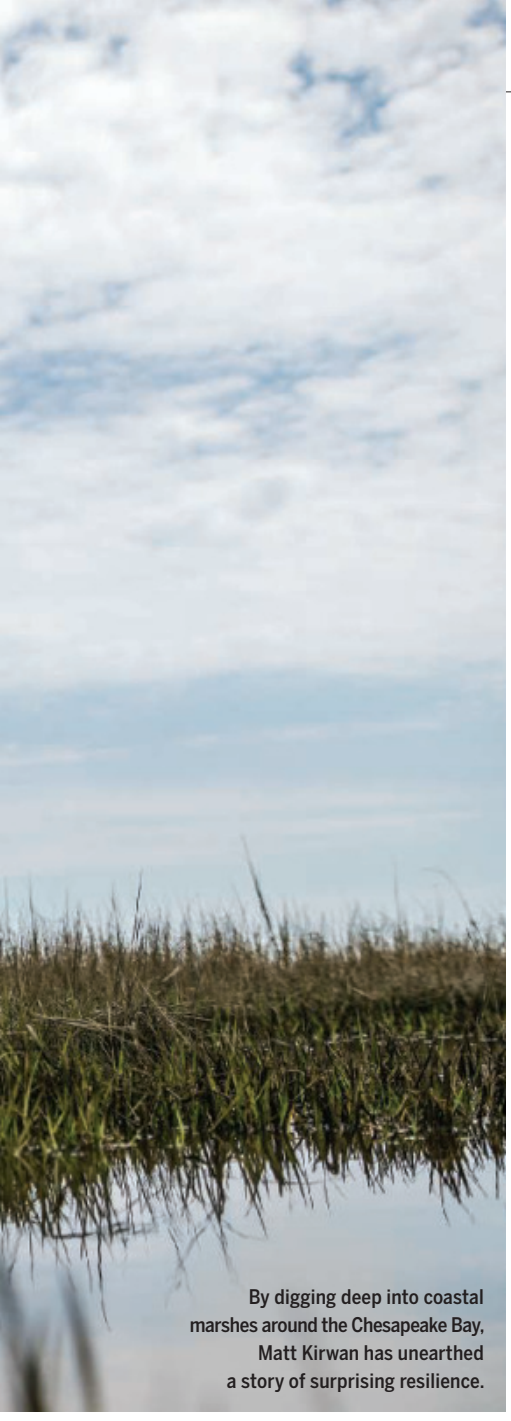
Coastal scientist Matt Kirwan has a sense of what it's like to flee from rising seas. More than 100 years ago, Kirwan's great-great-grandfather owned a farm close to this sprawling wetland refuge near the Chesapeake Bay, a key annual migration stop for hundreds of thousands of geese, ducks, and other waterbirds. Joseph Josiah Robbins sold his homestead in 1909, family members say, after salty, brackish waters invaded his fields, stunting and killing his crops.

Kirwan's ancestors were early climate refugees, he says, even if nobody called them that. "There are large tracts of farmland that were usable a generation ago but no longer are."

Seas are rising especially fast near the Chesapeake, where geologists say Earth's crust is tilting into the ocean, amplifying the effects of a warming climate. But many more people around the world could soon share his ancestors' experience, says Kirwan, who works at the Virginia Institute of Marine Sciences (VIMS). The rate of global sea level rise is increasing as ice from Greenland and Antarctica melts and warmer seawater expands. By 2050, sea levels could surge by 10 to 25 millimeters per year, according to climate modelers, up from just 5 to 6 millimeters per year here at the refuge and about 3 millimeters per year globally.

Saltwater is forecast to slosh into places it hasn't been since long before humans arrived. In the continental United States alone, just 1 meter of relative sea level rise could allow high tides to submerge as much as 49,000 square kilometers of now dry land—an area the size of Vermont and New Hampshire combined. Other countries have it worse. Egypt and Bangladesh could lose almost one-fifth of their habitable land. More than 200 million people could face flooding risk by 2050, studies suggest.

Humans aren't the only species at risk. Rising seas could also drown huge swaths



By digging deep into coastal marshes around the Chesapeake Bay, Matt Kirwan has unearthed a story of surprising resilience.

PHOTO: MICHAEL SNYDER

of coastal wetlands, destroying habitats that are among the most ecologically valuable on Earth. Besides providing homes for plants, fish, and many other organisms, wetlands trap vast amounts of carbon that might otherwise escape into the atmosphere and contribute to planetary warming. They also buffer some of the world's largest cities from violent, damaging storm surges. Some 20% to 90% of today's tidal wetlands could be lost by century's end, depending on how fast oceans rise, studies have suggested.

Kirwan, however, has argued that such forecasts are needlessly bleak. Studies that he and colleagues have conducted—many around the Chesapeake Bay—have shown that in the right conditions, tidal marshes can build themselves up to keep pace with rising seas, while also migrating inland as water creeps up coastlines. Some coastal wetlands could even expand as seas rise, the studies suggest—if people don't block their paths with seawalls, levees, and other infrastructure designed to hold back the flood.

"Marsh vulnerability tends to be overstated," Kirwan and colleagues wrote in a 2016 paper in *Nature Climate Change*, concluding that sea level rise does not pose "an immediate, catastrophic threat to many marshes."

Kirwan's findings are "very compelling," says ecologist Keryn Gedan of George Washington University, and have "given the marsh community a ray of hope."

But many researchers are skeptical, and some are challenging Kirwan's results. Models suggesting wetlands can keep pace with high rates of sea level rise are "seriously flawed," says ecologist Neil Saintilan of Macquarie University. "This idea of wetlands expanding under accelerated sea level rise, it basically violates some of the most basic geologic theory," says Torbjörn Törnqvist, a geologist at Tulane University. "It's just not going to happen."

The debate isn't just academic. Forecasts of how coastal wetlands will respond to rising water will help determine how aggressively the United States and other nations work to save these critical ecosystems, and which policies they adopt. "We should set a goal as ambitious as possible, given the facts," says Jeffrey Peterson, a retired policy adviser at the Environmental Protection Agency. "But we don't know what the facts are."

EARLY ONE MORNING this spring, Kirwan and two other VIMS researchers—then-undergraduate Emily Hall and lab manager Tyler Messerschmidt—mustered on a spit of land along the Severn River in southern Virginia. It's a soggy landscape where roads have names such as Ditch Bank and Low Ground. Hall and Messerschmidt loaded imposing data-collection gear onto their backs and trekked into what may be one of the world's youngest marshes.

Less than 100 years ago, farmers grew crops here. Those farmers, like most people through history, saw wetlands not as a boon, but as an impediment. In the United States, more than half of the nation's original wetlands were drained for agriculture or development. At this site, earthen levees 1 meter or so high—built by enslaved people, local lore has it—were erected to keep fields dry. "This is how people fought against flooding in the past," Kirwan said as he stood atop one of the still-intact levees.

In a low-lying pool, Hall and Messerschmidt erected a tripod topped by a sensor that collected hyperaccurate measurements of the marsh's topography. As the unit's beeps and boops mixed with bird song, the pair walked a transect through the marsh, battling sharp, stiff grasses. Every 5 meters, Hall jammed a metal shank called a peat corer into the muck and gave it a firm twist to cut out a plug of soil. The top few centimeters were full of marsh grass roots that hadn't decomposed. Below that, however, was mineral soil, revealing that the site was dry land until recently.

A core from Virginia's Goodwin Island reveals when dry land became marsh. Carbon absorbed by marsh plants can be stored in soil for centuries.

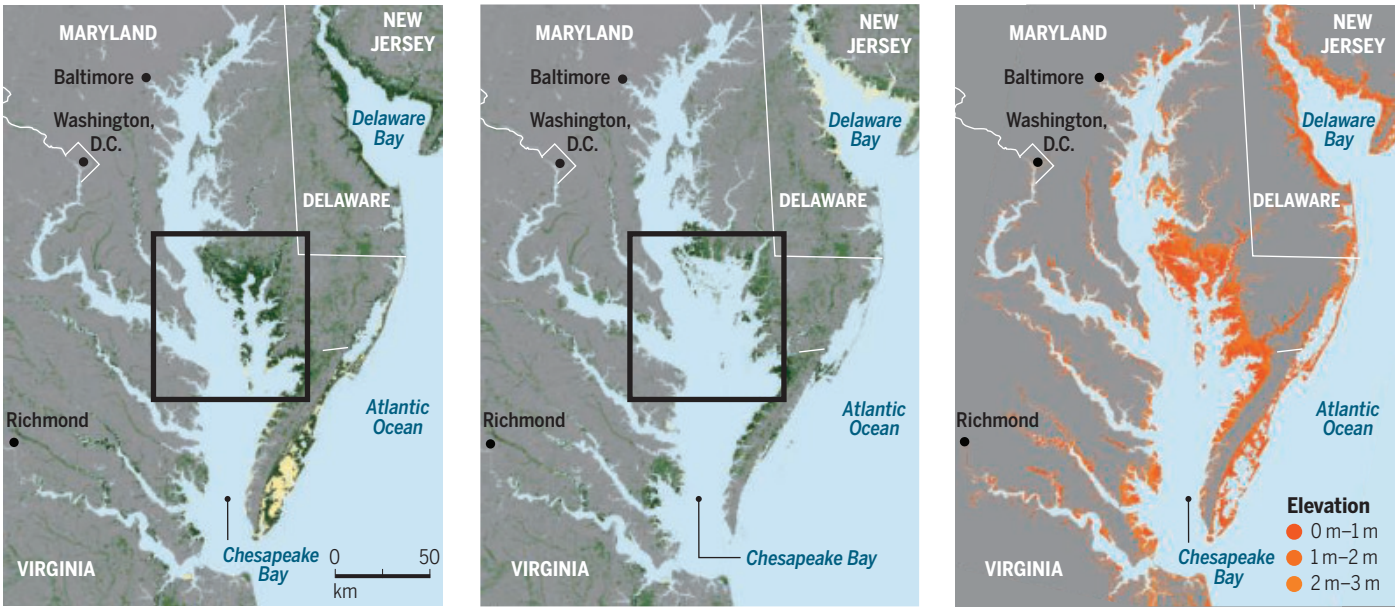
Kirwan's disarmingly informal manner and often-battered field clothes belie his status as one of the world's most respected coastal experts. He got his start as an undergraduate at the College of William & Mary, just down the road in Williamsburg, Virginia. After a doctorate at Duke University and a stint at the U.S. Geological Survey, he returned to William & Mary and VIMS in 2011 as a professor, and the Chesapeake Bay's wetlands became his laboratory. Thanks to childhood visits to family in the area, Kirwan is as comfortable bantering with hunters who flock to places like the Blackwater refuge as with fellow scientists.

Through his research, Kirwan has developed a view of coastal wetlands as dynamic, resilient ecosystems that have danced an eonslong tango with the rising and falling ocean. And some of his findings have challenged conventional thinking about how future sea level rise could change the planet's coastlines.

Wetland or water?

The fast-sinking Chesapeake Bay region offers a window into future scenarios for tidal wetlands. Some computer models suggest vast swaths of wetlands could disappear in coming decades as oceans surge. But other models forecast that coastal wetlands could maintain or even expand their areas by building soils vertically and migrating to higher ground that is currently dry.

● Marsh ● Swamp ● Beach



A marshy coast
The bay region, including the area around the Blackwater National Wildlife Refuge (upper center in box), currently supports extensive tidal marshes.

If wetlands can't keep up
Sea level rise of 1 meter could drown many wetlands according to widely used models that assume they have limited ability to build soils or migrate horizontally.

Potential migration zones
If tidal wetlands are able to build up and migrate inland as seas rise, they could take over coastal areas that are currently well above sea level.

Kirwan's first big splash was a 2010 study in *Geophysical Research Letters* showing that just two factors—the tidal range (the elevation difference between low and high tide) and the amount of sediment in the water—determined how much sea level rise salt marshes around the world could withstand. In places with large tidal ranges, plants near the high tide line will remain partly above water even as seas rise, whereas in places with less tidal variation, such as the Blackwater reserve, even a little bit of additional water can fully drown marshes. Meanwhile, in places where plentiful sediment settles out of water, marshes can build new ground at rates up to an astounding 10 centimeters per year, modeling by Kirwan and colleagues suggested, enabling them to outrun even high rates of sea level rise.

A few years later, in an influential 2013 paper published in *Nature*, Kirwan and Patrick Megonigal of the Smithsonian Environmental Research Center expanded on those findings to argue that most wetlands globally were keeping up with sea level rise just fine. One key, they wrote, is a biophysical feedback loop: Wetland plants slow incoming ocean water, causing fine sediment it carries to drop out and accumulate. As sea level rises and wetlands flood more often, the plants trap more sediment, building up the soil surface. Meanwhile, marsh plant roots die, but mostly don't decompose, further raising the ground.

Kirwan also found that wetlands' ability to migrate landward had been dramatically underappreciated. Many scientists, he and Gedan say, assumed most migrating wetlands would encounter either a natural barrier such as a steep slope or a bluff, or a human barrier such as a seawall. But Kirwan found many areas lacked such impediments and could accommodate enormous marsh migrations. In 2017, for example, his research team reported that since 1850, the flat, fast-sinking Chesapeake region has gained as much marsh as it has lost: more than 40,000 hectares, or more than twice the area of Washington, D.C.

One of Kirwan's research sites, Goodwin Island at the mouth of the York River, provides a powerful illustration. There, researchers documented a marsh that has been creeping uphill for more than 600 years. And Kirwan's team has found that even the newest marshes provide important ecosystem services, such as absorbing and storing carbon.

On Goodwin Island and elsewhere, those new marshes often occupy land that was forest until saltwater poisoned the trees. Such ghost forests have been expanding along the U.S. east coast for 100 years, Kirwan and Gedan pointed out in a 2019 paper in *Nature Climate Change*. Although these dead forests can look apocalyptic, the marshes that replace them actually provide more benefits for nature and people, and should be welcomed, Kirwan says. "From a pure ecological perspective, ghost forests are a win."

WHAT HAS STIRRED the most debate, however, are Kirwan's forays beyond quantifying past change to forecasting the future, when sea level rise is likely to accelerate rapidly. In 2016, Kirwan and colleagues critiqued forecasts of massive coastal wetland loss in a paper published in *Nature Climate Change*. The study, titled "Overestimation of marsh vulnerability to sea level rise," argued that some widely used models downplayed marshes' ability to build up soil as oceans rise. Models that account for those soil-building processes, the researchers argued, suggest marshes could withstand ocean rise of up to a half-meter per decade—among

GRAPHIC: K. FRANKLIN/*SCIENCE*

the direst scenarios for the coming century.

Three years ago, Kirwan and Mark Schuerch, a geographer at the University of Lincoln, went further. Using models, they predicted that, through 2100, coastal wetlands globally could suffer little or no net loss at all, and under some assumptions could even increase their area by up to 60%.

Such rosy forecasts offered a reason for optimism in a field where nearly every forecast is bleaker than the last. One group of researchers, for example, used data from Kirwan and others to predict that thanks to their ability to build vertically, tidal wetlands in the United States will pull carbon dioxide from the air at faster rates as oceans rise.

But Kirwan's positive outlook didn't sit well with some experts. Randall Parkinson, a coastal geologist at Florida International University, argued that Kirwan's 2016 analysis failed to account for the fact that coastal wetlands sink over time as plant roots partially decompose and soils compact under the weight of newly deposited sediment. Those processes ultimately reduce wetlands' surface elevation, Parkinson says, making them more vulnerable to sea level rise than Kirwan's team had estimated. Many of today's wetlands are already on their way to drowning as sea level rise accelerates, he says.

To bolster his argument, Parkinson points to the most recent period when oceans rose quickly: the end of the last ice age. About 20,000 years ago, as massive glaciers began to melt, seas surged; they ultimately rose some 120 meters in 13,000 years. Parkinson collected sediment profiles that were deposited off the coast of Florida during that surge. They showed that mangrove forests—the tropical analog of salt marshes—survived only along narrow coastal strips and were constantly on the move. In contrast, oceans have inched upward far more gently over the past 7 millennia, allowing mangroves and coastal marshes to expand dramatically.

Other recent studies have added weight to Parkinson's challenge. Last summer, for example, Saintilan and colleagues published an analysis in *Science* of sediment cores from 78 coastal sites around the globe. It found that mangroves virtually disappeared everywhere whenever sea level rise exceeded about 7 millimeters per year—a rate many regions could see again in just a few decades.

A study of tidal marshes around the United Kingdom that Benjamin Horton, director of the Earth Observatory of Singapore, helped conduct found a similar survival threshold, as did a 2020 study of historical wetlands in Louisiana led by Törnqvist. "The future of the wetlands in this region is not looking very good," Törnqvist says.

In April, Törnqvist and others took their most direct aim at Kirwan's work, detailing in *AGU Advances* their case for why many of today's wetlands will soon drown. One big problem, they wrote, is that many coastal wetlands aren't receiving nearly enough sediment to build soil at the rate needed. Although marshes now at the high tide line might be able to hang on for a few decades, they will collapse once they have burned through this "elevation capital," Törnqvist argues.

Kirwan has welcomed the critique. The *AGU Advances* study is "fantastic," he says. "It's probably the best challenge to some of my own work that I've read. ... I sent it around to everybody in our lab group." But he is not convinced that the fate of ancient wetlands foreshadows what will happen in coming decades. "We don't know a lot about what a marsh [represented by] a sediment core [from] 8500 years ago ... actually looked like in nature," he says. Lower temperatures and atmospheric carbon dioxide levels in the past, for example, meant plants probably grew more slowly, potentially making it harder for marshes to keep their heads above water.

Horton, who has co-authored papers with both Kirwan and his critics, says more data from more places are urgently needed. "There's a lack of empirical data to predict the resilience or vulnerability of an increasingly valuable ecosystem."

Still, Kirwan acknowledges that if waters rise too fast relative to land, as is happening in Louisiana, for example, wetlands will no longer be able to keep up by building vertically. In those cases, they will have to migrate horizontally, colonizing new areas now subject to daily tides.

WHETHER SUCH inland migrations will happen, however, is largely in the hands of people. In and around cities, many marshes and mangrove forests are, or soon will be, squeezed between rising seas and human infrastructure, such as seawalls, levees, roads, and houses, that block the path to higher ground.

That reality has led some wetland advocates to call for restrictions on development in current uplands that could become wetlands. In Norfolk, Virginia, a nonprofit group called Wetlands Watch is even urging city officials to create incentives for landowners to abandon and demolish structures once migrating wetlands approach within a certain distance.

On a recent rainy spring day, as the Lafayette River spilled over a battered seawall onto streets in Norfolk's low-lying Larchmont

A historical map shows land that was once farmed near the Severn River in Virginia. Many of those fields are now salt marsh thanks to rising sea levels.

PHOTO: MICHAEL SNYDER

The encroachment of salty water killed trees near the Blackwater National Wildlife Refuge in Maryland, creating a so-called ghost forest. Such dead forests can be an ecological win, says coastal researcher Matt Kirwan, because marsh grasses that take over the site build soil carbon and provide other benefits.

PHOTO: MICHAEL SNYDER

neighborhood, Skip Stiles and Mary-Carson Stiff, two leaders of Wetlands Watch, visited a property they see as a likely future wetland. Already, marsh grasses had encroached into the neat lawn behind the three-story house, whose balcony affords a fine view of the river. Stiff, a lawyer, had started discussions with the property owner. But so far, she says, neither that property owner nor any other has committed to forfeit their house to a moving marsh.

The idea may be an easier sell in rural places where land is cheaper and less extensive infrastructure blocks wetlands' path inland. At the Blackwater refuge, for example, federal officials have already bought dry land likely to be overtaken by marsh. The effort was catalyzed by an analysis showing that although some 2000 hectares of refuge marshes have drowned and been replaced by open water since the preserve opened in the 1930s, new marshes have taken over more than 1300 hectares of dry land in and around the refuge.

Those invasions impressed refuge biologist Matt Whitbeck, who began to identify areas where the Fish and Wildlife Service could purchase uplands to create wetland retreat corridors and maintain critical bird habitat. The realization that marshes were readily occupying higher ground and could continue to provide ecosystem services there, Whitbeck says, "was a huge shift in our thinking."

HUMANS MAY EVEN be able to play a positive role in marsh migration. Gedan, for example, noticed that the first salt-tolerant plants to invade previously dry land are often weedy shrubs with few ecological or human benefits. To build a better marsh, she and colleagues have begun a restoration experiment in former farm fields on a fast-sinking peninsula south of the Blackwater refuge. Earlier this year, marsh grasses and other vegetation the researchers had planted were just starting to green up, giving Gedan hope. Farmland like this, she explained, "for a long time has been considered impermeable to marsh migration," because farmers often used levees and other techniques to keep them from becoming waterlogged.

If such efforts to turn farmland into high-value marsh succeed, experts say paying farmers to give up soggy fields could help wetlands migrate and enhance habitat for species such as black ducks. In Maryland, one state-funded pilot project along these lines is already underway. But some farmers there are still building berms and installing tide gates, sometimes illegally, to keep marshes out. And countries such as Germany and the Netherlands, which heavily armored their coasts long ago, have yet to offer much land back to the sea. "It is a very difficult discussion to have, to give this land up," Schuerch says.

Ultimately, if Kirwan is right, many wetlands may overcome both rising seas and humanmade barriers. Along the Severn River, for example, Kirwan and his colleagues are finding that the old levees, durable though they were, ultimately failed. Inside and outside the walls, marsh grasses have colonized the former fields with nearly equal vigor. Young crabs skitter about. The marsh, though just a few decades old, seems to be doing its job—a good thing for nature, and for the people whose houses dot the marsh's edge.

"There's always been this assumption that people wouldn't let marshes move; they wouldn't give up their land," Kirwan says. "What I've tried to show is, not only would we let it happen, we already have." ■

Gabriel Popkin is a journalist in Mount Rainier, Maryland.

CALL FOR PAPERS

SPJ
SCIENCE PARTNER JOURNALS

Energy Material Advances

Energy Material Advances is an online-only, Open Access journal published in affiliation with **Beijing Institute of Technology (BIT)** and distributed by the **American Association for the Advancement of Science (AAAS)**. The journal publishes, research articles, review articles, short communications, perspectives, and editorials. *Energy Material Advances* covers multiple fields from cutting-edge material to energy science, investigating theoretical, technological as well as engineering aspects.

Submit your research to *Energy Material Advances* today!
Learn more at **spj.sciencemag.org/energymatadv**

The Science Partner Journal (SPJ) program was established by the American Association for the Advancement of Science (AAAS), the nonprofit publisher of the *Science* family of journals. The SPJ program features high-quality, online-only, Open Access publications produced in collaboration with international research institutions, foundations, funders and societies. Through these collaborations, AAAS furthers its mission to communicate science broadly and for the benefit of all people by providing top-tier international research organizations with the technology, visibility, and publishing expertise that AAAS is uniquely positioned to offer as the world's largest general science membership society.
Visit us at: **spj.sciencemag.org**

@SPJournals **@SPJournals**

OPEN ACCESS

ARTICLE PROCESSING CHARGES WAIVED UNTIL JULY 2023

DIVERSITY AND INNOVATION

Mothers of invention

Women are more likely than men to invent for women, but obstacles limit their participation in the innovation system

By **Fiona Murray**

In 1988, ophthalmologist Patricia Bath was granted US patent 4,744,360 for a laser-based "Apparatus for ablating and removing cataract lenses." Cataracts have disproportionately burdened women, and Bath had long been concerned about differences she saw among patient populations. She is not alone among innovators in focusing on solving problems that they see around them. But as one of the few female inventors of her generation, she had a rare vantage point. On page 1345 of this issue, Koning *et al.* (*1*) document a similar pattern on a large scale. They show that female inventors are more likely to produce patents to solve problems that specifically or disproportionately affect women (such as menopause and fibromyalgia). Beyond recognizing the loss of human talent that arises when women are underrepresented in innovation, this finding highlights the types of problems (and solutions) that are overlooked in the current system with its support of a homogeneous group of inventors.

The idea that people with distinctive life experiences—because of their gender, race, socioeconomic background, education, or nationality (to name a few)—will be at-

Massachusetts Institute of Technology, Sloan School of Management, Cambridge, MA, USA. Email: fmurray@mit.edu

PHOTO: LUISA DORR

Ophthalmologist Patricia Bath received US patents related to laser treatment for cataracts, which disproportionately affect women.

tracted to explore different problems might not be a surprise. Diverse inventors "see" the world differently. They will explore a solution space differently, making unusual connections linking otherwise disparate insights (*2*). What Koning *et al*. highlight is that diverse inventors also identify with different problems or missions—overcoming strategic blind spots that will otherwise occur if inventors are drawn from very similar backgrounds. Their approach, applied to over 440,000 US patents, is an entirely new way to map the problem spaces (i.e., applications) explored by inventors. This mapping convincingly shows the connection between inventor identity and invention domain across the life sciences (constituting over 10% of US patenting activity from 1976 to 2010). The authors show that all-female inventor teams are 35% more likely to innovate in areas of women's health than all-male teams (and female-majority teams are 18% more likely to have a women's focus than male-majority teams).

It is important to celebrate the particular problem focus and distinctive solutions that diverse innovators bring to the economy. Yet, to put these results into a wider perspective, Koning *et al*.'s analysis finds that women are represented on only 25% of US life sciences patents during a period when women have come to make up almost 50% of US PhDs in the life sciences. Such low levels of representation in the life sciences—from invention and patenting to start-up founding and fundraising to corporate board membership—illustrate the work that remains to overcome systemic barriers to diversity and inclusion in the life sciences (*3*).

These patterns of participation hold across the wider US innovation economy: Only 10% of patent inventors are women (*4*). Further along the innovation pipeline, less than 20% of tech start-up companies have a female cofounder, and female chief executive officers receive less than 7% of venture capital (*5*). At the same time, just as Koning *et al*. illustrate how female innovators are more likely to patent in female-focused areas, women are also more likely to found companies in some subfields such as educational technology. This "sorting" across the economy is echoed at the Massachusetts Institute of Technology, where we find that women are more likely to focus on mission- or purpose-driven sectors: health security, food security, and financial inclusion.

The concentration of women in certain fields provides new insights into the continued underrepresentation of women and minorities in innovation (*6*) but also raises new questions about the drivers that lead innovators into particular problem domains and missions. It is only when evaluating these dynamics at a system-wide level that greater diversity and inclusion in innovation can be effectively supported and the associated welfare losses in the economy stopped.

To understand the wider system context, it is necessary to consider two parts of the innovation system: One encompasses the domains prioritized (and funded) across the innovation landscape (and the mechanisms that go into defining those priorities). Another comprises the individuals selected as inventors and innovators to deliver those priorities (and the range of mechanisms that go into selecting key projects and people). At the intersection of these two parts of the innovation economy is a system in which the allocation of diverse talent to key domains may not simply be a preference for solving certain problems. Instead, it plausibly arises as system-wide beliefs about the innovation priorities for the economy and beliefs about the competence of innovators across fields, combine and misallocate talent.

Consider what funders are searching for—how problems and missions are scoped and prioritized both in early-stage science and in later-stage translation. It is notable that women's health has, until recently, been viewed as a marginal area of academic or commercial investment where few venture funders prioritized women's health or markets for female-oriented products.

A key system-level driver of these choices has been women's underrepresentation in the highest levels of decision-making (where markets or diseases are prioritized for investment) and in grant and investment making (where specific choices are made). Being absent shapes attention, or lack thereof, to these critical problems, leading to what Criado-Perez has called "invisible women" (*7*). More broadly, a lack of women in decision-making roles has likely shaped decisions—by predominantly male investors—to overlook key problems and markets. The same system dynamics are true for underrepresented minorities. Not having diverse voices at the table defining problem spaces perpetuates a system that narrows the projects and problems that are prioritized.

If female innovators identify particularly strongly with certain missions, or if they have specific solutions to these problems, then the lack of funding priority accorded to these areas will contribute to the low levels of women in the innovation economy. What funders search for—which investments they fund as priorities—shapes who is attracted into innovation. This system can only be overcome by expanding who sits at the table to set priorities.

Another aspect of the innovation system that must be attended to in driving diversity is more familiar: identifying, evaluating and selecting people to pursue innovation priorities. Many studies show that women (and other underrepresented groups) often receive lower evaluation scores when they submit research proposals, resumes, or business pitches—not because of a lack of quality, but because of assumptions being made about their capabilities (*8*), or because of the questions they are asked (*9*). To overcome these system-wide issues, it is necessary to consider the processes by which projects are selected: how solicitations are written, how venture pipelines are sourced, who is at the table during the evaluation process, what criteria are used, and how conversations are structured. Research shows that a wide range of voices needs to be heard in the selection process (*10*).

The criteria for selection are an especially critical part of how a system shapes diversity. In grant-making settings, decision-makers frequently rely on measures that are generally considered to be objective indicators of quality, e.g., citations, but are, in fact, biased in favor of over- not underrepresented groups (*2*). Thus, even signals of "merit" that are used to support innovation selections perpetuate a system that builds cumulative advantage for some groups over others. For example, if women receive smaller grants or smaller funding rounds early in their careers (*11*). This will have a cumulative long-term impact on productivity (in science or in commercialization), plausibly explaining the lower rates of invention observed by Koning *et al*. and across the economy. A "quick fix," e.g., training aimed at unconscious bias, is not enough. Considering who is at the table making selection decisions and the criteria they use are essential to system change.

The interaction between what is prioritized and who is selected has important, unintended, cumulative, system-wide consequences that magnify inequality at every stage—from grant funding, to patenting, to translation and venture funding. Women have lower levels of confidence when they consider entering some domains (*12*), causing them to shy away from certain fields—a dynamic that is magnified if they do not observe role models (*13*) or decision-makers with whom they identify. Similarly, if women believe that they have a disadvantage when pursuing funding in traditionally male-focused fields—whether those are male-focused diseases in life sciences or traditional sectors of the economy—they may be less likely to apply in the first place from "lack of fit" (*14*). These beliefs are not speculative; evidence from the start-up community shows that women's proposals were valued on average at

only two-thirds the value of equivalent proposals by males in traditionally male-focused business sectors. Men faced no such "valuation discount" in female-focused sectors (*15*). The outcome of such dynamics reinforces the representation patterns seen in life sciences patenting with women selecting more heavily into women's health.

When overlaying upon this system of selection a context in which diverse problem domains and missions are often underfunded or overlooked (for lack of effective diversity within priority setting bodies), then the drivers of underrepresentation on many dimensions in the innovation economy become clearer—albeit more daunting.

The logic of this analysis suggests a system in which the allocation of diverse talent to key domains may not simply be a preference for solving certain problems. As well, it arises as the result of system-wide beliefs about innovation priorities and beliefs about the competence of innovators across fields. If these beliefs are manifest early on in the innovation process, they are likely replicated at each subsequent stage, leading to cumulative disadvantage for underrepresented groups of women and other minorities and considerable welfare losses.

We must celebrate the positive impact of attracting some of the most talented women life scientists into critical and previously under-explored research areas such as women's health. Similarly, in areas of mission-driven innovation, the sense of purpose that drives underrepresented groups is a powerful force for change. But simply using traditional approaches to prioritize and select projects and people without appreciating the subtle dynamics at play across the entire system stifles the allocation of diverse talent across the entire innovation economy and does not bring the best and brightest to bear on the most pressing challenges. ■

REFERENCES AND NOTES

1. R. Koning *et al.*, *Science* **372**, 1345 (2021).
2. B. Hofstra *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **117**, 9284 (2020).
3. W. Ding *et al.*, *Acad. Manage. J.* **56**, 1443 (2013).
4. M. Delgado, F. Murray, in *Entrepreneurship and Innovation Policy and the Economy*, J. Lerner, S. Stern, Eds. (National Bureau of Economic Research, 2021).
5. Pitchbook, All In: Female Founders and CEOs in the US VC Ecosystem (Pitchbook, 2020).
6. A. Bell *et al.*, *Q. J. Econ.* **134**, 647 (2018).
7. C. Criado-Perez, *Invisible Women: Exposing Data Bias in a World Designed for Men* (Abrams, 2019).
8. A. W. Brooks, L. Huang, S. W. Kearney, F. E. Murray, *Proc. Natl. Acad. Sci. U.S.A.* **111**, 4427 (2014).
9. D. Kanze, L. Huang, M. A. Conley, E. T. Higgins, *Acad. Manage. J.* **61**, 586 (2018).
10. K. W. Phillips, *Sci. Am.* **311**, 42 (2014).
11. D. F. M. Oliviera *et al.*, *JAMA* **321**, 898 (2019).
12. E. Cech, B. Rubineau, S. Silbey, C. Seron, *Am. So. Rev.* **76**, 641 (2011).
13. M. Olsson, S. E. Martiny, *Front. Psychol.* **9**, 2264 (2018).
14. M. Heilman, *Res. Organ. Behav.* **5**, 269 (1983).
15. D. Kanze, M. A. Conley, T. G. Okimoto, D. J. Phillips, J. Merluzzi, *Sci. Adv.* **6**, eabd7664 (2020).

10.1126/science.abh3178

IMMUNOLOGY

Swarming motility in host defense

Neutrophils exhibit self-control to resolve infection and tissue damage

By **Briana L. Rocha-Gregg**[1] *and* **Anna Huttenlocher**[1,2]

Swarming is a collective movement that has far-reaching implications in biology and beyond. Swarms involve groups of individuals coordinating their behavior in a self-organizing process. Examples include swarming motility and quorum sensing in bacteria (*1*), foraging in ants, and the defensive actions of honey bees. In the animal immune system, swarms of white blood cells called neutrophils respond to microbial threats and tissue damage by traveling en masse to affected tissues. Although much work has been done to understand the initiation of neutrophil swarms (*2*), less is known about how they are terminated. On page 1303 of this issue, Kienle *et al.* (*3*) reveal that neutrophils limit their swarms by deactivating their own receptors that detect self-produced swarm signals. This desensitization limits neutrophil aggregation and, crucially, promotes pathogen clearance and resolution of acute inflammatory responses.

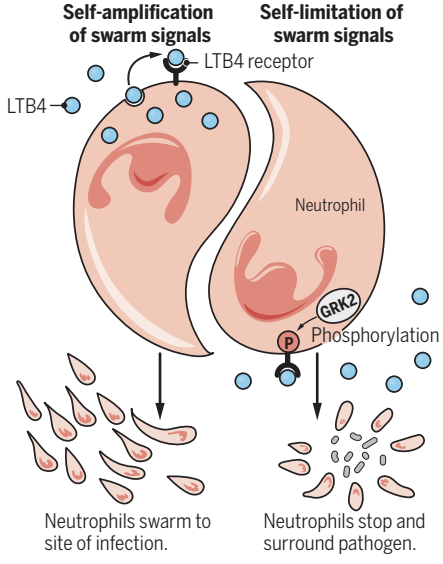
In response to infection or injury, neutrophils often travel long distances through complex chemical landscapes and diverse physical architectures. The long-range signals that initiate neutrophil recruitment to injured or infected tissues are well characterized (*4*) and include the modified lipid leukotriene B4 (LTB4) and chemokines such as interleukin-8 and CXCL2 (C-X-C motif chemokine ligand 2). As they migrate, neutrophils produce and secrete these same attractants, which amplify the recruitment signal and ensure that continued waves of neutrophils reach tissue damage through swarming migration (*5*). Indeed, a recent study identified a quorum-sensing type of neutrophil swarming toward zebrafish wounds through the local propagation of calcium signals and LTB4 release (*6*). This self-amplification is reminiscent of amoeba cells (*Dictyostelium discoideum*) and metastatic cancer cells. Studies of their migration through microfluidic mazes identified the diffusion and breakdown of attractants (*7*) as sources of self-generated chemoattractant gradients that contribute to these coordinated, self-propagated behaviors.

Given that neutrophil swarms are self-reinforced, how are they terminated? This is an intriguing question because recruitment signals persist even as neutrophils simultaneously arrive and leave damaged tissues—the latter by reverse migration (*4*). There has been growing interest in identifying pathways and mechanisms that promote resolution of neutrophilic inflammation. A number of neutrophil extrinsic, anti-in-

[1]Department of Medical Microbiology and Immunology, University of Wisconsin-Madison, Madison, WI 53706, USA. [2]Department of Pediatrics, University of Wisconsin-Madison, Madison, WI 53706, USA.
Email: huttenlocher@wisc.edu

Self-balancing act

Neutrophils swarm to affected tissue to sequester and kill invading microbes. Pathogen clearance depends on the balance between self-amplification and self-limitation of swarm signals [such as leukotriene B4 (LTB4)]. As neutrophils reach afflicted tissues, the concentration of swarm signals increases. As a result, cognate receptors are desensitized by G protein–coupled receptor kinase 2 (GRK2)–mediated phosphorylation. This ensures that neutrophils pause their motility and sequester and kill pathogens.



GRAPHIC: N. DESAI/*SCIENCE*

flammatory factors, including lipoxins and resolvins (*8*), are generated at injury sites and dampen neutrophil-mediated inflammation. Additionally, the production of reactive oxygen species (ROS) by host cells is necessary for defense against pathogens but also provides a cue that limits inflammation fueled by neutrophils. Chronic granulomatous disease (CGD) patients deficient in ROS production have sustained sterile (pathogen-free) inflammation, and their neutrophils form larger neutrophil swarms in vitro (*9*). Macrophages, another white blood cell present in wounds, cloak damage signals and prevent neutrophil swarming (*10*). They can also promote the reverse migration of neutrophils from sites of tissue damage (*11*).

Kienle *et al.* have identified an elegant, neutrophil-intrinsic mechanism that limits swarm size: negative regulation of the receptors that recognize self-produced swarm signals (see the figure). Notably, only receptors for intermediate-target attractants, like LTB4, are affected. Receptors for end-target attractants like C5a (complement component 5a) are unaffected. As a result, neutrophils remain sensitive to exogenous signals that promote functions necessary for pathogen killing. This self-limiting mechanism occurs through the activity of a cytoplasmic GRK-family protein, GRK2 (G protein–coupled receptor kinase 2). GRK proteins are cytoplasmic enzymes that phosphorylate activated G protein–coupled receptors. This phosphorylation results in receptor desensitization and, in some cases, internalization. Internalized receptors can be degraded or returned to the cell surface, as is the case with the CXCL2 receptor but not the LTB4 receptor. In this way, a cell can dynamically alter its sensitivity to various ligands.

It is tempting to think that by increasing swarms, more neutrophils would reach the wound, and the magnitude of their combined defenses would easily overcome the threat. This is not always the case. A particularly surprising observation by Kienle *et al.* is that persistent swarming did not result in better control of infection in mice with neutrophils lacking GRK2. In these animals, both increased cell speed and larger neutrophil clusters were observed at the wound. Without an adequate pause in motility, these neutrophils could not mount a successful defense response. Thus, a cell-intrinsic mechanism ensures that neutrophils successfully transition from the recruitment phase to the defensive phase. The study also suggests that a bigger swarm is not necessarily better in clearing pathogens. This is similar to what is seen in patients with CGD, where increased neutrophil swarming is associated with impaired microbial killing, although these effects may not be related.

Not all cellular swarms are beneficial to the body. Excessive or inappropriate neutrophilic inflammation is associated with debilitating diseases, including CGD and other autoinflammatory disorders. Additionally, the collective migration of cancer cells can drive metastasis, and self-propagating swarming may promote this behavior. Given the variety of contexts in which swarms occur, insights into their termination are of great general interest as well. Insect swarms can be both beneficial (bees) and devastatingly costly to agriculture (locusts). For decades, engineers and computer scientists have worked to incorporate aspects of swarm intelligence into technological applications. Robot swarms show promise in a variety of contexts, including environmental remediation (*12*).

At first glance, the model of self-control suggested by Kienle *et al.* seems deceptively simple. Long-range alarm signals trigger self-propagating neutrophil swarms that converge at sources of infection or injury. However, these swarms are intrinsically transient if receptors become desensitized as ligand concentrations increase close to target sites. The selectivity of this response enables neutrophils to prioritize signals that induce effector functions essential for clearing the infection and limiting collateral tissue damage. It is unclear how neutrophils prioritize signals to arrive at and depart from damaged tissues simultaneously (*4*). A tissue factor—myeloid-derived growth factor—has recently been identified that promotes neutrophil reverse migration and limits neutrophil inflammation (*13*). The signals that orchestrate and integrate these types of complex behaviors in homeostasis and disease are still largely unknown and likely involve the combined influence of self-propagated and exogenous cues. ■

REFERENCES AND NOTES

1. S. Mukherjee, B. L. Bassler, *Nat. Rev. Microbiol.* **17**, 371 (2019).
2. K. Kienle, T. Lämmermann, *Immunol. Rev.* **273**, 76 (2016).
3. K. Kienle *et al.*, *Science* **372**, eabe7729 (2021).
4. S. de Oliveira, E. E. Rosowski, A. Huttenlocher, *Nat. Rev. Immunol.* **16**, 378 (2016).
5. T. Lämmermann *et al.*, *Nature* **498**, 371 (2013).
6. H. Poplimont *et al.*, *Curr. Biol.* **30**, 2761 (2020).
7. L. Tweedy *et al.*, *Science* **369**, eaay9792 (2020).
8. S. E. Headland, L. V. Norling, *Semin. Immunol.* **27**, 149 (2015).
9. H. Alex *et al.*, *Nat. Commun.* **11**, 1 (2020).
10. S. Uderhardt, A. J. Martins, J. S. Tsang, T. Lämmermann, R. N. Germain, *Cell* **177**, 541 (2019).
11. S. Tauzin, T. W. Starnes, F. B. Becker, P. Y. Lam, A. Huttenlocher, *J. Cell Biol.* **207**, 589 (2014).
12. Y. Zhang, K. Yan, F. Ji, L. Zhang, *Adv. Funct. Mater.* **28**, 1806340 (2018).
13. R. A. Houseright *et al.*, *J. Cell Biol.* **220**, e202103054 (2021).

10.1126/science.abj3065

SIGNALING

Preparing macrophages for the future

Temporal dynamics of a key immune transcription factor shape the epigenome and future cell responses

By **Nagarajan Nandagopal, Ashwini Jambhekar, Galit Lahav**

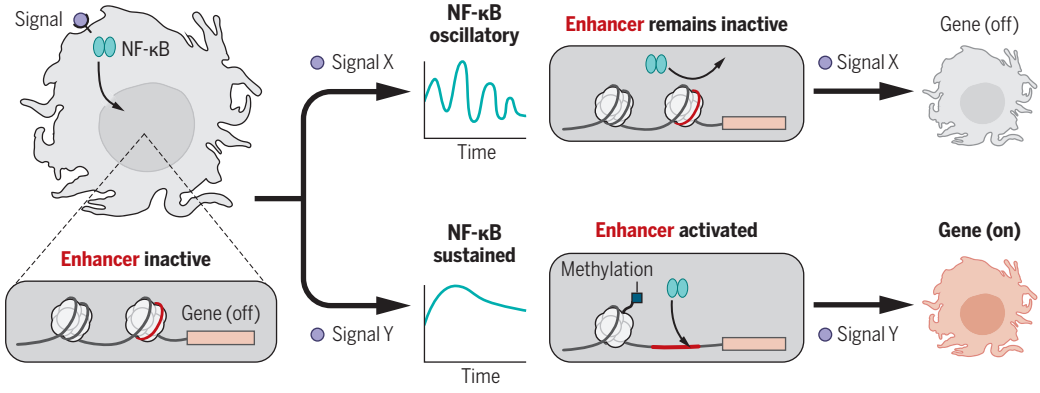
Cells live in complex environments and must respond appropriately to extracellular signals. Such responses often involve regulating the expression of hundreds of genes through transcription factors (TFs). Many TFs are activated by multiple signals and regulate the expression of distinct genes in response to each. How extracellular information is "encoded" in TF activity and subsequently "decoded" to orchestrate gene expression is a fundamental question in biology. Intriguingly, some TFs such as nuclear factor κB (NF-κB) and p53 encode signaling information in their temporal dynamics (*1*). Studies have shown that signaling dynamics can be used to control the induced expression levels (*2*), types (*3*), or ratios (*4*) of genes. On page 1349 of this issue, Cheng *et al.* (*5*) report a previously unknown role for TF dynamics: They show that NF-κB dynamics not only control how genes respond in the present but also reconfigure the cell to control gene expression in response to future stimulation.

In macrophages, which act as sentinel cells of the innate immune system, NF-κB dynamics was shown to encode signal identity. For example, activation by toxins released from invading bacteria leads to sustained NF-κB activity, whereas activation by inflammatory signals from other immune cells leads to oscillations in NF-κB activity (*6*, *7*). Rather than focusing directly on which genes' expression are induced by NF-κB, and by how much, Cheng *et al.* analyzed the NF-κB "epigenome," a set of factors that influences the potential for expression of genes to be induced upon TF activation.

Department of Systems Biology, Blavatnik Institute, Harvard Medical School, Harvard University, Boston, MA, USA. Email: galit@hms.harvard.edu

Signal-specific nuclear factor κB dynamics

In macrophages, the transcription factor nuclear factor κB (NF-κB) can be activated either with oscillatory or sustained temporal dynamics, depending on the signal. Under oscillatory dynamics, inactive enhancers remain nucleosome-bound. By contrast, sustained dynamics can remodel enhancers through nucleosome displacement and histone 3 Lys^4 (H3K4) methylation. After a second exposure to the same signal, enhancers that were activated by sustained NF-κB dynamics induce expression of their associated genes, whereas these genes remain inactive under oscillatory dynamics.



The status of "enhancers," DNA sequences bound by TFs to control gene expression, is a particularly important aspect of the epigenome. In some cases, enhancers are bound by nucleosomes, which obstruct TF binding. Such enhancers cannot activate genes until the nucleosomes are displaced. Other enhancers are readily available for TFs to bind and facilitate gene expression and are considered to be active. The set of genes whose expression can be induced and their expression levels depend on how these two classes of enhancers are distributed in the genome. This layer of regulation provides a means for different cell types to regulate different sets of genes by using the same TFs. It also enables a cell to change its response to the same signal and thus adapt to the environment.

Cheng *et al.* investigated the status of NF-κB–bound enhancers in macrophages derived from mice after stimulating the cells with signals previously shown to activate NF-κB with different dynamics. They found that signals that led to sustained NF-κB activity increased the number of active enhancers compared with signals that produced oscillatory activity. To directly test whether this difference was due to NF-κB dynamics, the authors blocked normal oscillations by disengaging a well-characterized negative-feedback loop in the NF-κB pathway, which did not alter other features of the response such as activation intensity. Removing NF-κB oscillations in this manner increased the number of active NF-κB–dependent enhancers. Thus, although both oscillatory and non-oscillatory NF-κB dynamics induce gene expression through already active enhancers, non-oscillatory activation also reconfigures the epigenome by activating additional enhancers.

What effect does this have on the cell? Because the epigenome defines the set of genes whose expression can be induced by NF-κB, one scenario is that by increasing the number of active enhancers, non-oscillatory NF-κB dynamics change the set of genes that could respond to future stimulation of the pathway. Indeed, analysis of gene expression after a second phase of sustained NF-κB activation showed induction of hundreds of genes not induced upon repeat stimulation of oscillatory NF-κB (see the figure). Thus, it appears that the functions of NF-κB as an inducer of gene expression and as a modifier of the epigenome (likely in conjunction with other proteins) are separable according to its temporal dynamics. On the basis of modeling, the authors suggest that this separation is achieved through the many steps involved in unwrapping inactive enhancers from nucleosomes, which demands persistent nuclear NF-κB—that is, non-oscillatory dynamics.

"...NF-κB dynamics not only control how genes respond in the present but also reconfigure the cell to control gene expression in response to future stimulation."

Modifying the epigenome provides NF-κB with the ability to record its activation and affect future cellular responses. Other forms of this phenomenon, broadly called epigenetic transcriptional memory, have been described (*8*), including in macrophages (*9*). It will be interesting to compare enhancer activation by NF-κB to other mechanisms of transcriptional memory, in terms of stability, fidelity, and lifetime. At a mechanistic level, it remains to be determined how NF-κB activity duration translates to nucleosome displacement and how the proposed multistep reaction mechanism compares with classic kinetic proofreading (*10*) or circuit-based mechanisms for duration sensing (*11*). The responses of target genes to other TFs are sensitive to oscillation frequency and duration (*12*); the relevance of the proposed mechanism to these pathways will be worth investigating.

At the level of the cellular response, the functional consequences of organizing NF-κB target genes into multiple cohorts are currently unclear. Does it lead to macrophage adaptation to particular immune threats, as suggested previously (*9*, *13*)? Maybe this response can be "tuned" by changing the duration or intensity of the initial stimulus. The discovery by Cheng *et al.* thus opens up several new lines of inquiry that will help to better understand how NF-κB orchestrates specific responses to different stimuli. More generally, other oscillatory TFs have known roles in regulating the epigenome (*14*, *15*). It will be fascinating to see whether similar principles apply to decoding dynamics and controlling the activation of target genes in other systems. ■

REFERENCES AND NOTES

1. J. E. Purvis, G. Lahav, *Cell*. **152**, 945 (2013).
2. K. Lane *et al.*, *Cell Syst*. **4**, 458 (2017).
3. J. E. Purvis *et al.*, *Science* **336**, 1440 (2012).
4. L. Cai, C. K. Dalal, M. B. Elowitz, *Nature* **455**, 485 (2008).
5. Q. J. Cheng *et al.*, *Science* **372**, 1349 (2021).
6. M. W. Covert, T. H. Leung, J. E. Gaston, D. Baltimore, *Science* **309**, 1854 (2005).
7. D. E. Nelson *et al.*, *Science* **306**, 704 (2004).
8. A. D'Urso, J. H. Brickner, *Curr. Genet*. **63**, 435 (2017).
9. S. L. Foster, D. C. Hargreaves, R. Medzhitov, *Nature* **447**, 972 (2007).
10. J. J. Hopfield, *Proc. Natl. Acad. Sci. U.S.A.* **71**, 4135 (1974).
11. J. Gerardin, N. R. Reddy, W. A. Lim, *Cell Syst*. **9**, 297 (2019).
12. M. D. Harton, W. S. Koh, A. D. Bunker, A. Singh, E. Batchelor, *Mol. Syst. Biol*. **15**, e8685 (2019).
13. M. G. Netea *et al.*, *Nat. Rev. Immunol*. **20**, 375 (2020).
14. A. Hafner, M. L. Bulyk, A. Jambhekar, G. Lahav, *Nat. Rev. Mol. Cell Biol*. **20**, 199 (2019).
15. N. Yissachar *et al.*, *Mol. Cell*. **49**, 322 (2013).

ACKNOWLEDGMENTS

This work was supported by National Institutes of Health grant R35GM139572. N.N. was supported by a fellowship from the Damon Runyon Cancer Research Foundation.

10.1126/science.abj2040

GRAPHIC: K. FRANKLIN/*SCIENCE*

NEUROSCIENCE

White matter and human behavior

Genetic variants influence structural connectivity with intriguing clinical implications

By **Christopher M. Filley**

One of the most enduring themes in human neuroscience is the association of higher brain functions with gray matter. In particular, the cerebral cortex—the gray matter of the brain's surface—has been the primary focus of decades of work aiming to understand the neurobiological basis of cognition and emotion. Yet, the cerebral cortex is only a few millimeters thick, so the relative neglect of the rest of the brain below the cortex has prompted the term "corticocentric myopia" (*1*). Other regions relevant to behavior include the deep gray matter of the basal ganglia and thalamus, the brainstem and cerebellum, and the white matter that interconnects all of these structures. On page 1304 of this issue, Zhao *et al.* (*2*) present compelling evidence for the importance of white matter by demonstrating genetic influences on structural connectivity that invoke a host of provocative clinical implications.

Insight into the importance of white matter in human behavior begins with its anatomy (*3*–*5*) (see the figure). White matter occupies about half of the adult human brain, and some 135,000 km of myelinated axons course through a wide array of tracts to link gray matter regions into distributed neural networks that serve cognitive and emotional functions (*3*). The human brain is particularly well interconnected because white matter has expanded more in evolution than gray matter, which has endowed the brain of *Homo sapiens* with extensive structural connectivity (*6*). The myelin sheath, white matter's characteristic feature, appeared late in vertebrate evolution and greatly increased axonal conduction velocity. This development enhanced the efficiency of distributed neural networks, expanding the transfer of information throughout the brain (*5*). Information transfer serves to complement the information processing of gray matter, where neuronal cell bodies, synapses, and a variety of neurotransmitters are located (*5*). The result is a brain with prodigious numbers of both neurons and myelinated axons, which have evolved to subserve the domains of attention, memory, emotion, language, perception, visuospatial processing, executive function (*5*), and social cognition (*7*).

White matter is important in clinical medicine, and knowledge of its normal structure and function informs the clinical interpretation of acquired brain lesions to which it is vulnerable. Because all brain regions are interconnected by white matter, an understanding of the genetic architecture of normal tracts is crucial for advancing the neurobiology of human behavior (*2*). Combined with data from major initiatives such as the Human Connectome Project, launched in 2010 by the National Institutes of Health with the goal of mapping all the long-distance connections in the brain (*8*), the role of genetics is pivotal in building a more complete understanding of normal white matter and its connectivity.

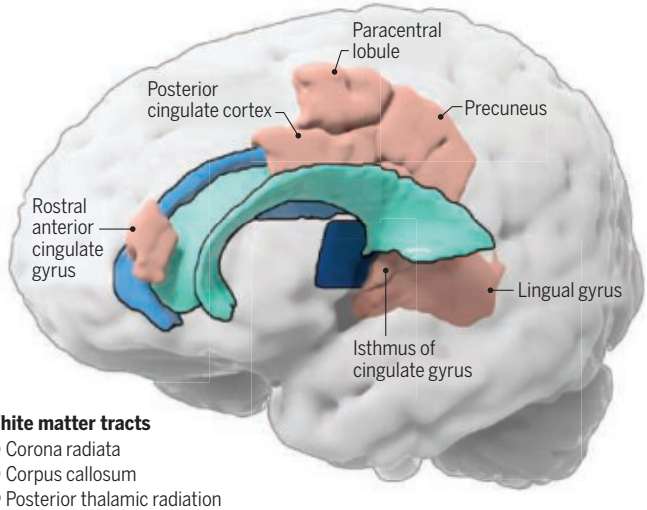
Damage to white matter can disrupt any brain function and produce profound clinical consequences. Neurology has long appreciated the effects of such damage on sensory and motor function, but more-recent work has amply demonstrated the capacity of white matter lesions to affect cognition and emotion (*4*, *5*, *9*–*11*). White matter pathology with neurobehavioral consequences can occur in a wide variety of acquired and genetic disorders affecting myelin and has also been recognized to occur in neurodegenerative disorders such as Alzheimer's disease (AD) that have long been thought to originate in cortical gray matter (*4*, *5*).

A particularly instructive category of white matter disorder is toxic leukoencephalopathy (TL), in which toxic white matter injury can produce devastating clinical sequelae (*9*). The best example of TL is toluene leukoencephalopathy, in which long-term abusers of paint fumes sustain diffuse damage to lipid-rich myelin, owing to the lipophilicity of toluene, and often develop a disabling cognitive syndrome called white matter dementia (*10*). As toluene abuse causes widespread myelin loss while sparing the gray matter, this disorder illustrates the profound cognitive effects of selective white matter damage (*9*–*11*).

Zhao *et al.* add to our knowledge of white matter with an analysis of genetic and neuroimaging data from nearly 44,000 individuals within five data resources. Focusing on 21 predefined tracts, they identified 109 loci associated with white matter microstructure. Genetic correlations were observed between white matter microstructure and a wide spectrum of diseases, and genetic variation was found to alter the function of oligodendrocytes—the glial cells responsible for myelination. Among the many diseases identified was AD, adding to mounting evidence that white matter dysfunction may be key to its pathogenesis. An influential "myelin model" postulates early white matter involvement in late-onset disease (*12*), and, even in early-onset autosomal dominant AD, microstructural white mat-

White matter tracts

Select white matter tracts in the human brain are shown. These fibers connect gray matter regions to mediate complex human behaviors. The corpus callosum connects the cerebral hemispheres to enable cognitive and emotional information transfer, the corona radiata contributes to motor functions, and the posterior thalamic radiation is involved in visual processing.



GRAPHIC: V. ALTOUNIAN/*SCIENCE*. (DATA) CHENG *ET AL.* (5) AND RESEARCH IMAGING INSTITUTE, UTHSCSA

Behavioral Neurology Section, Marcus Institute for Brain Health, Department of Neurology, University of Colorado School of Medicine, Aurora, CO 80045, USA. Email: christopher.filley@cuanschutz.edu

ter changes have been observed to precede symptom onset (*13*). Other neuropsychiatric conditions also implicate white matter, including schizophrenia, depression, attention deficit hyperactivity disorder, autism, amyotrophic lateral sclerosis, glioma, and stroke. Relevant to both stroke and AD, white matter hyperintensities often seen on magnetic resonance imaging scans of older people (*5*) have been associated with a locus on chromosome 17 (*14*).

Zhao *et al.* found that many commonly used centrally-active medications exert effects on genes associated with white matter microstructure—an observation that may lead to improvements in the treatment of many brain diseases. The pharmacology of drugs used for neuropsychiatric disorders is not well understood, and knowledge of the interactions of these drugs with white matter neurobiology may substantially bolster the clinician's armamentarium.

The emerging recognition of white matter and its contribution to human behavior will advance medicine as well as neuroscience. Considering both environmental and genetic factors clarifies the structure and function of normal and abnormal tracts, and this knowledge promises in turn to improve the diagnosis and treatment of people in whom white matter dysfunction may be disturbing neurobehavioral capacity. Moreover, the understanding of AD and many disabling neuropsychiatric disorders may be transformed by a focus on microstructural pathology in myelinated tracts. Broadly, a more detailed understanding of the relationships between white matter and behavior will surely expand knowledge of the brain. A complete portrait of the structural basis of cognition and emotion cannot neglect the white matter because it interacts so intimately with its gray matter counterpart. ■

REFERENCES AND NOTES

1. J. Parvizi, *Trends Cogn. Sci.* **13**, 354 (2009).
2. B. Zhao *et al.*, *Science* **372**, eabf3736 (2021).
3. J. L. Saver, *Stroke* **37**, 263 (2006).
4. J. D. Schmahmann, D. Pandya, *Fiber Pathways of the Brain* (Oxford Univ. Press, 2006).
5. C. M. Filley, R. D. Fields, *J. Neurophysiol.* **116**, 2093 (2016).
6. K. Zhang, T. J. Sejnowski, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 5621 (2000).
7. Y. Wang, I. R. Olson, *Trends Cogn. Sci.* **22**, 504 (2018).
8. M. F. Glasser *et al.*, *Nat. Neurosci.* **19**, 1175 (2016).
9. C. M. Filley, B. K. Kleinschmidt-DeMasters, *N. Engl. J. Med.* **345**, 425 (2001).
10. C. M. Filley, G. M. Franklin, R. K. Heaton, *Neuropsychiatry Neuropsychol. Behav. Neurol.* **1**, 239 (1988).
11. J. D. Schmahmann, E. E. Smith, F. S. Eichler, C. M. Filley, *Ann. N. Y. Acad. Sci.* **1142**, 266 (2008).
12. G. Bartzokis, *Neurobiol. Aging* **32**, 1341 (2011).
13. M. Á. Araque Caballero *et al.*, *Brain* **141**, 3065 (2018).
14. M. Fornage *et al.*, *Ann. Neurol.* **69**, 928 (2011).

10.1126/science.abj1881

A new model helps explain ice cliff collapse, such as the ice calving from the Perito Moreno Glacier, Argentina, in 2010.

PHOTO: GABRIEL ROJO/MINDEN PICTURES

GLACIOLOGY

Is the marine ice cliff hypothesis collapsing?

An improved rheologic model shows that glacier retreat may not always be quite so quick

By **Nicholas R. Golledge**[1] *and* **Daniel P. Lowry**[2]

Marine margins, where ice sheets flow from land into the ocean, become exposed to environmental conditions and internally generated forces markedly different from those governing the flow of ice further inland. Prior research has suggested that if the retreat of such margins formed tall ice cliffs, they may become structurally unstable and collapse, leading to further retreat (*1*). Although the veracity and importance of the "marine ice-cliff instability" are still uncertain (*2*), increasing the accuracy of simulated ice cliff processes is a key challenge because the Greenland and Antarctic ice sheets could raise global sea level by 65 m. On page 1342 of this issue, Bassis *et al.* (*3*) developed a model that reliably captures the complex behavior of ice cliffs as they deform and fracture. In doing so, they find that marine-terminating parts of Antarctica may be less vulnerable than previously suggested to rapid and irreversible collapse (*4*, *5*).

Ice sheets terminating in the ocean lose mass through melting either at the surface or on the underside of floating ice, and through calving of icebergs. For the calving fronts of thick glaciers to remain stable, they need to be grounded in deep water (*6*). Thicker ice or shallower water increases the height of the ice cliff exposed above water, and where this exceeds the threshold for structural stability (~100 m), the cliff may collapse through slumping (*7*) (see the fig-

[1]Antarctic Research Centre, Victoria University of Wellington, Wellington 6140, New Zealand.
[2]GNS Science, Avalon, Lower Hutt 5011, New Zealand.
Email: nicholas.golledge@vuw.ac.nz

ure). The rate at which such calving proceeds is governed by the yield strength of ice (*8*), as well as the geometric configuration of the glacier front and the environmental forcing.

Experimentation has also revealed that the critical height required for collapse of a marine ice cliff depends on how rapidly it becomes exposed (*9*). The sudden break-up of buttressing ice shelves, such as witnessed in the Antarctic Peninsula over recent decades, could expose a new ice cliff within days or weeks. This "preconditioning" means that cliffs could fail at lower heights than if ice shelf collapse proceeded more slowly. A "speed limit" for how quickly an ice shelf can collapse has been recently computed (*10*), but one of the key difficulties in accurately modeling or predicting any ice cliff failure that then follows is that the calving process itself comprises two distinct modes of failure. Thick and undamaged ice has a high tensile strength and tends to fail through shearing (deformation) (*11*), whereas rapidly sliding glaciers are prone to tensile failure (fracturing) because of their extensional flow regime and the intersection of crevasses that develop as the glacier is "stretched."

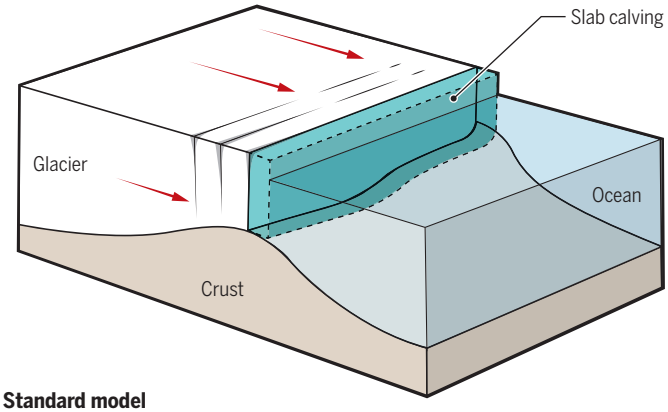
This dual mode of failure is where Bassis *et al.* make such an important advance. The authors' "m-ice" model simulates ice as a viscous material whose deformation rate is described by a power law only until its yield strength is reached, at which point it then deforms rapidly. Accumulation of plastic strain in the failed ice reduces its strength, allowing it to deform even more rapidly, progressively localizing failure within discrete bands. Thus, although not explicitly modeling the brittle processes of ice fracturing, the m-ice model uses a pioneering composite rheology that allows the flow and failure modes of ice cliff calving to be captured in a single continuum framework.

Bassis *et al.* use their model to explore the environmental and geometric controls on how a range of marine-terminating glacier margins might evolve. The authors find that resistive forces at the ice front, caused either by sea ice or the calved debris that typically chokes narrow fjords during winter months, can slow or even completely prevent the retreat of an ice cliff. But even without this kind of buttressing, Bassis *et al.* conclude that ice cliffs may be inherently stable if the speed of ice flow or slope of the bed underlying the ice are within certain bounds. Specifically, the authors find that the upstream gradient of ice thickness exerts a first-order control on the tendency of an ice cliff to reach a critical (collapse-prone) height, and that the rate at which ice flows toward the calving front exerts a secondary control. For the majority of thickness gradients considered, faster-flowing ice allows ice margins to advance, whereas slower-flowing ice produces cliff retreat. Where ice thickness gradients are greater than about 30 m per kilometer, ice cliffs tend to collapse regardless of the inflowing ice velocity.

The idea that brittle failure of vertical ice cliffs could lead to catastrophic collapse of a large part of the West Antarctic Ice Sheet has recently been put forward and is used to project sea level contributions from Antarctica of as much as 1 m by 2100 (*4*), under a future greenhouse gas emissions scenario that leads to an increase in radiative forcing of 8.5 W/m^2. These studies have done much to catalyze and focus research efforts within the glaciological community, and to highlight to policy-makers the deep uncertainty associated with Antarctic ice sheet retreat in particular. But parameterized models that use extrapolations from sparse observations of a poorly understood process mean that resulting predictions of retreat rates vary considerably between model versions (*1*, *4*, *5*), so the impact of ice cliff collapse on future sea level rise remains uncertain (*2*, *12*). What is certain, however, is that if the processes involved in ice cliff collapse are to be included in models used for future sea level projections, they need to incorporate robust physics and be well-validated by comparison to observations. Bassis *et al.* have made a substantial leap forward in both of these areas with the m-ice model that offers exciting avenues for better understanding ice cliff behavior. ■

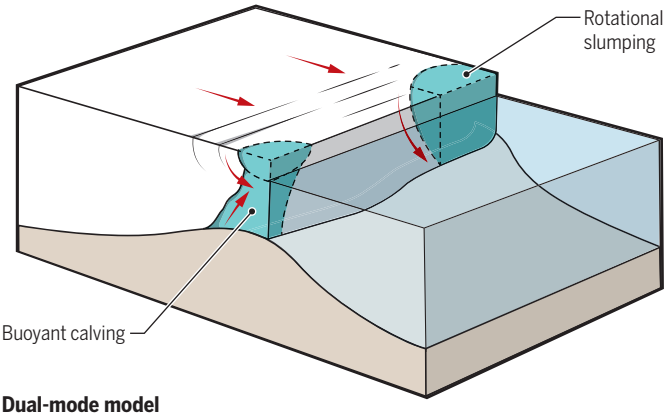
A refined mechanism of marine ice cliff retreat

The manner in which ice cliffs fail is important for projecting glacier retreat in a warming environment.



Standard model

Standard ice cliff modeling assumes no vertical difference in the internal stress. This leads to failure at a critical height above the ocean surface as a result of tensile stresses. The whole ice column retreats by a time-averaged horizontal wastage rate.



Dual-mode model

Dual-mode modeling simulates differences in the internal stress regime within the vertical ice column. This continually tracks the accumulation of plastic strain during deformation and allows zones of damaged ice to develop, which localize subsequent failure. The failure occurs by slumping, or by slumping and buoyant uplift of submerged blocks if the water is deeper.

GRAPHIC: C. BICKEL/*SCIENCE* BASED ON N. GOLLEDGE/ANTARCTIC RESEARCH CENTRE

REFERENCES AND NOTES

1. D. Pollard, R. M. DeConto, R. B. Alley, *Earth Planet. Sci. Lett.* **412**, 112 (2015).
2. M. Meredith *et al.*, IPCC Special Report on the Ocean and Cryosphere in a Changing Climate, H.-O. Pörtner *et al.*, Eds. (IPCC, 2019), pp. 203–320.
3. J. N. Bassis *et al.*, *Science* **372**, 1342 (2021).
4. R. M. DeConto, D. Pollard, *Nature* **531**, 591 (2016).
5. R. M. DeConto *et al.*, *Nature* **593**, 83 (2021).
6. J. N. Bassis, C. Walker, *Proc. R. Soc. A* **468**, 913 (2012).
7. B. R. Parizek *et al.*, *Geology* **47**, 449 (2019).
8. J. Bassis, L. Ultee, *J. Geophys. Res. Earth Surf.* **124**, 2036 (2019).
9. F. Clerc, B. M. Minchew, M. D. Behn, *Geophys. Res. Lett.* **46**, 12108 (2019).
10. A. A. Robel, A. F. Banwell, *Geophys. Res. Lett.* **46**, 12092 (2019).
11. Y. Ma, C. S. Tripathy, J. N. Bassis, *Geophys. Res. Lett.* **44**, 1369 (2017).
12. T. L. Edwards *et al.*, *Nature* **566**, 58 (2019).

ACKNOWLEDGMENTS

This work was funded by the Royal Society Te Apārangi contract RDFVUW1501; New Zealand Ministry for Business, Innovation and Employment contracts RTUV1705 ("NZSeaRise"); and ANTA1801 ("Antarctic Science Platform").

10.1126/science.abj3266

RETROSPECTIVE

Marye Anne Fox (1947–2021)

Groundbreaking organic chemist and educator

By **Robert E. Continetti**

Marye Anne Fox, a leading force in the development of organic photochemistry, died on 9 May at the age of 73. Fox conducted groundbreaking research with applications in renewable energy and environmental chemistry, and she was a gifted university administrator and national leader. A pioneering woman in a male-dominated field, Fox freely gave advice to early-career women about how to become successful chemists while balancing both career and family. She was an outstanding educator who taught chemistry at all levels from middle school to graduate school, and she was an effective advocate for broadening inclusion in chemistry education and beyond.

Born in Canton, Ohio, on 9 December 1947, Fox was inspired by Sputnik to pursue a career as a scientist. She received her BS in chemistry from Notre Dame College in Ohio in 1969, her MS degree in chemistry from Cleveland State University in 1970, and her PhD in chemistry in 1974 from Dartmouth College in New Hampshire. In 1976, Fox joined the faculty at the University of Texas at Austin as the first female assistant professor and rose through the ranks; she was named to her first endowed chair in 1986 and was appointed as the first vice president for research in 1994. She then served as chancellor at both North Carolina State University and University of California San Diego (UCSD) before her retirement in 2012.

Fox made fundamental discoveries related to the mechanisms of chemical reactions that were initiated by light absorption or the application of electric fields. She built a career that went beyond the traditional bounds of condensed-phase organic chemistry to enable applications in biological and materials chemistry. A hallmark of her research was a desire to understand and exploit the chemistry of electronically excited states and charge transfer processes in molecules and molecular assemblies. She is credited with the origination of organic photoelectrochemistry, a technique that can purify organic-contaminated materials by exploiting wide bandgap semiconductor substrates to control the initiation of oxidative reactions. Fox aimed to address the grand challenge of efficient solar energy conversion, and her work on photocells that produce both electrical current and hydrogen gas achieved that goal by facilitating the development of clean energy sources.

Moving into ever-larger systems while maintaining a focus on mechanisms, Fox's laboratory gained insights into the role that electron–hole pair formation plays in conductivity at semiconductor–liquid interfaces in a variety of materials. This work yielded a panoply of tools for influencing interfacial electron transfer over long distances. The contributions made by Fox and her co-workers will have a lasting impact on these important fields.

In addition to being a distinguished researcher and administrator, Fox was a gifted teacher and mentor. She was a certified secondary school chemistry and mathematics teacher in Ohio, and her teaching experiences in a disadvantaged school in East Cleveland shaped her understanding of education, the effort required for classroom instruction, and the principle that communicating clear expectations for strong student performance was essential. In her academic career, she supervised the graduate research of more than 75 students and many postdocs, and she coauthored a widely adopted textbook on organic chemistry that focused on the mechanistic approach, the hallmark of her own research program.

As the scientific impact of her research grew, Fox was increasingly called on to serve a broader mission. As vice president for research at the University of Texas at Austin, she supported the creation of new interdisciplinary institutes and improved technology transfer. As the 12th chancellor of North Carolina State University, she promoted the expansion of research, the educational mission, and the commercialization of intellectual property for the benefit of society. A science teaching laboratory building on that campus is now named in her honor.

In 2004, Fox became the first woman to serve as permanent chancellor of UCSD. We met while I was serving as chair of the Department of Chemistry and Biochemistry. I found her to be an inspiring role model who persevered through challenging times with vision and a steady hand. She had a calm demeanor that lent itself well to engagement, and she was a gifted administrative leader. She led the university during an era of extraordinary campus growth while dealing with unprecedented financial challenges. She used her position to make UCSD one of the greenest campuses in the nation and to combat racism on campus. She engaged in dialogue with student leaders and empowered the campus community to address difficult issues. By instituting systemic changes in leadership with a focus on improving the campus climate, she paved the way for the establishment of a vice chancellor for equity, diversity, and inclusion to promote sustained progress. She also championed programs to support outreach to underrepresented communities in the broader San Diego region.

Fox served as an adviser to Governor George W. Bush of Texas and later served the nation on the National Science Board and the President's Council of Advisors on Science and Technology. She was elected, in 1994, to both the US National Academy of Sciences and the American Academy of Arts and Sciences and, in 1996, to the American Philosophical Society. In 2010, she was awarded the National Medal of Science by President Barack Obama, the highest honor bestowed by the United States on a scientist.

Fox was a scientific pioneer in physical organic chemistry, an exceptional educator, and a gifted academic leader. She was motivated by a desire to discover and build, and thereby leave the world a better place. She was a role model for aspiring young scientists, especially women. Her efforts were fruitful, and the discipline of chemistry, institutions of higher education she served, and society overall continue to benefit immensely from her contributions. ■

Department of Chemistry and Biochemistry, University of California San Diego, La Jolla, CA 92093, USA.
Email: rcontinetti@ucsd.edu

10.1126/science.abj7318

PHOTO: JIM COLT/UC SAN DIEGO PUBLICATIONS

BOOKS *et al.*

Nahua people, shown here celebrating Earth Day in 2016, still speak Nahuatl, the language of the Aztecs.

PHOTO: MARVIN RECINOS/AFP/GETTY IMAGES

ANTHROPOLOGY

Examining the Aztecs

Skilled naturalists and healers, the Mesoamerican society left behind a rich cultural legacy

By **Andrew Robinson**

In Mexico, the Aztecs are far from lost, argues archaeologist and anthropologist Frances Berdan in *The Aztecs*, her contribution to the Lost Civilizations series. The current Mexican flag shows an eagle and a serpent on top of a cactus—a clear reference to a famous 16th-century Aztec depiction of the founding of their capital city, Tenochtitlan (now the site of Mexico City), in 1325. Moreover, the Aztec language, Nahautl, is still spoken by 2 million people. The Nahautl word for "email," for example, is "tepozmecaixtlatiltlahcuilloli" ("apparatus where writing is delivered to your face").

Knowledge of the pictographic and allusive Aztec script—which is no longer used, unlike its spoken counterpart—is to a great extent lost, however. The script has been reconstructed several times by various scholars, most recently by Gordon Whittaker in his painstaking book *Deciphering Aztec Hieroglyphs* (*1*). Spanish sources have proved crucial to the decipherment. The eagle-on-a-cactus hieroglyph, for instance, appears in the first folio of a postconquest Aztec manuscript known as the *Codex Mendoza*—created during the rule of Antonio de Mendoza, viceroy of New Spain from 1535 to 1550, following the Spanish conquest of the Aztec empire in 1521—in which the Spanish annotator comments that the image stands for Tenochtitlan, which "means, in our Castilian, 'prickly pear cactus growing on a stone.'"

We are lucky that the codex survived, notes Berdan. After its completion around 1541, it is believed to have been carted by mule train down from the mountains to Veracruz, where it was loaded onto a Spanish treasure ship headed to Spain. The ship apparently fell victim to a seaborne attack by French privateers, and the codex ended up in France in the hands of a cleric and cosmographer to King Henry II. Having passed through four subsequent owners, all of whom fortunately acted as good stewards, it eventually came into the collections of the Bodleian Library in Oxford in 1659. Published in the mid-19th century, it appeared in 1992 in a definitive facsimile edition edited by Berdan and anthropologist Patricia Rieff Anawalt.

From this and other early colonial sources, we learn that the Aztecs had a strong interest in the natural world, including the heavens, although there is no evidence that they separated the natural from the supernatural. "They drew on their predecessors' fount of knowledge based on millennia of celestial observations," writes Berdan, including solar and lunar movements, patterns of eclipses, the repeated appearance of comets, and the occurrence of meteor showers. Using this information, they calculated the lengths of the lunar month and the solar year and the synodic period of Venus. Instead of a man in the Moon, the Aztecs saw a rabbit. (According to their creation mythology, the gods hurled a rabbit onto the Moon to dim it.)

Aztec physicians developed treatments for headaches, stomachaches, coughs, fevers, parasites, skin sores, insomnia, and unstable mental states, as well as for snakebites, broken skulls, and severed noses. According to one study, 85% of 118 plants used by the Aztecs that are "ethnohistorically identified with curative properties" are efficacious in modern medical terms (*2*). The hot sap of the maguey (agave), for example, was applied to wounds and is known today to inhibit bacterial growth. "Without the concept of bacteria at hand, the Aztecs could not phrase this cure in those terms, but their experience taught them that it worked," notes Berdan.

Most of this knowledgeable and accessible introduction to the Aztecs—the fruit of a lifetime's study—is concerned with matters such as food and drink, textiles and dress, pottery and art objects, dwellings and architecture, the social divisions of society, trade and the economy, religion and mythology, and, inevitably, the notorious Aztec penchant for human sacrifice. This latter custom was integral to Aztec myths and ceremonies. "Humans were burdened with a debt to their gods for their very existence," and they believed they must repay it with their blood—and sometimes with their lives.

The Aztecs
Frances F. Berdan
Reaktion Books, 2021.
232 pp.

Some excellent illustrations add to the volume's readability, yet one cannot help but notice the surprising absence of any substantial discussion of Aztec script. For that, the interested reader must turn instead to Whittaker's highly detailed study (*1*). ■

The reviewer is the author of *Writing and Script: A Very Short Introduction* (Oxford University Press, 2009).
Email: andrew@andrew-robinson.org

REFERENCES AND NOTES

1. G. Whittaker, *Deciphering Aztec Hieroglyphs* (Univ. of California Press, 2021).
2. B. Ortiz de Montellano, *Aztec Medicine, Health, and Nutrition* (Rutgers Univ. Press, 1990).

10.1126/science.abi5921

FOOD POLICY

The economist's guide to feeding the world

A new tome offers a road map for sustainable food production policies

By **Izabela Delabre**

In *The Economics of Sustainable Food,* editor Nicoletta Batini, a leading expert in designing macroeconomic strategies to deal with issues at the nexus of climate change and public health, argues that macroeconomic policy has largely overlooked food systems—a perplexing oversight, given that the agriculture-food system is the largest industry in the world and considering the substantial threats that unsustainable food systems pose to economies and people. Batini argues that we need a "Great Food Transformation" to support healthy and sustainable food systems that meet targets outlined in both the United Nations Sustainable Development Goals and the Paris Agreement.

The book brings together a collection of essays on how to make food production sustainable, highlighting important differences between proposed implementation in advanced economies and in less-advanced economies. Examples of the "sustainable farming trends" discussed include small and polyfunctional farming (in contrast to monocultural models promoted by agribusiness); urban, vertical, and controlled-environment farming; regenerative ocean farming; and alternative protein cultivation (e.g., plant-based "meat" and cultured animal tissues). The essays emphasize the benefits of these trends, with minimal discussion of the complex politics underlying their implementation. Nevertheless, they succeed in highlighting important alternatives to our current production practices.

The volume also proposes ways of "greening" food demand through a diverse range of economic policies. These include rethinking taxes, subsidies, standards and quality metrics, labeling and marketing regulations, and broader structural reforms.

The authors are sensitive to contexts. Multiple essays explicitly argue that those currently following meat-centered Western diets must move toward plant-based nutrition for the benefit of their health, to reduce global emissions, to increase land use efficiency, and to help make healthy food more affordable globally. In less-advanced economies, food systems will need to meet nutritional needs, address existing issues of access, and improve sustainability while ensuring socioeconomic development. In an essay titled "Greening food demand in less-advanced economies," Divya Mehra and colleagues go so far as to argue that low- and middle-income countries will require a substantial increase in greenhouse gas emissions and water use to achieve healthy diets, which will necessitate a faster shift to plant-forward diets in more-advanced economies to ensure equity.

The Economics of Sustainable Food: Smart Policies for Health and the Planet
Nicoletta Batini, editor
Island Press, 2021. 320 pp.

In 2015, 1.6 billion tons of food were lost or wasted, about a third of all the food that was produced. Despite growing calls to address food waste, the factors that contribute to it are highly complex and context dependent. In their essay "Eliminating food waste," Geeta Sethi and colleagues argue that "FLW [food loss and waste] realities vary so significantly across regions that they might be thought of as completely separate phenomena, affecting different types of food for different reasons—and even with different consequences." Although greater attention is being paid to addressing FLW, more investment in empirical research is needed to identify the most appropriate strategies for reducing waste according to context, commodity, and stage in the supply chain.

The volume's final essays focus on conserving land, sea, mammals, and insects to ensure food security. The contributions in this section examine the big-picture links between biodiversity, climate, and food systems, drawing on fascinating case studies from a range of geographical contexts. By highlighting global environmental issues, policy frameworks, and economic measures, the authors illuminate how policy-makers and stakeholders who focus on the food sector can play a meaningful role in supporting broader biodiversity initiatives.

A particular strength of this volume is the numerous case studies cited by the authors that demonstrate the feasibility of alternatives to unsustainable food system models, countering mainstream political economic narratives, which often bemoan the lack of a better system. This is a timely and important book that makes bold statements to which macroeconomists and development practitioners should pay attention. ■

10.1126/science.abi9163

The reviewer is at the Department of Geography, Birkbeck, University of London, London WC1E 7HX, UK. Email: i.delabre@bbk.ac.uk

Caterers discard expired ingredients in March 2020. The causes of food waste are highly context dependent.

PHOTO: YANG HUAFENG/CHINA NEWS SERVICE/GETTY IMAGES

Illegal trade and compromised habitats threaten the survival of the European eel (*Anguilla anguilla*).

PHOTO: WIL MEINDERTS/MINDEN PICTURES

Edited by **Jennifer Sills**

European eel population at risk of collapse

The European eel (*Anguilla anguilla*) population has declined by 98% since 1980 (*1*). The trade of European eels requires permits from the Convention on International Trade in Endangered Species (CITES), and importing and exporting the species have been illegal in the EU since 2010 (*2*). However, illegal export of the eels to Asia continues, and they have been found in major supermarkets across Hong Kong, Malaysia, and China (*3*, *4*). Without additional regulations, the European eel population is at risk of collapse (*5*).

Adult European eels breed in the Sargasso Sea. Their larvae travel thousands of kilometers over the course of 2 years to reach the European fresh and brackish waters (*1*), where they mature. After 5 to 20 years, they migrate back to the dense sargassum weed. This complicated metamorphic life cycle puts the population under substantial stress, which is exacerbated by global warming, biotope degradation, pollution, and heavy fishing (*6*–*8*).

In addition to a ban on importing and exporting eels, EU countries must further regulate the fishing of these eels in all their metamorphic stages. In the Baltic countries, fishing for European eels is forbidden from November to February to ensure their population growth, even though these eels are active in Baltic waters from October to November (*9*). Denmark, which harvested 180 tons of commercial eels in 2020 alone, has no restrictions on nor inventory of recreational fishing (*10*). In 2019, the total EU harvest of commercial adult and juvenile eels was more than 2000 tons and 55 tons, respectively, and another 1625 tons were lost to hydropower plants and their pumps (*11*).

To protect the European eel from further decline, the EU must implement either a complete ban or restrictive quotas on harvesting. Conserving this species will help to meet Sustainable Development Goal 14, Life below Water (*12*).

Christian Sonne[1,2]* Wan-xi Peng[2], Aage K. O. Alstrup[3], Su Shiung Lam[4,2]
[1]Aarhus University, Roskilde, Denmark. [2]Henan Agricultural University, Zhengzhou, China. [3]Aarhus University, Aarhus, Denmark. [4]Universiti Malaysia Terengganu, Terengganu, Malaysia.
*Corresponding author. Email: cs@bios.au.dk

REFERENCES AND NOTES

1. K. Aarestrup *et al.*, *Science* **325**, 1660 (2009).
2. L. Musing *et al.*, "Implementation of the CITES appendix II listing of European eel *Anguilla anguilla*" (2018); https://cites.org/sites/default/files/eng/com/AC/30/E-AC30-18-01-A1.pdf.
3. J. L. Richards *et al.*, *Sci. Adv.* **6**, eaay0317 (2020).
4. P. Galey, M. Billing, "Eel trafficking in the EU, the world's 'biggest wildlife crime,'" Phys.org (2018); https://phys.org/news/2018-11-eel-trafficking-eu-world-biggest.html.
5. E. D. Enbody *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2022620118 (2021).
6. International Council for the Exploration of the Sea (ICES), *ICES Sci. Rep.* **3**, 1 (2021).
7. W. Dekker *et al.*, *ICES J. Mar. Sci.* **73**, 5 (2016).
8. S. Pacariz *et al.*, *Ecol. Freshw. Fish.* **23**, 86 (2014).
9. The Fisheries Secretariat, "Eel migration report provides insights but also highlights data gaps" (2020); www.fishsec.org/2020/05/15/eel-migration-report-provides-insights-but-also-highlights-data-gaps/.
10. C. P. Schirmer, "The number of young eels dropped by 98 percent: Still, they continue to be fished," Danmarks Naturfredningsforening (2021); www.dn.dk/nyheder/antallet-af-unge-al-faldet-med-98-procent-alligevel-bliver-de-stadig-fisket/ [in Danish].
11. ICES, "European eel (*Anguilla anguilla*) throughout its natural range" (2020); www.ices.dk/sites/pub/Publication%20Reports/Advice/2020/2020/ele.2737.nea.pdf.
12. UN Department of Economic and Social Affairs, Sustainable Development, The 17 Goals (2015); https://sdgs.un.org/goals.

10.1126/science.abj3359

Disability in space: Aim high

In February, the European Space Agency (ESA) announced the Parastronaut Feasibility Project (*1*), a plan to make every reasonable effort to send astronauts with disability to space. This idea is a step in the right direction, but ESA is framing the decision as a way to be more inclusive rather than a way to make all astronauts safer and more effective. This short-sighted perspective may explain why the fine print of the announcement limits "acceptable" disabilities to lower-limb deficiencies. By not including a range of disabilities, ESA is discounting the potential of people with disabilities, ignoring the fact that disabilities are in great part barriers imposed by society (*2*), and missing an opportunity to better prepare all

astronauts to adapt to in-flight trauma. To live up to its own standards, provide inspiration to the next generation [including the 15% (*2*) who have a disability], and improve mission safety, the ESA should broaden its criteria for eligible disabilities.

ESA is planning to make minor adjustments to existing launch vehicles, an approach that hardly ever leads to useful design (*3*). Furthermore, adaptation to a single disability is a dead end: When a person with a different disability is selected, the whole adaptation process would have to be repeated. As a result, future spacecraft will be as inaccessible in design as the International Space Station and launch vehicles are today.

Instead, development in space should be with, not for, people with disabilities. Integrating astronauts with disabilities that they have learned to live with into spaceflight programs today may help save a non-disabled astronaut with mission-acquired trauma tomorrow. For example, during a proof-of-concept ground mission in 2017 (*4*, *5*), a blind analog astronaut with significant hand deformities helped identify design flaws in a habitat, leading to improvements that benefited his non-disabled successors (*6*).

ESA should consider disabilities that include blindness, deafness, upper leg deficiencies, upper limb deficiencies, paraplegia, multi-morbidity, and even cognitive or neurological (sensory and motor) deficiencies. All of these would help prepare for health issues that have been previously observed in astronauts [e.g., (*7*–*12*)].

Christiane Heinicke[1]*, Marcin Kaczmarzyk[2], Benjamin Tannert[3], Aleksander Wasniowski[4], Malgorzata Perycz[5], Johannes Schöning[6]

[1]Center of Applied Space Technology and Microgravity (ZARM), University of Bremen, 28359 Bremen, Germany. [2]Department of Building Engineering, Faculty of Civil and Environmental Engineering and Architecture, Rzeszow University of Technology, 35-084 Rzeszow, Poland. [3]Department of Applied Media Informatics, Hochschule Bremen–City University of Applied Science, 28199 Bremen, Germany. [4]Space is More, 52-420 Wrocław, Poland. [5]Computational Biology Lab, Department of Artificial Intelligence, Institute of Computer Science of the Polish Academy of Sciences, 01-248 Warsaw, Poland. [6]Department of Human-Computer Interaction, University of Bremen, 28359 Bremen, Germany. *Corresponding author. Email: christiane.heinicke@zarm.uni-bremen.de

REFERENCES AND NOTES

1. ESA, "Parastronaut feasibility project" (2021); www.esa.int/About_Us/Careers_at_ESA/ESA_Astronaut_Selection/Parastronaut_feasibility_project.
2. World Health Organization, "World report on disability" (2011); www.who.int/disabilities/world_report/2011/report.pdf.
3. J. Boys, *Doing Disability Differently: An Alternative Handbook on Architecture, Disability, and Designing for Everyday Life* (Routledge, 2014).
4. C. Heinicke, M. Kaczmarzyk, M. Perycz, A. Wasniowski, "Dealing with a physically disabled crew member: Lessons learned by the crew of the ICAres-1 mission" (European Planetary Science Congress, Berlin, Germany, 2018); https://meetingorganizer.copernicus.org/EPSC2018/ EPSC2018-592.pdf.
5. A. Wasniowski, M. Perycz, W. Jarosz, J. Dudzinska, "Addressing disability in space: ICARES-1 Mars analog mission" (69th International Astronautical Congress, Bremen, Germany, 2018).
6. A. Mintus, L. Orzechowski, N. Cwilichowska, "Overview of updated design and research potential," LunAres Analog Research Station (71st International Astronautical Congress, 2020).
7. M. Roy-O'Reilly, A. Mulavara, T. Williams, *npj Microgravity* **7**, 5 (2021).
8. S. L. Johnston III, K. T. Smart, J. M. Pattarini, in *Principles of Clinical Medicine for Space Flight*, M. R. Barrett, E. S. Baker, S. L. Pool, Eds. (Springer, 2019), pp. 327–353.
9. A. G. Lee *et al.*, *npj Microgravity* **6**, 7 (2020).
10. L.-F. Zhang, A. R. Hargens, *Physiol. Rev.* **98**, 59 (2018).
11. D. L Baisden *et al.*, *Aviat. Space Environ. Med.* **79**, 629 (2008).
12. F. A. Cucinotta, M. Alp, F. M. Sulzman, M. Wang, *Life Sci. Space Res.* **2**, 54 (2014).

COMPETING INTERESTS

M.P. served in 2019 as scientific director for Lunares Research Station, a habitat in which the Innovative Concepts Ares–1 mission was conducted in 2017.

10.1126/science.abj7353

Time to amend the US Clean Water Act

On 3 June, the National Oceanic and Atmospheric Administration released its forecast for the size of this summer's hypoxic zone (known as the "dead zone") in the Gulf of Mexico (*1*). The dead zone, which is largely the result of pollution from agricultural and urban runoff, recurs every summer primarily because the Clean Water Act has failed to address this kind of pollution. When the Clean Water Act was passed in 1972 (*2*), unregulated pollution from sites such as factories and sewage treatment plants—point source pollution—was causing widespread water quality degradation across the nation (*3*). Today, the leading cause of compromised water quality is rainwater or snowmelt that picks up contaminants as it flows across the land surface and carries them into bodies of water—nonpoint source pollution—as happens in the Gulf of Mexico and countless other inland and coastal waters (*4*). Because the primary cause of pollution has changed, the Clean Water Act should be updated.

In 1987, the Clean Water Act was amended to provide limited funding to address nonpoint source pollution (*5*). Outcomes have been mixed and difficult to track (*6*). To address nonpoint source pollution, the Act promotes voluntary actions by those generating the pollution. Agricultural runoff, a substantial source of nonpoint source pollution, is largely exempted from the Act's provisions (*7*). Given that climate change is expected to exacerbate water quality problems that are driven by nonpoint source pollution, particularly nutrient runoff (*8*), the strategies and resources put in place by the 1987 amendment are insufficient.

The Clean Water Act's weak approach to nonpoint source pollution hinders its ability to improve water quality and protect aquatic ecosystems. To address agricultural runoff and other types of nonpoint source pollution in the face of a changing climate, Congress should amend the Act to provide regulatory tools and innovative market-like policies [e.g., (*9*)] to complement the voluntary approaches that alone have proved inadequate. An amendment would also provide the opportunity to address threats to water quality such as pharmaceutical compounds, nanomaterials, and other micropollutants (*10*) that were not recognized 35 years ago when the Act was last updated.

Todd V. Royer
O'Neill School of Public and Environmental Affairs, Indiana University, Bloomington, IN 47405, USA. Email: troyer@indiana.edu

REFERENCES AND NOTES

1. National Oceanic and Atmospheric Administration, "NOAA forecasts average-sized 'dead zone' for the Gulf of Mexico" (2021); www.noaa.gov/news-release/noaa-forecasts-average-sized-dead-zone-for-gulf-of-mexico.
2. Federal Water Pollution Control Amendments, 33 U.S.C. §§1251 et seq. (1972); www.govinfo.gov/content/pkg/USCODE-2017-title33/html/USCODE-2017-title33-chap26.htm.
3. R. W. Adler *et al.*, *The Clean Water Act 20 Years Later* (Island Press, Washington, DC, 1993).
4. US Environmental Protection Agency, "Polluted runoff: Nonpoint source (NPS) pollution" (2020); www.epa.gov/nps.
5. US Environmental Protection Agency, Polluted Runoff: Nonpoint Source (NPS) Pollution, "319 Grant Program for States and Territories" (www.epa.gov/nps/319-grant-program-states-and-territories).
6. US Government Accountability Office, "Nonpoint source water pollution: Greater oversight and additional data needed for key EPA water program" (2012); www.gao.gov/products/gao-12-335.
7. J. G. Laitos, H. Ruckriegle, *Vermont Law Rev.* **37**, 1033 (2013).
8. E. Sinha *et al.*, *Science* **357**, 405 (2017).
9. P. J. Bohlen *et al.*, *Front. Ecol. Environ.* **7**, 46 (2009).
10. R. P. Schwarzenbach *et al.*, *Science* **313**, 1072 (2006).

10.1126/science.abj3438

ERRATA

Erratum for the Report "Worldwide evidence of a unimodal relationship between productivity and plant species richness," by L. H. Fraser *et al.,* *Science* **372**, eabj5097 (2021). Published online 28 May 2021; 10.1126/science.abj5097**

Your Legacy to Science

AN ESTATE GIFT TO THE
AMERICAN ASSOCIATION FOR THE ADVANCEMENT OF SCIENCE



Since 1848, our founding year, the American Association for the Advancement of Science (AAAS) has been deeply committed to advancing science, engineering and innovation around the world for the benefit of all people.

By making AAAS a beneficiary of your will, trust, retirement plan or life insurance policy, you become a member of our 1848 Society, joining Thomas Edison, Alexander Graham Bell and the many distinguished individuals whose vision led to the creation of AAAS and our world-renowned journal, *Science*, so many years ago.

Unlike many of its peers, *Science* is not for-profit. Your estate gift would provide long-term financial stability and durable annual income that will support operations and competitive innovation for years to come. **This support is vital.**

"As a teacher and instructor, I bear responsibility for the younger generations. If you have extra resources, concentrate them on organizations, like AAAS, that are doing work for all."

—Prof. Elisabeth Ervin-Blankenheim, 1848 Society member

If you intend to include AAAS in your estate plans, provide this information to your lawyer or financial adviser:

Legal Name: American Association for the Advancement of Science
Federal Tax ID Number: 53-0196568
Address: 1200 New York Avenue, NW, Washington, DC 20005

If you would like more information on making an estate gift to AAAS, cut out and return the form below or send an email to philanthropy@aaas.org. Additional details are also available online at **www.aaas.org/1848Society**.

1848 SOCIETY
AAAS

cut here ✂

Yes, I would like more information about joining the AAAS 1848 Society.

PLEASE CONTACT ME AT:

Name: _____

Address: _____

City: _____ State: _____ Zip code: _____ Country: _____

Email: _____ Phone: _____

RETURN THIS FORM TO:
AAAS Office of Philanthropy and Strategic Partnerships • 1200 New York Avenue, NW • Washington, DC 20005 USA

OUT OF HARM'S WAY

By **Brad Wible**

Climate-induced relocation. Managed retreat. Facilitated migration. Climate-related displacement. These terms describe aspects of what may be beyond words to millions of people whose homes, communities, and cultures are, or could soon be, profoundly threatened by climate change. These people must decide whether, where, when, and how to move, and to thrive, out of harm's way. Communities facing an influx of uprooted people are no less challenged. Although no single term has been legally defined and universally accepted to describe people on the move as a result of environmental drivers, the global community has made commitments to support them.

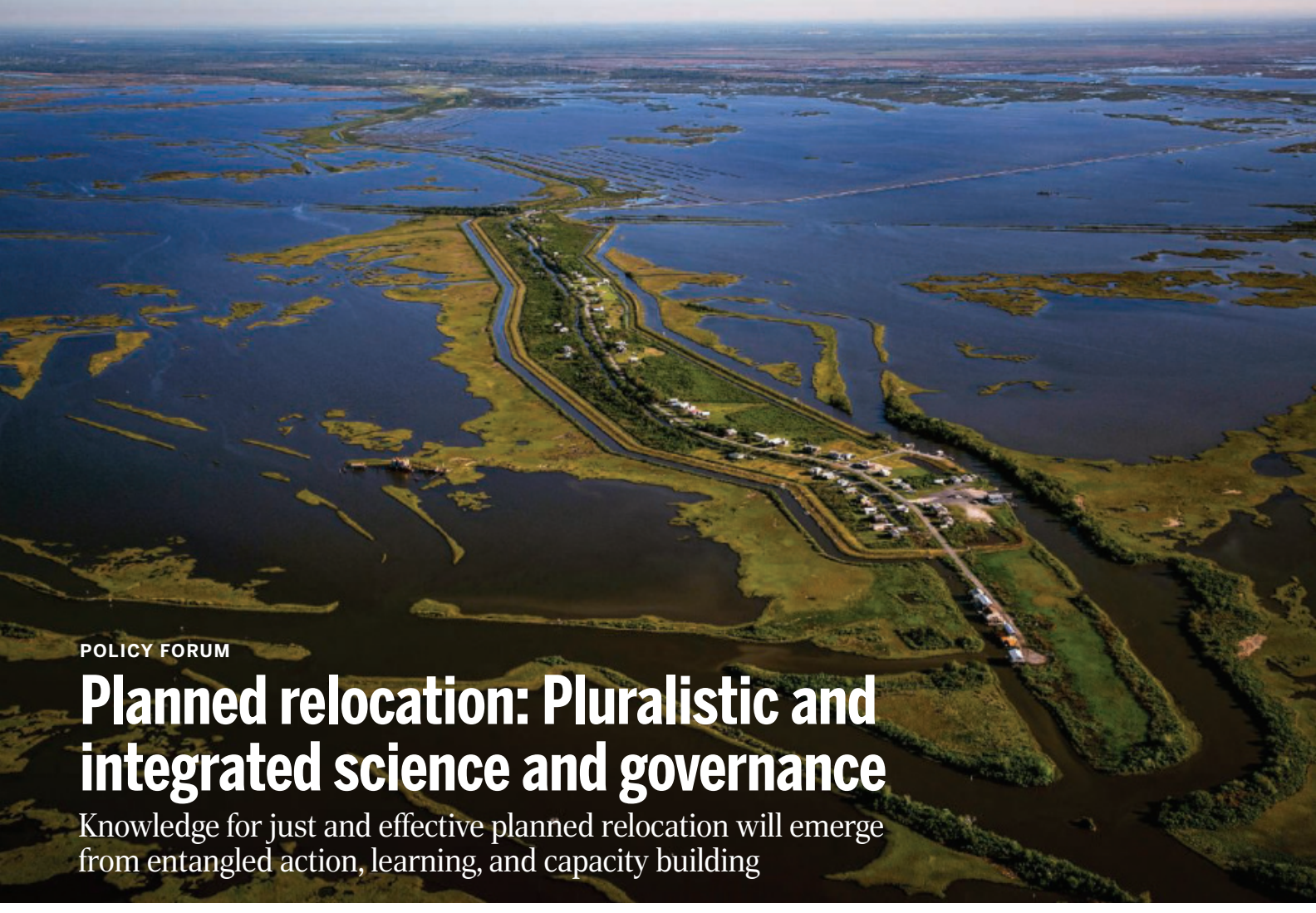
This special issue examines how research can engage with and support communities and governments navigating this uncertain landscape. Researchers are interweaving science and governance for community decision-making, improving integration of global and local analyses of human habitability, estimating the largely unmeasured costs of human displacement, mapping policy pathways toward retreat from rising seas, preparing destination regions to be migrant friendly, incorporating design and decision support into retreat planning, and more.

The direction, quality, legitimacy, and impact of such research may depend not only on the creativity of its formulation and the rigor of its execution but also on its assumptions about who makes decisions. This is because many groups whose lives will be disproportionately disrupted by climate change have also historically experienced hardships under decisions imposed upon their communities by outsiders. Thus, we must consider not only what science can do but also how science is done, and by whom.

PHOTO: ANDREW BURTON/GETTY IMAGES

INSIDE

Faced with permafrost melt and erosion, and after many years of discussion and planning, ethnically Yup'ik people of Newtok, Alaska, have begun to relocate. The process of settling in a new village, Mertarvik, is underway. Final judgment of the process and outcomes lies with Newtok's citizens.



POLICY FORUM

Planned relocation: Pluralistic and integrated science and governance

Knowledge for just and effective planned relocation will emerge from entangled action, learning, and capacity building

By **R. H. Moss[1,2], P. M. Reed[3], A. Hadjimichael[3], J. Rozenberg[4]**

Although relocation of human populations is nothing new, global environmental changes such as climate change, sea level rise, and land use change are increasing the likelihood of relocation for potentially millions of people, especially in coastal regions. Globally, sea level rise alone could place 340 million people on land projected to be below annual flood levels by 2050 (*1*). The need for relocation will increase because of such risks, the lack of funding for protection and accommodation strategies, and/or the reality that sea walls and other measures will eventually be ineffective. Thus, current approaches to planned relocation such as buyouts for individual households are likely to be "woefully inadequate" in the future (*2*). We discuss how science, governance, and their interactions need to evolve to make planned relocation a strategic option that leaves people, communities, and the environment better off. The starting point is to acknowledge that relocation involves a physical transition away from locations exposed to global change hazards, as well as the need for transformation of institutions, social networks, cultural associations, economic relationships, and other aspects of a community's way of life.

Given that relocation is a life-altering change, organizations such as the United Nations (UN) High Commission on Refugees mandate that it needs to be planned and implemented with meaningful engagement of affected parties and carried out to improve (or at least maintain) their quality of life. To ensure responsiveness to changing conditions and preferences, relocation should be part of a flexible, nested, and interconnected set of adaptation strategies that also include coping (reactive, short-term risk-reduction measures) and incremental adjustments (measures to increase resistance and/or resilience) (*3*). How to combine these different measures into a strategic portfolio of policies and actions places demands on science and governance to support open-ended adaptive planning processes that manage trade-offs across interests, uncertainties in knowledge, and institutional ambiguity created by overlapping jurisdictions, authorities, and expertise.

Planned relocation is a complex social dilemma that involves many structural, perceptual, economic, and interpersonal dynamics that discourage collective action. It will involve resolving fraught questions such as what decision processes are used, who relocates (and when), how are they compensated, where will they move, what assistance is provided (and to whom) in receiving communities, how abandoned wastes and environmental legacies are remediated, and how agreements are monitored and enforced. There is no single best approach to move a community—stakeholders with conflicting objectives will see it differently even when they share basic world views. The interaction of social and environmental triggers and lack of a preferred pathway make planned retreat similar to other global change dilemmas. But the potential scope, existential character of needed transformations, and complexity of governance challenges make it especially demanding.

[1]Joint Global Change Research Institute, Pacific Northwest National Laboratory, College Park MD, USA. [2]Andlinger Center for Energy and Environment, Princeton University, Princeton NJ, USA. [3]School of Civil and Environmental Engineering, Cornell University, Ithaca, NY, USA. [4]World Bank, Washington, DC, USA. Email: rmoss@princeton.edu

PHOTO: JOSH HANER/THE NEW YORK TIMES/REDUX

A $48 million grant from the US Federal government in 2016 is aimed to assist relocation of the residents of Isle de Jean Charles, Louisiana, including members of the Biloxi-Chitimacha-Choctaw tribe.

GOVERNANCE PARTNERSHIPS

Despite the immensity of the challenge, it is vital now to constructively engage science and governance to plan physical transitions and socioeconomic transformations that reduce risk and make people, communities, and the environment better off. Here, we offer several ideas for improving governance partnerships in developing strategies for planned relocation.

Eliminate perverse incentives and establish inclusive governance

Existing institutions and processes of governance will be stretched to address the challenges of planning and implementing relocation in a way that meets basic humanitarian principles and good practices. This is because current mixes of policies, institutions, and relationships are responsible for producing the prevailing distribution of privilege and vulnerability in society. Although climate change plays a role, it amplifies present challenges that are an amalgam of past governance, entrenched inequities, and norms. The sheer potential scale of relocation globally is beyond anything our modern global society has experienced. For example, the megacity Jakarta is actively considering relocation because of growing climate hazards, aquifer subsidence, and the density of a highly vulnerable low-income population. These challenges are not limited to the developing world, as evidenced by the mounting annual damages and recovery costs of climate extremes on populations in the United States.

Improving governance will require addressing structural inequalities and many perverse incentives and behavioral dynamics that continue to drive people to settle in areas exposed to hazards. Innovations are needed to address organizational silos, poor planning and risk communication, psychological attachments to place, and dependence on continued occupation for tax revenues. These challenges can be exacerbated with well-intentioned coping strategies (e.g., the "levee effect" that reduces accurate perception of risk). In the United States, for example, federal programs including subsidization of beach nourishment, the National Flood Insurance Program, and the federalization of natural disaster recovery encourage settlement of risky areas. Planned relocation toolkits (*4*) are beginning to emerge that orient the challenge within domestic legal frameworks and international organizations (e.g., the UN Office for Disaster Risk Reduction) and the experiences garnered from existing national efforts (e.g., Fiji's efforts to move 46 villages).

Making and implementing decisions in which communities voluntarily relocate will require inclusive, deliberative processes that emphasize transparency, engagement, trust building, accountability, and an interactive approach for engaging with science. Policy or legislative frameworks are critical to defining long-term targets and providing credible commitments to maintain the continuity of objectives across institutions and political mandates (*5*). Strategies will need to accommodate changing circumstances (new scientific evidence, technological change, new preferences) and the management of implementation tactics based on expert advice, monitoring and reporting, and accountability. In most countries, new institutions and funding are required to improve access to expert advice, coordination, and consultation. Governance frameworks for relocation will need to include periodic communication about future risks, engagement with private sector and civil society, and oversight mechanisms to monitor and enforce the implementation of agreed plans.

ACTION-ORIENTED KNOWLEDGE

Diverse perspectives in problem framing

Defining the problem and its context is the central challenge posed by planned relocation. Framing a problem establishes what is prioritized (and what is treated as unimportant), what the objectives are, and what questions will be asked and answered. Framing is often contested, and to avoid marginalizing communities, it needs to incorporate diverse perspectives, start from the specific local context of ongoing systemic challenges, enhance stakeholders' agency, and bring together diverse sources of knowledge (*6*, *7*).

It is particularly challenging to carefully analyze the diverse stakeholders and the types of knowledge that are pivotal to understanding and framing planned relocation (e.g., capturing perspectives from the relocating, receiving, and remaining populations). Problem framing could consider the need for expertise, tactical engagement, and sustained advocacy to catalyze plans into transformative actions (*6*, *8*). In addition, emerging innovations in computational social science and "coproduction" of research (in which stakeholder communities are involved in different aspects of the scientific process) offer opportunities for formalizing stakeholder analysis. Analyses could improve stakeholder identification, categorization, and relationship (power) mapping.

Account for power dynamics

Decades of research in planning, public administration, sustainability science, and science and technology studies have examined how to improve the relevance and effectiveness of science to inform planning and policy for a wide range of social, environmental, and sustainability challenges. Several prominent strands of this work focus on coproduction as being more than a means to produce science, providing a mechanism to generate public goods, services, and institutions (*7*). Accordingly, the design of coproduction processes is not just about how the interactions of policymakers, stakeholders, and scientists affect the usability of science. It is also about the process of social change—how epistemologies, social and cultural norms, institutions and policies, and power relationships among communities and stakeholders interact to determine who is involved in the process, which types of knowledge are seen as legitimate, what is produced, and what outcomes result.

For challenges as fraught as planned relocation, this more expansive approach provides a foundation for codeveloping knowledge and action. It requires engaging multiple perspectives on values and knowledge where the actors involved in coproduction of planned retreat must work together to explore normative and political differences inherent in their different visions of the future (*6*).

A critique of coproduction processes is that they can depoliticize discourse by using scientific arguments to evoke universalized ideas of what is "best." They can be structured as if all participants have an equal role when in fact governments, large nongovernmental organizations, and economic interests have disproportionate power and greater opportunities for participation (*7*). This is not just a process issue but can also affect the outcomes of coproduction—for example, favoring the use of narrow cost-benefit framings that conclude that protective measures such as beach nourishment or construction of sea walls are economically justified only for high-value assets.

Empirically informed awareness of the diverse roles and dimensions of power in coproduction and social change offers an avenue for rebalancing problematic relationships that lead to inequality or exclusion, or at least avoiding their unintended consequences (*7*). Modest steps such as providing funding to enable underserved communities to participate in coproduction, or formalizing the participation of Indigenous advisory councils, can also help level the playing field (*9*).

Diversify knowledge sources and types

To support planned relocation, science needs to deliver not just technical solutions but also

knowledge of how to relocate and transform communities, including the willingness and capacities of different groups and institutions to support fundamental change over time (*6*). Providing this knowledge will require a transdisciplinary approach to research that broadens the array of scientific disciplines and other sources of knowledge engaged. Government bodies and stakeholders (e.g., real estate interests, businesses, community-based organizations) will need to be integrated into research not just as "users" but as knowledge holders and experts in community needs, preferences, norms, and evolving capacity to implement solutions. When relocation involves Indigenous communities, rather than integrating traditional knowledge into Western science, scientists involved in coproduction arrangements should foster mutual respect on the multiple ways of knowing, by engaging in tribal avenues, such as regional newsletters and talking circles at tribal meetings (*9*, *10*).

Informing social and economic transformation will require research into the capacities and values of different populations and institutions. This requires understanding issues such as what will motivate people to make changes, the capacity of individuals and institutions to act on their preferences, and how current conditions and path dependencies affect the viability of future options (*6*). It will be necessary to "think critically about outcomes as well as processes, about institutional and process designs, [and] about power and performance" (*11*).

Amiya Brunet sits outside her home in Isle de Jean Charles, Louisiana, 7 April 2016.

PHOTO: JOSH HANER/THE NEW YORK TIMES/REDUX

Sample from a range of plausible futures to evaluate decision options

Science can better inform action if it stops trying to predict inherently unpredictable phenomena. Currently, many decision-makers frame their questions to scientists as "what will happen," and scientists respond with "projections" (possibilities based on assumptions about future radiative forcing), which are often used as predictions. This framing, in addition to putting science in the dangerous position of speculating, is not necessarily as helpful to decision-makers as "what if" questions about the consequences of options under many plausible futures. Science can be more useful by changing the objective of collaboration from "predict then act" to the exploration of hypothetical questions about what short-term actions would be consistent with long-term objectives and perform well under a diverse range of plausible futures (*12*).

As a specific example, the State of Louisiana has been confronting sea level rise, land subsidence, accelerating losses of coastal lands, and increasing risks from storm surge. The state has initiated an innovative and collaborative planning process that budgets $50 billion in a portfolio of projects to be adaptively implemented over the next 50 years (*13*). Unlike traditional cost-benefit–driven risk planning efforts based on a specific expected future ("what will happen"), the Louisiana master plan has engaged broad stakeholder participation to map what project investments hold immediate benefits while providing flexibility to confront a broad range of plausible future scenarios that could reshape their investment priorities as well as future stakeholder needs ("what if" planning).

This approach recognizes that many types of uncertainty will impede judgment and decision-making (*12*). The natural stressors that can trigger the need for evacuation are uncertain because they are emergent, compounding, and cascading outcomes of complex human–environment interactions. But the implications of changes in future values and behaviors are also uncertain and arguably just as important for evaluating decision options. Even in well-documented historical instances of relocation, it is difficult to understand how outcomes emerged from the actions taken—let alone anticipate with any certainty how desired outcomes arise from future actions (*14*).

One important opportunity is to more widely apply decision-making under deep uncertainty (DMDU) methods (*12*). These exploratory approaches draw on local-scale stakeholders' knowledge of the key factors and dynamics (human and natural) and provide a promising mechanism for informing planned relocation. Models and scenarios serve as focal points to build shared understanding about the potential implications of the different values and options preferred by stakeholders.

Social learning to build local capacity

Relocation is a complex process that will benefit from expanding the range of intermediaries and services available to facilitate production and application of knowledge. Those involved will need to know not only what scientifically robust sources of information are available for the hazards they face, but also how this information should be used to assess vulnerability, revise flood maps or zoning, evaluate financial risks to reset insurance rates and bond ratings, redesign infrastructure systems, update capital improvement and other plans, or establish thresholds and monitoring systems to trigger the next phase of agreed measures. Much attention has focused on providing climate scenarios and data, but to meet the needs of relocation, the range of services must be expanded. Needed services include not only identifying good practices in engineering, financial risk, and other technical analyses but also supporting transformation, capacity building, and

establishment of standards for different types of deliberative and analytic processes.

Research, case studies, and pilot projects are testing approaches to meet these challenges, and a useful next step is to organize evaluation and social learning to establish good practices and technical guidance. One option is to incorporate evaluation into assessments such as the Intergovernmental Panel on Climate Change and the US National Climate Assessment to establish a knowledge foundation for climate services. This would create standards for services delivered through international organizations, the private sector, academia, and public agencies (to ensure availability of services for underserved, low-income communities) (*15*). Another is an open-source wiki for climate solutions that would enable a more diverse range of knowledge holders to interact and curate guidance on good practices on an ongoing basis, emphasizing sources of credible information.

Another opportunity is to expand the use of intermediaries—individuals and institutions that facilitate interactions between stakeholders and experts (*8*). Many intermediary skillsets are necessary for the different stages of deliberative planning, financing, tactical implementation, and ex-post monitoring of relocation actions. Given the potential for contested needs and values, it is important that intermediaries be aware of how they can unintentionally affect power relationships or outcomes—for example, by using types of knowledge, analysis metrics, or visualizations that favor the perspectives of one group or another. A "critical pragmatic approach" highlights the importance of this awareness and of designing and critically evaluating deliberative processes where conflicts between parties are not reduced to simple consensus-driven debates (*11*). A variety of measures are needed to increase the number and efficacy of intermediaries, including professional certification; greater recognition, including in promotion and tenure processes; and increased funding.

Harness emerging innovations in community science and data analytics

Innovations in community science, sensing, and data analytics hold great promise in providing insights for planned relocation if privacy, equity, and other concerns such as maladaptive applications of generic algorithmic or sensing tools are addressed (*15*). Combining these innovations with monitoring investments in socioeconomic data offers the potential to better capture the interdependent evolution of human and natural systems that shape the experiences and prospects of populations facing relocation. For example, high-resolution models of flooding magnitude and extent might be available for an area, but data are missing on how inequities in agency and justice interact with exposure to hazards to shape the prospects of using planned relocation to improve people's lives.

These innovations will increase the utility of standard modes of multidisciplinary scientific research that combine hazard predictions, engineering, financial, and other analyses to inform technical solutions that contribute to physical transitions. Additional methodological advances that have not yet been fully exploited include improved projections of hazards at various spatial scales; research on coastal habitat loss and nature-based solutions; new data sources, indicator-based assessments, and demographic modeling to identify vulnerable populations; and practice standards for using global change risk analytics in engineering and other professions. This contextualized technical knowledge can provide insights for sequencing transitional risk reduction and protection measures.

REALIZING JUST RELOCATION

Revolutionizing the role of science to focus on conditions that will affect the ability of society to identify just relocation pathways, build agency, and implement strategies under uncertainty will require a "pluralistic and integrated approach to action-oriented knowledge" (*6*). Such an approach will increase confidence in the ability of communities to successfully navigate planned relocation on the massive scales at which it is likely to be required. It must build a more ethical and responsible approach that serves those affected. ■

REFERENCES AND NOTES

1. S. A. Kulp, B. H. Strauss, *Nat. Commun.* **10**, 4844 (2019).
2. J. Carey, *Proc. Natl. Acad. Sci. U.S.A.* **117**, 13182 (2020).
3. N. Chhetri, M. Stuhlmacher, A. Ishtiaque, *Environ. Res. Commun.* **1**, 015001 (2019).
4. E. Ferris, *Int. Organ. Migr.* (2017).
5. World Bank, "World Bank Reference Guide to Climate Change Framework Legislation" (Washington, DC 2020); https://openknowledge.worldbank.org/handle/10986/34972.
6. G. Caniglia *et al.*, *Nat. Sustain.* **4**, 93 (2021).
7. C. Wyborn *et al.*, *Annu. Rev. Environ. Resour.* **44**, 319 (2019).
8. P. Kivimaa, W. Boon, S. Hyysalo, L. Klerkx, *Res. Policy* **48**, 1062 (2019).
9. P. Cochran *et al.*, in *Climate Change and Indigenous Peoples in the United States: Impacts, Experiences and Actions*, J. K. Maldonado, B. Colombi, R. Pandya, Eds. (Springer, 2014; https://doi.org/10.1007/978-3-319-05266-3_5), pp. 49–59.
10. N. Latulippe, N. Klenk, *Curr. Opin. Environ. Sustain.* **42**, 7 (2020).
11. J. Forester, *Plann. Theory* **12**, 5 (2013).
12. V. A. W. J. Marchau, W. E. Walker, P. J. T. M. Bloemen, S. W. Popper, *Decision Making under Deep Uncertainty : From Theory to Practice* (Springer Nature, 2019); https://library.oapen.org/handle/20.500.12657/22900.
13. J. R. Fischbach, D. R. Johnson, D. G. Groves, *Environ. Res. Commun.* **1**, 111001 (2019).
14. K. de Koning, T. Filatova, *Environ. Res. Lett.* **15**, 034008 (2020).
15. R. H. Moss *et al.*, *Weather Clim. Soc.* **11**, 465 (2019).

10.1126/science.abh3256

POLICY FORUM

Assessing human habitability and migration

Integrate global top-down and local bottom-up analyses

By **Radley M. Horton[1], Alex de Sherbinin[2], David Wrathall[3], Michael Oppenheimer[4]**

Habitability loss is increasingly recognized as an important dimension of climate risk assessment and one with complex linkages to migration. Most habitability assessments, like climate risk assessments more generally, are based on "top-down" approaches that apply quantitative models using uniform methodologies and generalizable assumptions at global and regional scales, privileging physical sciences over social science–informed understandings of local vulnerability and adaptive capacity. Many assessments have focused on a single climate hazard threshold (such as permanent inundation or the 1-in-100-year flood), and a subset have implied that outmigration may be one of the few viable adaptation responses (*1*). There is a risk that such climate determinism minimizes the potential for human agency to find creative, locally appropriate solutions. Although top-down modeling can serve a useful purpose in identifying potential future "hot spots" for habitability decline and potential outmigration, only by integrating "bottom-up" insights related to place-based physical systems and social contexts, including potential adaptive responses, will we arrive at a more nuanced understanding. This integrated framework would encourage development of policies that identify the most feasible and actionable local adaptation options across diverse geographies and groups, rather than options that are deterministic and one-size-fits-all and encourage binary "migrate or not" deci-

[1]Lamont-Doherty Earth Observatory, Columbia University, New York, NY, USA. [2]Center for International Earth Science Information Network, Columbia University, New York, NY, USA. [3]College of Earth, Ocean, and Atmospheric Sciences, Oregon State University, Corvallis, OR, USA. [4]Department of Geosciences and School of Public and International Affairs, Princeton Univeristy, Princeton, NJ, USA. Email: rh142@columbia.edu

sions. We propose a set of recommendations centered around building the research and assessment knowledge base most needed to inform policy responses around habitability loss and migration.

DEFINING AND ASSESSING HABITABILITY

We define habitability as the environmental conditions in a particular setting that support healthy human life, productive livelihoods, and sustainable intergenerational development. Climate change may undermine one or more of the following associated, interacting, dimensions of habitability: basic human survival (*2*), livelihood security (*3*), and societies' capacity to manage environmental risks (*4*). Rapid rates of climate change and departures from historical variability ranges can increase risks, especially when coupled with nonclimate stressors. In such instances, threats to habitability may be evident in changing flows of human migration, whether forced or voluntary (*5*).

Most habitability assessments have relied on outputs from top-down models. This approach is conducive to system-level prediction, producing quantitative outputs that are globally comparable, such as single physical hazard thresholds that are either assumed or empirically based. Much recent work reflects a blend of long-term, high-resolution historical climate data where available, combined with projections across a large suite of global climate models driven by multiple representative concentration pathways (RCPs) representing trajectories of greenhouse gas concentrations. Another critical element is inclusion of extreme events, often expressed as a frequency of occurrence or a magnitude associated with a given recurrence period. In turn, top-down demographic and economic models, which form the basis for the shared socioeconomic pathways (SSPs) projecting global socioeconomic trajectories, provide a picture of future population and development that can also inform projections of people and assets at risk. Climate projections can also drive sectoral impact assessments—for example, empirically by extending historical statistical relationships between climate variability and the affected sector. More commonly, projections from standardized climate simulations drive sectoral impact models that dynamically simulate key features, such as crop growth. Top-down migration models use relative changes in sectoral impacts across regions along with other information as a means of projecting future population flows. Thus, these models project responses to habitability changes in regions where varying conditions may lead to outmigration, inmigration, or both.

The standardized nature of top-down methods facilitates comparisons—for example, of regions most at risk of crossing habitability thresholds associated with a climate hazard, and when. The top-down perspective can also reveal large-scale trends and interconnected features of global systems. However, there are several limitations. First, local and regional geophysical and sector-specific factors can drive hazards and risks at scales missed by global analyses. Second, less-modeled, place-specific characteristics of populations, such as health and socioeconomic status, shape both exposure and vulnerability. Third, adaptation choices and activities are embedded in historical context and culturally specific individual and community values and objectives that cannot easily be incorporated in models. Fourth, high-impact outcomes—associated, for example, with compound extreme events and abrupt changes in climate, ecological, and social systems—may be underestimated because of top-down model limitations such as the inability to credibly resolve evolving correlation structures across variables, space, and time, and key system sensitivities and feedbacks within and across systems (*6*). For example, climate phenomena teleconnected across great distances may lead to "breadbasket" failures in key food-producing regions and price shocks that can seriously reduce food security among vulnerable populations far away from the regions experiencing the climate stress.

Fortunately, top-down approaches are increasingly being paired with bottom-up approaches that offer a specificity that can help address these challenges. Bottom-up conceptual and/or computational modeling of complex adaptive systems can be designed to simulate the local experience of losing habitability over time. In the breadbasket case above, models of local responses can be paired with global models of international food trade that set boundary conditions. For example, agent-based models (ABMs) set up simulations with agents empirically calibrated to behaviorally respond to changing environmental conditions: the loss of assets and livelihood opportunities, threats to life, and changing structure of social networks. Modeling can be trained on local data to understand and predict important feedbacks at higher spatial and temporal resolution than is possible with global models. ABMs can be calibrated to examine a range of individual-actor preferences and test the effect of local decision-making to plausibly depict trade-offs among adaptation options, including migration (*7*). As another bottom-up example, qualitative information can be coproduced with diverse stakeholders, including subject matter experts, to explore high-impact scenarios and local solutions that will be missed by top-down approaches. Of course, bottom-up approaches have their limitations as well. For example, their specificity makes it difficult to compare across geographies and groups, and individual methodological decisions can appear arbitrary. Furthermore, bottom-up computational models such as ABMs are still limited by a lack of empirical data with which to calibrate model parameters.

CLIMATE HAZARDS AND HABITABILITY

Here, we walk through the habitability challenges of two climate hazard examples, demonstrating the strengths and limitations of top-down approaches and how bottom-up perspectives lead to different policy-relevant insights.

Sea level rise and extreme sea level events

Recent years have seen growing complexity and nuance in assessments. Global assessments have supplemented climate model outputs by considering a broad range of sea level change components and including, for example, expert elicitation as a means of estimating low-probability, high-consequence outcomes (*8*). High-spatial-resolution digital elevation models and consideration of changes in the frequency and intensity of societally relevant metrics such as recurrence intervals and extreme values of coastal high water have been integrated into global products. Using many of the above advances, Kulp and Strauss estimated that the number of people exposed annually to coastal flooding under constant population could increase from 250 million people today to, by 2100, 310 million to 420 million under an intermediate scenario to 380 million to 630 million under a high-end scenario (*1*). Other studies have included changes in storms, hyper-local positive correlations between population density and subsidence, population projections consistent with SSP-RCP combinations, and assets at risk.

Additional refinements have focused on specific coastal locations, adding critical context at the expense of global information. For example, Storlazzi *et al*. framed their assessment of tipping-point risks to atolls around two metrics—annual overwash events that threaten infrastructure, and salinization of groundwater—that are specifically relevant for atolls given their small size, uniformly low elevation, and relative isolation and found that habitability is threatened in most atoll islands by the middle of the 21st century, far sooner than permanent-inundation–based studies would suggest (*9*). Some local studies have included dynamic interaction between coastal waters and adjacent landforms. Other local and regional studies have considered social dimensions of human vulnerability, as well as in situ adaptation, using empirically calibrated agent-based liveli-

hood decision models that span multiple climate, RCP, and SSP scenarios (*7*).

The three dimensions of habitability demonstrate why no single coastal flood metric threshold can be determined in a top-down way. For the direct survivability dimension, key factors include future flood control, feasibility of evacuation, and the stochasticity of individual storms. For livelihood, saline intrusion, for example, could benefit some sectors such as specialized aquaculture, even as it harms most sectors and people. And for the societal resilience dimension, large-scale factors such as levels of inequity, strength of governance and social networks, and quality of infrastructure will be critical. As sea levels rise and coastal flooding becomes more common, social, economic, and political factors in some locations will conspire to induce sudden loss of habitability far sooner than physical hazard–based thresholds such as permanent inundation would suggest, as risk perception and long-term economic viability shift. For example, increases in insurance premiums could negatively affect asset values and tax revenues, leading to deteriorating infrastructure and services. The timing of such threshold-crossing cannot be predicted on the basis of top-down models alone. In some instances, shocks can lead to rapid learning, adjustment, and in situ adaptation, at least temporarily.

Extreme heat

Most assessments of future heat hazards have considered temperature only, although recent efforts are increasingly adopting a compound events framework—for example, considering how co-occurring extremes of high temperature and high humidity can modulate threats to habitability. Humid heat is particularly harmful to human health and the ability to engage in outdoor activities. Sherwood and Huber described a wet bulb temperature of 35°C as a threshold above which humans could not survive beyond approximately 6 hours owing to physiological and thermodynamic limits on the ability to cool through perspiration (*2*). Model-based studies have projected that this threshold could be crossed in the Persian Gulf and South Asia during the second half of the 21st Century under a high-emissions scenario (*10*). However, a finer-scale study found that this threshold has already been briefly crossed multiple times in populous cities.

Although an absolute habitability threshold exists for the survivability dimension of extreme humid heat, some people will lose their ability to thermoregulate at much lower wet bulb temperatures. Mortality rates of the elderly, those with chronic health conditions, and those involved in strenuous activity rise dramatically well below the 35°C wet bulb threshold. In terms of the livelihood dimension, at ~3.5°C of global warming above pre-industrial levels, de Lima *et al.* project that in Sub-Saharan Africa and Southeast Asia increases in humid heat may decrease agricultural labor productivity by 30 to 50%, leading to larger agricultural sector impacts than are associated with direct temperature and CO_2 effects on crops (*11*). However, air conditioning and other adaptations will enable—indeed, have enabled—some people to continue to live in places that exceed the 35°C threshold. Such an outcome increases inequity because those with no option but to work outdoors, or no access to affordable air conditioning, would be forced to migrate. And even for those with air conditioning, the third dimension of habitability—society's capacity to manage environmental risks—will be tested in unforeseen ways because it will be critical that air conditioning not fail.

Sea level rise and extreme humid heat are far from the only climate hazards that have been assessed in the literature for potential habitability thresholds. For example, changes in surface moisture fluxes as mean precipitation and temperature shift are projected to have large impacts on dryland agriculture, fire regimes in forests, and water availability downstream from snow and glacier reservoirs. These and other hazards and impacts may overlap and interact across scales to affect habitability in complex ways, such as by potentially increasing the risk of conflict.

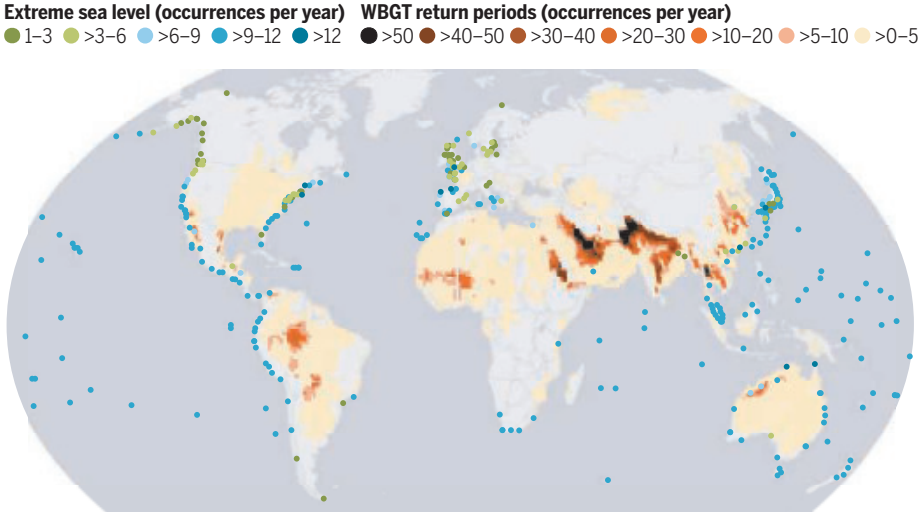
Areas where current-day rare extreme sea level and humid heat events will occur with high frequency by the end of the century under a high emissions scenario of sea level rise and warming are identified in the figure (*8*, *12*). The two metrics, corresponding to the current 1-in-100-year extreme sea level event and a wet bulb globe temperature of 33°C, respectively, are emblematic of top-down approaches. They thus represent an important point of entry for engagement with the bottom-up insights described above, as a step toward more nuanced habitability and migration assessments.

LINKING HABITABILITY TO CLIMATE MIGRATION

Migration may result from threats to survival, upended livelihoods, or the breakdown in the collective capacity to adapt (*5*). However, research on climate change and migration makes clear that an even broader set of factors undergird migration decision-making. A decision to move is ultimately a personal or household judgment on factors that include local habitability. Involuntary migration occurs when people lack agency about the key dimensions of mobility, including the timing, destination, or duration of mobility or whether to migrate at all. Where agency is extremely low, involuntary migration may take different forms, including temporary or

Frequent exceedance by 2100 of historically rare climate thresholds

Under the high-emissions scenario RCP8.5, at most coastal locations extreme sea level events historically defined as 1-in-100-year events are projected to range in frequency from once per year to more than 10 times per year due to the effects of sea level rise alone. Only point locations where historical event data are available are shown. Projected number of days per year by 2100 exceeding a 33°C wet bulb globe temperature (WBGT) in a high-emissions scenario are also depicted. Under standard assumptions of wind and solar radiation, a WBGT of 33°C corresponds to a wet bulb temperature of roughly 31.5°C. [Sea level data are from figure 4.12 in (*8*); WBGT data are from figure 3 in (*12*).]



GRAPHIC: N. DESAI/*SCIENCE* BASED ON HORTON *ET AL.*

permanent displacement and distress migration. Distress migration—mass migration or displacement related to rapid deterioration in local circumstances—is a humanitarian concern because of the need for emergency interventions to avoid poor outcomes. Distress migration has been a common phenomenon throughout history but has risen and fallen on the global policy agenda largely as a function of whether or not wealthy industrialized countries are destinations. Also of humanitarian concern is the phenomenon of involuntary immobility, in which people are unable to move without help—the population most likely to require assistance relocating under managed retreat programs. Avoiding distress migration and involuntary immobility in favor of safe and orderly migration, as advanced by the Global Compact on Migration, is now a global policy priority, and the Compact calls on governments to "strengthen joint analysis and sharing of information to better map, understand, predict, and address migration movements" as a result of climate change impacts—all of which are essential aspects of habitability assessment.

Many assessments posit some form of forced migration as an inevitable outcome of declining habitability. Yet, environmental stress rarely directly results in migration but works through a complex array of economic, demographic, social, and political proximate determinants that both initiate and sustain or modify flows. In any given population exposed to climate risks, different segments of the population respond to hazards differently and at different points in time, and as such, migration evolves with habitability through time. Whereas some may be able to migrate from deteriorating conditions without assistance, others may become immobile owing to limited options and insufficient resources, suffering progressive impoverishment and vulnerability unless social protection or planned relocation efforts are implemented (*5*). In situ adaptation, facilitated migration, and improving reception of migrants in (largely urban) destination areas are often more appropriate policies in these regions. Managed retreat has been proposed as a strategy for regions with declining habitability, but as a largely technical package of responses that includes buyouts, incentives, and planned relocation, among others, it does not currently translate well to most developing-world circumstances.

The relationship between habitability and migration may be counterintuitive, as illustrated by the lack of evidence for migration away from low-lying delta areas despite acute risks (*7*). Migration itself affects habitability for those who are unable or unwilling to leave increasingly vulnerable circumstances, either positively, such as through incoming remittances, or negatively, such as through outmigration of the working-age demographic stratum and subsequent changes in economic dynamism and livelihood options. Flows may begin owing to entrenched poverty and environmental risks and then be sustained as migrant social networks lower barriers for those who initially remained behind.

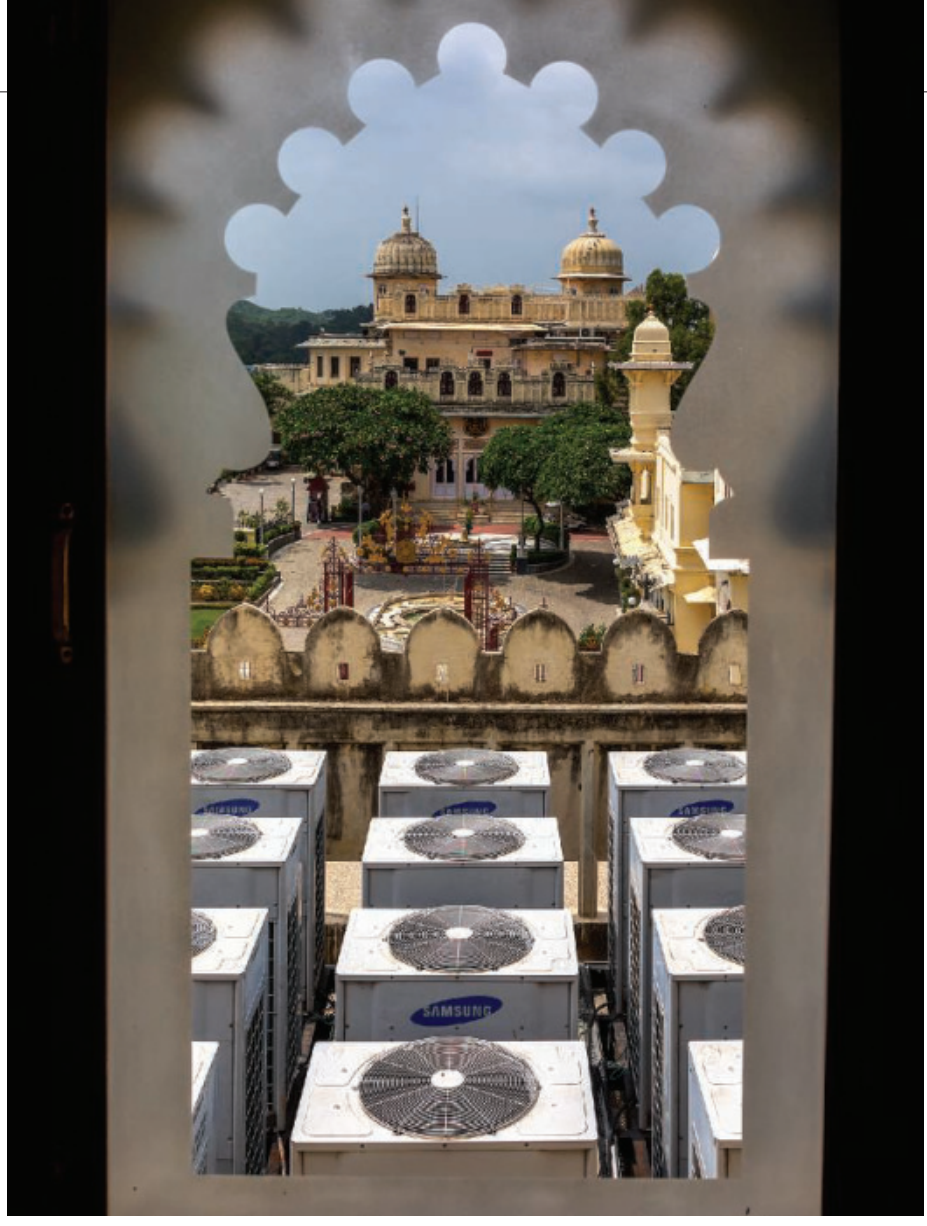
Although migration offers possibilities for advancing human well-being, as multiple dimensions of habitability are compromised, resulting forced migration will negatively affect human well-being. Migrants risk new constraints in urban informal settlements, and displaced persons may become permanently disconnected from their original communities and livelihoods in resettlement communities or refugee camps (*13*). Although top-down assessments oversimplify likely migratory responses to habitability declines, this does not necessarily imply that migration flows are overestimated. Multiple factors are driving migration in developing regions to varying degrees, including poor governance, perceived lack of opportunities, conflict, individual extreme events, and in some cases, climate-catastrophic discourses that add to a sense of hopelessness (*14*). Deeper and more contextualized understandings of migration dynamics aid in policy design, but the threats that result from declining habitability in combination with other drivers are real and may lead to substantial displacement of populations across a range of spatial scales.

A PATH FORWARD

Top-down, threshold-based habitability assessments can serve a critical role in helping to identify priority regions and groups for integrated bottom-up work while revealing interactions in global systems that cannot be gleaned from the bottom-up work alone. Integration not only leads to better predictions of when and where habitability may diminish but also can be used to inform adaptation responses that themselves help preserve or restore habitability. Bottom-up assessments by definition provide finer, local resolution, and their richness of detail means that they require diverse participation and methods. To date, most locales have not been subject to such integrated habitability assessment. We thus encourage transdisciplinary, long-term coupled top-down and bottom-up habitability assessment [for example, (*15*)] to complement and augment efforts such as the Inter-Sectoral Impact Model Intercomparison Project (ISI-MIP), which has contributed so much to our understanding of potential future climate impacts on sectors such as agriculture, water, ecosystems, and health. Initial model intercomparison could focus on what regions and groups face diminishing habitability under different model configurations. Particularly where models agree on potential habitability hot spots, bottom-up modeling experiments could be conducted and compared on specified challenges to human survival, livelihoods, and capacity to manage risk, although standardization would be needed. The Intergovernmental Panel on Climate Change (IPCC) and national efforts can also help to develop this still inchoate middle space between top-down and bottom-up approaches to habitability and migration. Migration is emerging as a cross-cutting theme throughout the current IPCC assessment, and a special report on habitability and migration would both advance the knowledge base and showcase emerging methodologies. As one example, a climate change detection and attribution dimension would help inform dialogues about loss and damage under the Paris Agreement. Likewise, a discussion on migration across the Reasons for Concern commonly used in IPCC assessments (*5*) would allow us to distinguish how climate-induced migration, distress or otherwise, is distinct from other forms of migration.

The complexity of the assessment challenge calls for a holistic, people-centric approach in which models, data aggregation, and ethnographic work are all advanced. Sectors such as engineering, hydrology, and reinsurance, that have historically been overreliant on physical models and hazard thresholds, operate at a scale that is ripe for habitability-relevant innovations at the interface between top down and bottom up. In this middle space, models could be used to examine policy scenarios instead of learning occurring exclusively from costly, time-consuming, real-world policy interventions that may put vulnerable people at risk. Greater communication among modelers will be key, and models must be validated with on-the-ground local research. To support migration and habitability modeling specifically, this would include data on when, where, and why people have moved or considered moving, how they define habitability, and the policy conditions that determine mobility outcomes (*14*). Furthermore, bottom-up research must account for the place-specific characteristics of populations—such as assets, livelihood opportunities, and social networks—that shape both exposure and adaptation. Investments in place-based social science thus help address data gaps, providing ground-truthing that will strengthen simulations of the outcomes of interventions. Investments in early-warning systems could help to anticipate where distress migration may happen, a key step in informing policy.

The shortcomings of adaptation planning and policy at current risk levels in wealthy countries hint at the global challenges ahead in a changing climate. In the United States,



Air conditioning, such as in Rajasthan, Udaipur, India, will enable some people to continue to live in places that exceed the 35°C threshold. Those with no access to affordable air conditioning could have little choice but to migrate.

PHOTO: ERIC LAFFORGUE/ART IN ALL OF US/CORBIS/GETTY IMAGES

for example, federal and local risk assessments—let alone policies—are not presently centrally coordinated or comparable. There is woefully insufficient funding available for bottom-up adaptation efforts from the better-financed federal level. Policies toward population mobility—whether planned, internal responses or immigration from other countries—vary from inconsistent over time to incoherent and sometimes inhumane. Coproduction of knowledge across diverse groups will be a precondition for any breakthroughs. In some instances, a starting point may be to bring preexisting top-down habitability and migration assessments to communities, provided that community feedback is collected and integrated iteratively and before key policy decisions are made. In other instances, stakeholder engagement may begin with fewer top-down, nonprobabilistic approaches that can be developed with communities, such as storylines and scenarios. Storylines and scenarios lend themselves to exploration of the uncertainties that most influence habitability locally (for example, the potential for changing correlation structures in models) and which adaptation strategies should be explored for which groups. Deeper stakeholder engagement, coupled with the other recommendations above, thus provides a foundation for colearning, iteration, and developing flexible approaches to the challenge of diminishing habitability.

To the extent that top-down, threshold-based approaches are used to define habitability universally, there is a risk of assuming a high likelihood of uniform outmigration or concluding with blanket policy recommendations around managed retreat. Basing assessments on nuanced definitions of habitability and integrating top-down with bottom-up approaches could encourage a broader range of policies tailored to specific locations and groups, including regions that have been put forth as likely receiving areas. A focus on the dimensions of habitability presented here, and bottom-up approaches, will invariably alter top-down projections of migration. Under wetbulb temperatures exceeding 35°C, high levels of outmigration from the Persian Gulf may be avoided if air conditioning is widely available and alternative livelihood options develop for those who would otherwise work outdoors. However, there will be regions where social tipping points and a sense of prevailing pessimism about the future—for example, owing to evolving risk perception or disinvestment by the private or public sectors—could contribute to outmigration far sooner and more suddenly than top-down habitability threshold–based methods would suggest. Global, regional, and national migration policies themselves will also play an important role in facilitating or impeding migration.

What is already clear is that climate change will result in shifting population distributions and that this process will overall be harmful to the most vulnerable, including those who may be "trapped" in deteriorating circumstances. For the reasons described here, and as a matter of climate justice, many semi-arid regions, much of the tropics, and some low-lying deltas and islands should be high priorities for integrated transdisciplinary work on habitability risks and major investments in adaptation. But only by taking into account the complexities described here will we avoid climate determinism and instead implement proactive policies on adaptation and migration that in particular will address the needs of the most vulnerable. ■

REFERENCES AND NOTES

1. S. A. Kulp, B. H. Strauss, *Nat. Commun.* **10**, 4844 (2019).
2. S. C. Sherwood, M. Huber, *Proc. Natl. Acad. Sci. U.S.A.* **107**, 9552 (2010).
3. T. Tanner *et al.*, *Nat. Clim. Chang.* **5**, 23 (2015).
4. J. Barnett, W. N. Adger, *Annu. Rev. Environ. Resour.* **43**, 245 (2018).
5. R. McLeman *et al.*, *Clim. Change* **165**, 24 (2021).
6. N. Simpson *et al.*, *One Earth* **4**, 489 (2021).
7. A. R. Bell *et al.*, *Environ. Res. Lett.* **16**, 024045 (2021).
8. M. Oppenheimer *et al.*, in IPCC Special Report on the Ocean and Cryosphere in a Changing Climate, H.-O. Pörtner *et al.*, Eds. (IPCC, 2019).
9. C. D. Storlazzi *et al.*, *Sci. Adv.* **4**, eaap9741 (2018).
10. C. Raymond, T. Matthews, R. M. Horton, *Sci. Adv.* **6**, eaaw1838 (2020).
11. C. Z. de Lima *et al.*, *Environ. Res. Lett.* **16**, 044020 (2021).
12. D. Li, J. Yuan, R. E. E. Kopp, *Environ. Res. Lett.* **15**, 064003 (2020).
13. A. Heslin *et al.*, in *Loss and Damage from Climate Change* (Springer, 2019), pp. 237–258.
14. H. Adams, S. Kay, *Environ. Sci. Policy* **93**, 129 (2019).
15. K. Grace, S. Siddiqui, B. F. Zaitchik, *Nat. Food* **2**, 1 (2021).

ACKNOWLEDGMENTS

The authors thank four anonymous reviewers and C. Lesk for comments and K. MacManus for assistance with the map figure. R.M.H. and A.d.S. were supported by the Columbia Climate School and its Earth Institute, and A.d.S. received funding from NSF award 1934978.

10.1126/science.abi8603

POLICY FORUM

Addressing the human cost in a changing climate

Displacement costs remain largely invisible, hindering effective action

By **Bina Desai[1], David N. Bresch[2], Christelle Cazabat[1], Stefan Hochrainer-Stigler[3], Reinhard Mechler[3], Sylvain Ponserre[1], Jacob Schewe[4]**

Climate change is leading to systemic and existential impacts, and evidence is mounting that these can result in the displacement of human populations. There is a rapidly growing demand for comprehensive risk assessments that include displacement and its associated costs to inform humanitarian response and national planning and coordination. However, owing to complex causation, missing and incomplete data, and the political nature of the issue, the longer-term economic impacts of disaster- and climate-related displacement remain largely hidden. Current approaches are rarely ex ante and prospective and do not consider systemic risk management. Not surprisingly, response-based approaches have shown mixed results, repeatedly demanding substantial resources while not addressing the root causes of displacement.

Climate change is not only affecting the intensity, frequency, and duration of hazards that trigger displacement but is also eroding already fragile livelihoods and ecosystems, acting as an aggravator of existing vulnerability and contributing to chronic poverty and conflict in affected countries (*1*). Although disaster risk reduction as a cross-sectoral issue has gained considerable attention over the past two decades, disaster displacement risk is still not fully integrated in national policies and planning. Out of 46 countries included in the 2020 Internal Displacement Index, most acknowledge disaster displacement in principle and have climate policies or national adaptation plans in place. However, only 27 recognize the link between the gradual impacts of climate change and displacement (*2*).

With an evidence-based, longer-term vision and investments, climate-related displacement—the forced movement of people in response to a hazard—can be averted and replaced by a range of measures such as planned relocation that is voluntary (at least to a large degree) and financially supported, or by building the resilience of at-risk populations, reducing vulnerability to such an extent that moving is not required. What is missing is a risk-informed framework for country-led, forward-looking approaches to make the case for substantial investment in effective risk reduction, durable solutions for those displaced, and the prevention of new displacement.

Applied risk science, using probabilistic models and large empirical datasets compiled over the years, combined with insights from local empirical research and community assessments, now offers the opportunity for a step change in informed decision-making. For example, the shift from deterministic disaster risk assessments, based on historical data, to state-of-the-art probabilistic modeling used by the insurance industry, calibrated with historical data but including randomness to encompass all possible scenarios, presents a notable advance in risk science that is yet to be fully applied to displacement risk. New tools and risk modeling platforms, such as CLIMADA run by ETH Zürich or CAPRA of the World Bank, can now be adapted for displacement risk assessments. Further, assessing the social and economic cost of displacement can provide incentives for transformational action and change, from mere response to disaster displacement to proactively addressing vulnerability and exposure, thereby reducing displacement risk.

"Applied risk science… now offers the opportunity for a step change in informed decision-making."

DISASTER DISPLACEMENT AND CLIMATE CHANGE

Disaster displacement is a global reality and everyday occurrence. Millions of disaster displacements have been systematically recorded since 2008—on average, 24.5 million new movements every year (*3*). Weather-related hazards account for almost 90% of all these displacements (*2*), with climate change and the increasing concentration of populations in areas exposed to storms and floods, coupled with socioeconomic drivers of vulnerability, meaning that more people are at risk of being displaced. Demographic, historical, political, and socioeconomic factors determine whether people can withstand the impacts of a physical hazard or environmental stressor or have to leave their homes. Climate change interacts with all of these factors, particularly where resources and the capacities of humans and systems are already stretched (*4*). For example, sea level rise results in loss of land in coastal areas and low-lying atolls of island states, forcing communities to retreat or leave the land altogether. Salinization can reduce crop yields, undermine arable land and freshwater availability, and force people to move. Increasing temperatures affect soil moisture and degradation, which make the soil susceptible to nutrient loss and erosion, thereby destroying the livelihood basis for rural communities. Glacial retreat and melt, loss of biodiversity, and land and forest degradation mean decreased ecosystem services and provisioning services, pushing people to move. Because climate change can also alter the intensity, frequency, and duration of hazard events, climate anomalies and more devastating sudden-onset disasters may follow.

Most of the impacts of climate change only result in displacement for those vulnerable to them. This essential point is repeatedly forgotten, with important policy implications (*5*). A prosperous farmer with access to drip irrigation and fertilizers, reliable buyers, loans, and insurance will not be as affected by changes in rainfall patterns as a smallholder subsistence farmer relying on the regularity of seasonal rains or a pastoralist in search of pasture for his herd. An urban dweller with an office job and regular income will not need to leave his home because of the loss of mangroves, which are providing sustenance to millions in coastal communities. Nonetheless, although individual vulnerability leads to a risk of adverse displacement outcomes, disaster and climate risks are increasingly becoming systemic because high-level and widespread impacts may ripple through social and economic networks, incurring

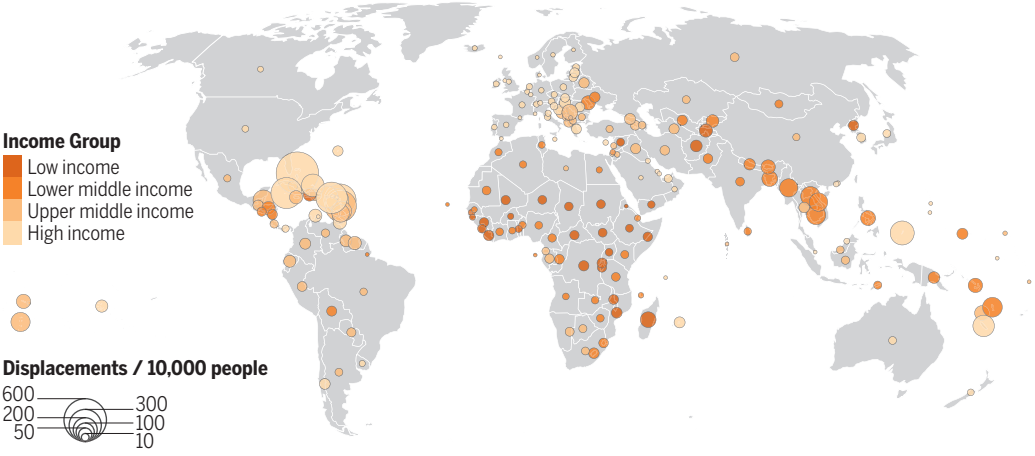
[1]Internal Displacement Monitoring Centre (IDMC), Geneva, Switzerland. [2]Eidgenössische Technische Hochschule (ETH) Zürich, Zürich, Switzerland. [3]International Institute for Applied Systems Analysis (IIASA), Laxenburg, Austria. [4]Potsdam Institute for Climate Impact Research (PIK), Potsdam, Germany. Email: bina.desai@idmc.ch

further adverse micro and macro impacts and disruptions (*6*).

Climate change is thus a displacement trigger in its own right (e.g., loss of coastlines through sea level rise and coastal erosion), a visible aggravator (e.g., when livelihoods are eroded because of soil degradation and loss of ecosystem services), and a hidden aggravator (e.g., increasing the intensity of cyclonic winds and shifting rainfall patterns that result in floods). But the impacts of climate change interact with broader changes in the physical and social environment, resulting in potentially rising costs associated with future displacement.

Global disaster displacement risk relative to population size

Average Annual Displacement (AAD) risk is a compact metric that represents the estimated effect, accumulated over a long time frame, of future small to medium and extreme events and estimates the likely displacement associated with them on a yearly basis for sudden-onset hazards such as tsunamis, cyclonic winds, storm surges, and riverine floods. See (*10*) for details. Each country's AAD risk relative to its population size is shown (expected annual displacements / 10,000 people). Country income group classification from the World Bank.



GRAPHIC: N. DESAI/*SCIENCE* BASED ON B. DESAI *ET AL.*

A HEAVY BURDEN…ON NOBODY'S BALANCE SHEET

Disaster displacement often undermines the welfare and well-being of affected individuals and communities and can also incur a substantial social and economic burden on countries. Although many countries have begun to plan for the risk of extreme events in one way or another, governments typically do not formally account for displacement risk and their associated costs in national development plans and annual budgets of line ministries. Even without taking into account the aggravating forces of climate change, there is growing evidence that displacement not only severely disrupts the lives of those forced to flee their homes but also has an economic impact on local communities and national economies (*7*). The direct cost of providing every internally displaced person (totaling more than 55 million in 2020) with support for housing, education, health, and security has been estimated at US$370 per person per year, accumulating to more than US$20.5 billion for 2020 (*2*).

These figures are mostly based on information available from protracted conflict-related displacement situations because the economic impacts of displacements linked with disasters and climate change usually go unrecorded. A key knowledge gap exists here because only limited event-based or nationally aggregated data is available on how long people remain displaced after a disaster, despite ample evidence that this type of displacement is often long-term and can become protracted (*2*). These impacts can add up to billions of dollars worldwide. Each time one person is displaced, even for a few days, costs arise for transportation, shelter, food and non-food items, and the loss of income if the person cannot continue their usual work. Adding in long-term consequences, such as lack of schooling, training, and on-the-job experience, increases this economic impact. These costs should be on national balance sheets but are instead most often borne by communities themselves, by local governments that have to divert already limited development funds to response, and by humanitarian actors. In the face of increasingly severe disaster- and climate-related displacement, these costs are only set to rise.

The highest economic impacts usually stem from the loss of income and the need to provide displaced people with shelter and health care. Disaster-resilient housing and livelihoods, as well as strong primary health care systems, are also where investments are needed most ahead of disaster events to reduce displacement and enable lasting solutions. By nature of its mandate, humanitarian response is not set up to invest in resilient livelihoods or infrastructure and service development.

It is not only low-income nations that are at risk of economic impacts due to displacement. During the 2019–2020 bushfires in Australia, the loss of economic production as a result of people missing just one day of work during displacement was estimated to be about US$510 per person (*8*). These costs add up, particularly if a disaster causes considerable housing destruction, which may delay people from returning to their homes for months. The cost of covering housing needs resulting from Australia's Black Summer bushfires was estimated to be between US$44 million and US$52 million for a year, posing a substantial financial burden, given that previous recovery efforts indicate that it can take people between 1 and 4 years to rebuild their homes (*8*).

ASSESSING DISPLACEMENT RISK

These numbers and examples from across the globe highlight that we need to get better at understanding and assessing the nature and scale of disaster displacement risk. The coverage and detail of relevant datasets have improved, and various models and approaches exist at regional and global scales, although their time frames, methods, and resulting estimates vary enormously. For example, the World Bank, using a gravity model and new data on climate change, water availability, and crop production, has estimated that slow-onset climate hazards such as water scarcity and declining crop yields could lead to more than 100 million additional internal migrants in Latin America, South Asia, and sub-Saharan Africa by 2050 should neither accelerating climate impacts nor unequal development be adequately addressed (*9*).

In many such assessments, there is a strong focus on environmental stressors and hazards, and on climate change's impacts on their intensity and frequency. This may have potentially resulted in inflated

numbers and certainly in an inflated perception regarding the role of climate change in the dynamics of human mobility and forced movements today and in the coming decades. Estimates from a probabilistic model that takes housing rendered uninhabitable as a proxy for displacement in sudden-onset disasters, such as floods and cyclones, suggest that an average of around 14 million displacements can be expected each year (a conservative approach that is highly likely to be an underestimate) (*10*). This displacement risk is heavily concentrated in the Asia-Pacific region, where both exposure and vulnerability are high. Even in relative terms—that is, numbers of potential displacements in relation to population size—displacement risk is high not only for South and East Asia but also for Pacific and other small island states (see the first figure).

Climate change as well as changes in population size and composition and of key social and economic indicators all affect how this displacement risk may change in the future. According to probabilistic, spatially explicit risk modeling that uses ensembles of climate models and hydrodynamic modeling to quantify flood hazard, is calibrated on past events, and incorporates commonly used climate change and development scenarios, rapidly increasing exposure due to population growth may be the largest driver of displacement risk in the future (*11*). Nevertheless, this strong role of population size should not overshadow the fact that the substantial increase related to climate change is not trivial (see the second figure). New assessments show that we can expect a 50% increase in displacement risk related to floods for each degree of temperature increase (*11*). Although, currently, various epistemic uncertainties need to be reckoned with, such projections serve to illustrate the future burden to consider in a rapidly warming and changing climate.

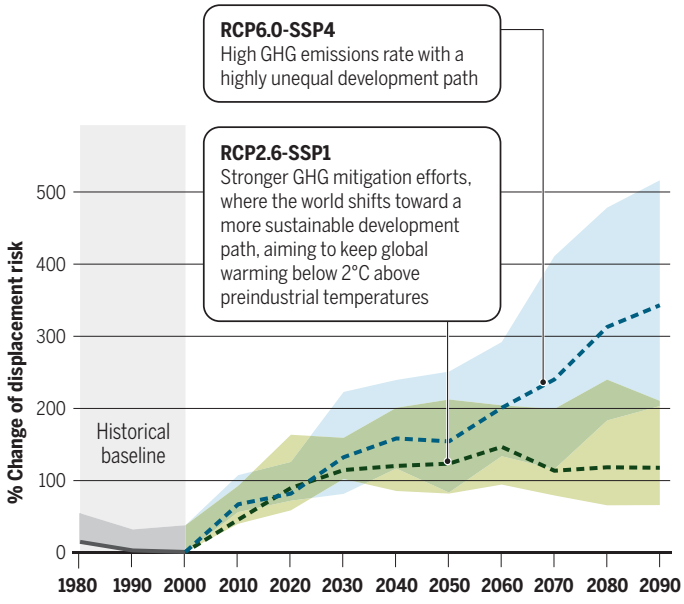
Beyond probabilistic and deterministic disaster displacement risk models, there are other modeling approaches that can increasingly be put to the task. Agent-based network models can assess individual-level impacts and costs through a bottom-up methodology that can reflect how shocks to one part of a system (community, economy, country, or region) can cascade through the whole system and also spill over into other systems (*12*). Further, a system dynamics approach can describe in a relatively comprehensive manner the relationships between a wide range of dimensions and indicators, although it requires granular datasets that are often unavailable and is highly cost- and labor-intensive to develop.

Finally, integrating risk estimates with analysis of public finance allows quantification of the relevance and "additionality" of internal displacement impacts on governments' (and often donors') budgets. First attempts at undertaking this analysis, adapting the International Institute for Applied Systems Analysis (IIASA) catastrophe simulation model (CatSim) in support of public financing strategies in pre- and postdisaster contexts, have shown that the cost of internal displacement can substantially increase national and global budget gaps (fiscal gaps) and the chance of budget crises (*13*). For example, in Bangladesh, a disaster with a return period of 50 years can be expected to incur costs related to internal displacement of nearly US$4.1 billion per year of subsequent displacement; a smaller magnitude but more frequent disaster with a return period of 10 years would incur more than US$1 billion. The estimated possible amount of funding that the country may be able to divert from existing development budgets and credit buffers adds up to just over US$1 billion of fiscal resilience, which means that Bangladesh is likely to be unable to cover the costs associated with internal displacement for events that occur every 10 years on average.

Further estimates of such costs can provide the basis for making the case for preventive action and for developing appropriate financial instruments such as national reserve funds, enhanced social protection schemes, and catastrophe bonds, as well as regional or global sovereign insurance pools (*14*). Beyond these first steps in developing basic estimates of the costs, further work is required to better understand who bears these and how benefits from improved policies would be distributed across different segments of society.

Change in flood displacement risk

Shaded areas show different scenarios of flood displacement risk based on a range of climate and hydrological models, relative to historical baseline. The width of the shading represents an estimate of the uncertainty induced by natural climate variability and limitations in current understanding of the climate system and hydrological systems. Dashed lines show the average values across models. Historical baseline is defined by the average flood hazard frequency and intensity from 1976 to 2005, combined with population data for 2000. RCPs reflect different trajectories of variation in atmospheric GHG concentrations. SSPs reflect different scenarios of global socioeconomic development. Modified from (*11*).



RCP, representative concentration pathway; GHG, greenhouse gas; SSP, shared socioeconomic pathway

INVESTING IN THE FUTURE

Comprehensive risk assessments that account for displacement risk and estimate its economic costs signal a need to improve coordination on budget allocations and cooperation in program execution across ministries and public and private sectors. This would enable the explicit inclusion of these contingent risks into budget stress-testing procedures and other risk-management planning processes. It would also provide incentives for managing risk with an ex ante approach, because it anticipates the ex post consequences and trade-offs involved in responding to shocks (*13*).

Risk assessments should help communities and local and national governments grappling with immediate displacement risk or the prospect of intensifying natural hazards or loss of territory or habitats. More financing must be made available for localized, granular displacement risk assessments, which municipalities can use to inform urban development plans, zoning regulations, and local building codes and for forward-looking, long-term planning for relocation where necessary.

Recent attempts at providing a measure for displacement risk and its impacts are

GRAPHIC: KELLIE HOLOSKI/*SCIENCE* BASED ON KAM *ET AL.* (*11*)

only the first step. In the coming years, further investment should build on the promises of longer-term risk modeling and couple its results with impact assessments so that countries can build displacement estimates into their multiyear development plans (*15*). Understanding needs and priorities in the decision-making processes of affected populations, institutional capacities, and socioeconomic dynamics, even if less systematically assessed, will be at least as important at indicating what the future holds. Given the scope and complexity of the problem, a pluralistic methodological setup is required to contribute to a better understanding of displacement risk and to inform effective policy and response under a broad range of circumstances. ■

REFERENCES AND NOTES

1. Intergovernmental Panel on Climate Change (IPCC), *Climate Change 2014: Synthesis Report. Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*, Core Writing Team *et al*., Eds. (IPCC, 2014).
2. Internal Displacement Monitoring Centre (IDMC), Global Report on Internal Displacement 2021 (2021); www.internal-displacement.org/global-report/grid2021.
3. IDMC, Global Internal Displacement Database; www.internal-displacement.org/database.
4. M. Brzoska, C. Fröhlich, *Migr. Dev.* **5**, 190 (2016).
5. Economics of Climate Adaptation (ECA), "Shaping climate-resilient development: A framework for decision-making. A report of the Economics of Climate Adaptation Working Group" (ECA, 2009); www.ethz.ch/content/dam/ethz/special-interest/usys/ied/wcr-dam/documents/Economics_of_Climate_Adaptation_ECA.pdf.
6. C. Raymond *et al*., *Nat. Clim. Chang.* **10**, 611 (2020).
7. C. Cazabat, L. Yasukawa, "Unveiling the cost of internal displacement. 2020 report," S. Ambrus, Ed. (IDMC, 2020); www.internal-displacement.org/sites/default/files/publications/documents/IDMC_CostEstimate_final.pdf.
8. E. du Parc, L. Yasukawa, "The 2019–2020 Australian bushfires: From temporary evacuation to longer-term displacement," J. Lennard, Ed. (IDMC, 2020); www.internal-displacement.org/sites/default/files/publications/documents/Australian%20bushfires_Final.pdf.
9. K. K. Rigaud *et al*., "Groundswell: Preparing for internal climate migration" (World Bank, 2018); https://openknowledge.worldbank.org/handle/10986/29461.
10. IDMC, "Global disaster displacement risk – A baseline for future work" (2017); www.internal-displacement.org/publications/global-disaster-displacement-risk-a-baseline-for-future-work.
11. P. M. Kam *et al*., *Environ. Res. Lett.* **16**, 044026 (2020).
12. A. Naqvi, F. Gaupp, S. Hochrainer-Stigler, *OR Spectrum* **42**, 727 (2020).
13. IDMC, IIASA, "Points of no return: Estimating governments' fiscal resilience to internal displacement" (IDMC, 2020); www.internal-displacement.org/sites/default/files/publications/documents/201903-fiscal-risk-paper.pdf.
14. J. Linnerooth-Bayer, S. Hochrainer-Stigler, *Clim. Change* **133**, 85 (2015).
15. S. Hochrainer-Stigler *et al*., *Int. J. Disaster Risk Reduct.* **24**, 482 (2017).

10.1126/science.abh4283

POLICY FORUM

Pathways to coastal retreat

The shrinking solution space for adaptation calls for long-term dynamic planning starting now

By **Marjolijn Haasnoot[1,2], Judy Lawrence[3], Alexandre K. Magnan[4,5]**

There is an urgent need to take coastal retreat more seriously as an option for adapting to sea level rise (SLR) and as a strategy capable of providing positive outcomes, if planned ahead. Early signs of such thinking are emerging. We demonstrate how exploring pathways to managed retreat adds value in the context of irreversible long-term SLR. Retreat is typically framed and understood as a single action, largely used after events rather than preemptively, and considered as a last resort. However, implementing managed retreat constitutes a multidecadal sequence of actions (i.e., across pathways) including community engagement, vulnerability assessment, land use planning, active retreat, compensation, and repurposing. This Policy Forum advances practical knowledge on what pathways to coastal retreat may look like and how they can pave the way for flexible and positive transformational adaptation, if started now.

SHRINKING SOLUTION SPACE

SLR globally accelerated from 1.4 mm/year (1901–1990) to 3.6 mm/year (2006–2015) and will continue to do so during this century (10 to 20 mm/year in 2100) (*1*). Sea levels could rise between 0.43 and 0.84 m globally by 2100, relative to 1986–2005, as a median estimate under low and high emission scenarios, respectively. However, a rise of 2 m by 2100 cannot be ruled out (*1*). There is also a clear commitment to SLR centuries into the future due to inertia in both the climate and ocean systems; for every degree of warming, sea levels will eventually rise ~2.3 m (*2*).

Inexorable SLR makes some degree of relocation of coastal residents, buildings, infrastructure, and activities inevitable, even if global warming is mitigated to 1.5° or 2°C. The necessity of paying more serious attention to pathways to managed retreat is becoming urgent (*3*). To begin with, observed coastal flooding is already reaching unacceptable levels for communities and infrastructure in many low-lying coastal settlements around the world (*1*), and unless adaptation starts now, in a few generations, more regions (e.g., small islands, parts of the US coast, major deltas) will be at risk of coastal flooding (*1*). Additionally, retreat requires decadal lead time to plan and implement equitably (*3*, *4*). Furthermore, many decisions taken today have a long legacy effect and create path dependencies, closing off some options in the future. For example, coastal defenses last for many decades and protected areas attract people and assets, which lead to expectations of further protection. On the other hand, creating space for wetlands to grow as sea levels rise provides a temporary buffer, keeping future options open for later development or a lower barrier to retreat.

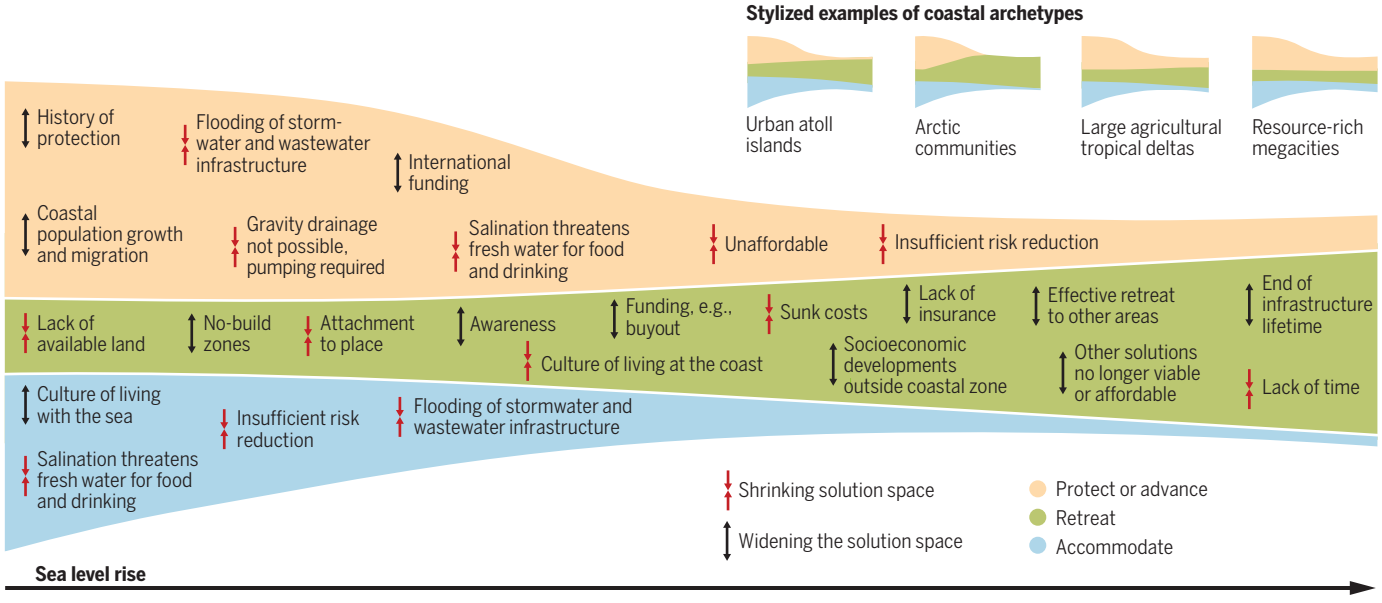
Ongoing and accelerating SLR, compounded with other climate-related changes (e.g., intensification of extreme events such as storms, heavy rainfall, and river flows) and increasing population at the coast, is already progressively shrinking the solution space of available adaptation options. Accommodation options (e.g., elevated buildings, early warning, and shelter) will not be enough to reduce coastal risks to acceptable levels under SLR-induced flooding and erosion. As sea levels rise, groundwater salinization will render water supplies unusable and limit food production to saline-tolerant crops. Nor will nature-based solutions, such as offshore reefs or wetland restoration, be likely to keep pace with combined climate change impacts (*1*) and human pressures that have reduced space and sediment supply to the coast. Such responses are therefore expected to be only temporary adaptations in many places (*5*).

Hard protection, either through holding the line (protect) or advancing seaward (advance) using levees, barriers, or artificial islands, can be beneficial, for example, in resource-rich megacities but also has limitations, as sustained and rapid SLR would make it increasingly difficult to extend infrastructure within available time frames (*6*). Also, hard protection will not be an affordable long-term solution for

[1]Deltares, Delft, Netherlands. [2]Faculty of Geosciences, Utrecht University, Utrecht, Netherlands . [3]New Zealand Climate Change Research Institute, Te Herenga Waka–Victoria University, Wellington, New Zealand. [4]IDDRI (Sciences Po), Paris, France. [5]LIENSs, La Rochelle University, La Rochelle, France. Email: marjolijn.haasnoot@deltares.nl

The evolving and shrinking solution space to address sea level rise

The colored areas show how the solution space to protect or advance, accommodate, and retreat changes as sea level rises. Different drivers and soft or hard limits shape this space. The figure highlights, first, a general narrowing of the solution space as a whole and, second, a change in the ratio between the three adaptation strategies, with retreat becoming dominant. This applies differently across coastal archetypes [derived from (*1*); see inset] owing to local contexts.



every community, nor will it address the impacts of rising groundwater and river flows in every coast (*6*) or the existing and increasing residual risks (e.g., when levees fail). In low-lying coastal areas across different geomorphologies and levels of development, retreat offers an alternative option (see the first figure) that ultimately removes vulnerability and risk in situ.

A DYNAMIC STRATEGY

Retreat is not easy, for various reasons, including attachment to place, high costs, lack of risk awareness, impacts on inland settlements, and political resistance (*3*). For example, retreat means sunk costs of existing investments in public infrastructure and private property and does not address the risk to cultural assets that cannot be relocated. However, among the reasons that make managed retreat beneficial is that it enables long-term change at the coast to be anticipated and planned for in an orderly way, which can minimize both stress on people and agencies and inequitable outcomes.

Exploring pathways can support staging retreat and help to break retreat into manageable steps over time, align it with maintenance or other social goals (e.g., economic development or environmental conservation), and implement retreat depending on how the future unfolds. This could help to overcome the societal resistance to retreat. Dynamic Adaptive Policy Pathways (DAPP) (*7*) planning is a practical approach developed to do exactly this and is increasingly used to support climate change adaptation decision-making. To date, DAPP planning has been used to address adaptation to SLR in several locations, including the Netherlands, the UK, the US, and New Zealand, where measures have included no-build zones and community and assets relocation (*5*, *6*, *8*). The long-term perspective puts retreat on the table next to protection and accommodation measures (see the first figure), avoiding increasing investments that eventually become higher sunk costs.

A first step in pathways planning is to assess the hazard, vulnerabilities, and uncertainties and to identify adaptation options. An adaptation option may fail to achieve objectives and/or may reach a performance limit or threshold (also referred to as an adaptation tipping point) as conditions change (e.g., SLR); a new or additional measure is then needed. Similarly, opportunities may arise (e.g., when infrastructure needs replacing or when people cannot tolerate SLR impacts and the need for retreat becomes obvious). The first figure presents some thresholds and opportunities for adaptation to SLR that change the solution space.

Next, by sequencing options, starting with low-regret and preparatory actions that can and/or need to be taken in the near term, pathways are designed while also testing options for their sensitivity to a range of SLR increments and to their path dependency. Pathways design is often done in a staged manner, with increasing depth of analysis. For part of the city of Miami, Florida, potential pathways were first developed using narratives, by asking stakeholders: What could be short-term, mid-term, and long-term adaptation options? What is the next option? Promising options and pathways were then further assessed using detailed models. In the Netherlands, a study assessed the solution space for multiple meters of SLR before exploring pathways. The study concluded that spatial planning that recognizes the consequences of long-term SLR is needed, because of the uncertain, potentially high SLR.

Monitoring is typically used to evaluate success of implementation but is also needed for detecting early warning signals on approaching thresholds and windows of opportunities for preemptive actions (e.g., new insights on future risks or new social values). This helps to identify when a decision to shift to another action is necessary. For adaptation to SLR, signals can be derived from climate drivers (e.g., mass loss from Antarctica, local SLR), impact signposts (e.g., flooding or freshwater availability) based on observations, and scientific studies and assessment [e.g., the Intergovernmental Panel on Climate Change (IPCC)] and, maybe more critically, from social, economic, and cultural signposts (e.g., insurance withdrawal, increased costs, and others developed with communities). Monitoring levels of (in)tolerable risk, increasing exposure to

GRAPHIC: KELLIE HOLOSKI/*SCIENCE* BASED ON M. HAASNOOT *ET AL.*

damage through population changes, and infrastructure aging could warn about potential lock-in or lock-out situations. Potential signals need to be evaluated for timeliness and reliability, while considering the required lead time for planning and implementation of next actions. This is problematic in a context of increasing and accelerating coastal risks, where physical and societal thresholds occur close together, with limited time left for implementation, and where communities are dependent on critical infrastructure, the functioning of which is already threatened. For example, in Florida, several water infrastructure thresholds are close or have been reached, where nuisance flooding is observed and the septic systems are being compromised by rising groundwater tables. New infrastructure with pumps and drainage can only buy a limited amount of time (*8*).

Beyond mapping the solution space that includes retreat, pathways thinking is also critical to supporting the design and implementation of the transition to retreat, as presented with the nested pathways in the second figure.

PRACTICAL PATHWAYS INSIGHTS

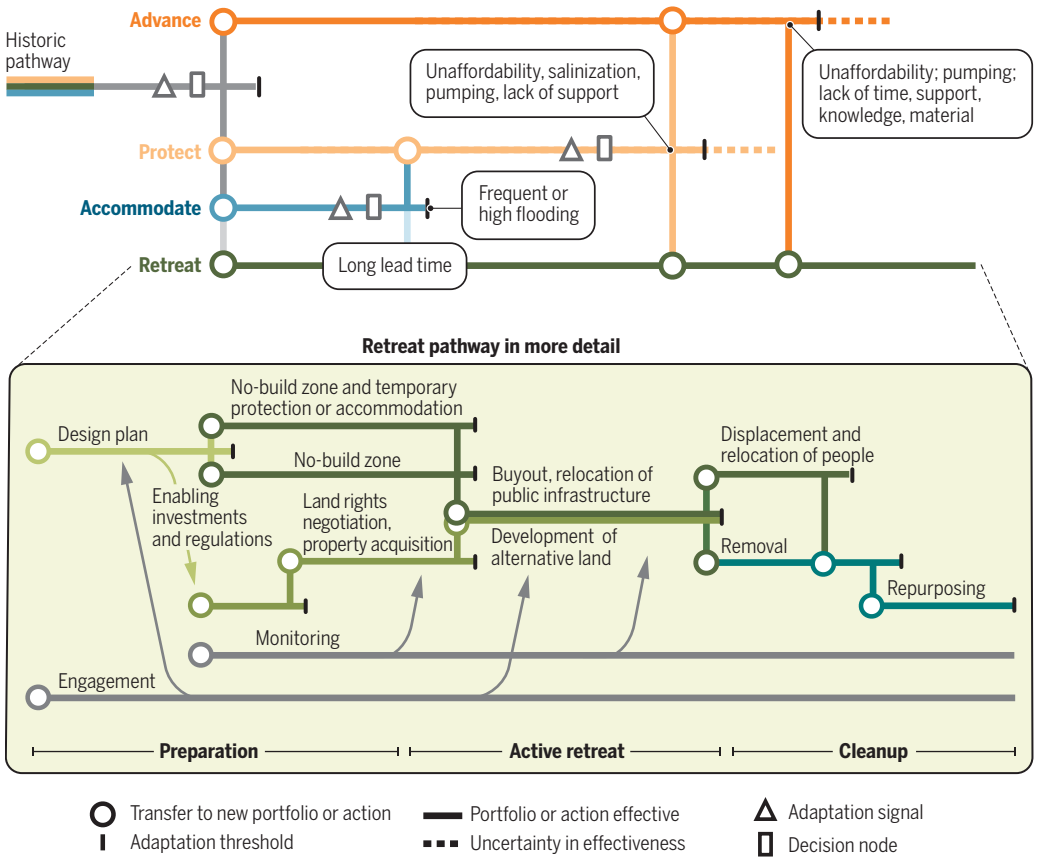
Although the relevance, extent, rate, and modalities of managed retreat will vary depending on SLR and local context, three generic steps can be highlighted across coastal settlements: preparation, active retreat, and cleanup (*5*). Enabling decision-makers to progressively prepare includes engagement to gain community understanding of the risks and to understand social values and vulnerabilities; planning to identify options, exploring pathways, and establishing monitoring plans to detect signals of opportunities (e.g., early moves, end of lifetime of infrastructure); funding for property acquisition and infrastructure provision in alternative areas; and adjustment of land use plans and regulations. These preparatory actions support active retreat, which comprises the acquisition of property, buyout, and removal of structures or relocation of houses, people, and economic activities. The last step, cleanup, comprises land rehabilitation and repurposing (e.g., for coastal amenity and recreational uses that can relocate readily) until that land is permanently flooded by the sea.

Because implementing managed retreat can take decades, it needs to be considered well ahead of any climate-induced societal and physical thresholds (*9*). The time needed depends on each society's willingness and ability to anticipate the climate risks and to act on them before observed impacts. Time is also needed to plan and engage with those affected about the urgency to start the retreat process now, so that individuals can make relocation decisions as opportunities arise. For example, in the Netherlands and New Zealand, retreat to enable river floodplain restoration was signaled well ahead of project implementation in anticipation of the effects of climate change (*5*, *10*), which gave time (25 and 10 years, respectively) for eventual removal of houses and purchase of at-risk properties on a voluntary basis. This contrasts with instances where retreat has been triggered after damaging climate events (e.g., after hurricanes Sandy in New York and Katrina in New Orleans (*4*, *10*); where protection proved ineffective and retreat was forced, creating additional community stress and costs [e.g., after a storm and mudslide in New Zealand (*11*)]; or where forced retreat to a flood-safe area was unsustainable because work was unavailable in the new location [e.g., in the Philippines (*12*)]. In the Carteret Islands, Papua New Guinea, resettlement of island populations created negative outcomes owing to a lack of economic opportunities in the relocation areas, land tenure conflicts with established populations, and disruptions to local communities that were not planned for (*13*). These examples illustrate the social consequences of retreat if it does not take a planned and staged pathways approach.

To determine when to start active retreat, one can assess under what conditions retreat is required because of limitations of other strategies, indicating the latest moment at which active retreat should be realized. Another way is to assess the conditions under which retreat becomes more beneficial than

Indicative adaptation pathways of retreat

Retreat is presented as a nested pathway within a broader pathways map, including advance, protect, and accommodate. Retreat comprises three stages: preparation, active retreat, and cleanup. Engagement and monitoring support planning and implementation (gray lines). After designing a plan, land use regulations and temporary measures can be implemented, followed by buyout. Enabling investments and regulations are precursor actions.



GRAPHIC: KELLIE HOLOSKI/*SCIENCE* BASED ON M. HAASNOOT *ET AL.*

other strategies accounting for flood risk, alignment with social goals, and costs. For example, Kool *et al*. (*14*) worked backward from an infrastructure threshold for SLR of 30 cm, at which point a gravity-based stormwater and wastewater system would need to be replaced by a pumped system. Before that point, the costs for a new system, its lifetime, and the opportunity costs to the community would need to be assessed against the costs and benefits of a retreat option that helps remove the ongoing impacts from SLR. Using pathways for adjacent locations, they identified opportunities for drainage system redesign to buy time for engagement with the community before eventual retreat. Such a strategy consisting of progressive steps can result in a beneficial transition that is supported by the community.

An increasing number of studies (*3*, *5*, *10*, *15*) provide lessons for developing robust pathways to coastal retreat: (i) engaging early with affected communities to build understanding of their risk tolerance, vulnerabilities, and values; (ii) enhancing the policy and public understanding of higher risk levels than in the past; (iii) early design of and contributions to design of funding mechanisms and regulations that can enable implementation of retreat; (iv) avoiding developments in places recognized as risky and where existing urbanization trends can be reversed through no-build zones and prohibited land uses; (v) considering locations for new developments or designing them to be movable; and (vi) considering whether buying time through temporary accommodation, protection, or nature-based measures will trigger greater risk exposure and therefore worsen the problem over time, or whether these approaches facilitate a transition to retreat.

NECESSARY ENABLERS

Inexorable SLR that will continue for centuries means that for many low-lying coastal areas worldwide, retreat is an inevitable adaptation action. If planned now and integrated with social, economic, and cultural goals, the anticipatory and dynamic pathways to retreat can be a positive approach to reduce coastal risks and minimize regret of investments and social inequities.

To allow retreat to be considered a serious option and implemented where appropriate, there are a number of necessary enablers that require further attention by the research and policy communities. These include: (i) improved understanding of how SLR is a changing risk over time that requires a shift from static to dynamic pathways decision-making and how this affects communities differently now than in the past; (ii) improved understanding of what managed retreat comprises and how it can be staged over time through monitoring and sharing experiences; (iii) development of policies and regulations that are grounded in anticipatory planning supported by sustainable funding arrangements; (iv) further development of analytical methods relevant to changing risk, such as for mapping the shrinking solution space and identifying if and when retreat will be needed; (v) further assessment of the effectiveness of the range of adaptation responses under alternative futures and how retreat can be integrated with wider societal goals; and (vi) enhancement of the role of political leadership in building community trust in preparation for managed retreat, and embedding commitment devices to maintain the long-term dynamic approaches for reducing SLR risks.

Notably, the development and the implementation of any retreat pathway fundamentally depends on the past trajectory of coastal risks; the present situation (governance, coastal strategy, observed impacts, individual and institutional values and attitudes toward climate-related risks); the envisioned future; and when and under what conditions adaptation opportunities and limits appear. Whatever the context considered, it is increasingly evident that the shrinking solution space for adaptation in low-lying coastal areas calls for long-term dynamic pathways planning now. ■

REFERENCES AND NOTES

1. M. Oppenheimer *et al*., "Sea level rise and implications for low-lying islands, coasts and communities" in *IPCC Special Report on the Ocean and Cryosphere in a Changing Climate*, H.-O. Pörtner *et al*., Eds. (IPCC, 2019).
2. A. Levermann *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **110**, 13745 (2013).
3. A. R. Siders, M. Hino, K. J. Mach, *Science* **365**, 761 (2019).
4. K. J. Mach *et al*., *Sci. Adv.* **5**, eaax8995 (2019).
5. J. Lawrence *et al*., *Curr. Clim. Change Rep.* **6**, 66 (2020).
6. M. Haasnoot *et al*., *Environ. Res. Lett.* **15**, 034007 (2020).
7. M. Haasnoot, J. H. Kwakkel, W. E. Walker, J. ter Maat, *Glob. Environ. Change* **23**, 485 (2013).
8. J. Obeysekera, M. Haasnoot, R. Lempert, *US CLIVAR Variations* **18**, 1 (2020).
9. S. A. Stephens, R. G. Bell, J. Lawrence, *Environ. Res. Lett.* **13**, 104004 (2018).
10. M. Hino, C. B. Field, K. J. Mach, *Nat. Clim. Change* **7**, 364 (2017).
11. C. Hanna, I. White, B. Glavovic, *Sustainability* **12**, 736 (2020).
12. J. See, B. Wilmsen, *Glob. Environ. Change* **65**, 102188 (2020).
13. J. Connell, *Aust. Geogr.* **43**, 127 (2012).
14. R. Kool, J. Lawrence, M. Drews, R. Bell, *Infrastructures* **5**, 92 (2020).
15. A. K. Magnan, V. K. E. Duvat, *Reg. Environ. Change* **20**, 119 (2020).

ACKNOWLEDGMENTS

We thank C. Kraan, S. McEvoy, and A. Reisinger for feedback and I. van den Broek for the figures. J.L. thanks the NZ Resilience National Science Challenge Enabling Coastal Adaptation Programme (GNS-RNC040) and NZ SeaRise Endeavour Programme (RTUV1705). A.K.M. thanks the French National Research Agency (STORISK ANR 15-CE03-0003 and "Investissements d'avenir" ANR-10-LABX-14-01).

10.1126/science.abi6594

POLICY FORUM

High-density population and displacement in Bangladesh

The strategy promotes "migrant friendly" towns and selective relocation abroad

By **Mizan R. Khan, Saleemul Huq, Adeeba N. Risha, Sarder S. Alam**

Among the many adverse impacts of climate change in the most vulnerable countries, climate change–induced displacement increasingly caused by extreme weather events is a serious concern, particularly in densely populated Asian countries. Reports by the Intergovernmental Panel on Climate Change (IPCC) project a grim picture for South Asia, the most populous region on Earth, home to about one-quarter of global population, with the highest poverty incidence. A combination of poor socioeconomic indicators and increased frequency and intensity of cyclones and floods renders the region extremely vulnerable. Meanwhile, slow-onset climate hazards, such as sea level rise, salinity intrusion, water stress, and crop failures gradually turn into larger disasters. Within South Asia, Bangladesh stands as the most vulnerable: 4.1 million people were displaced as a result of climate disasters in 2019 (2.5% of the population), 13.3 million people could be displaced by climate change by 2050, and 18% of its coastland will remain inundated by 2080 (*1*). We describe how, faced with such natural and human-made adversities, Bangladesh can stand as a model of disaster management, adaptation, and resilience.

The Paris Agreement goal of keeping the temperature rise at 1.5°C or well below 2°C compared to pre-industrial times may not be achieved, given the lack of ambitious mitigation. As a result, the number of people estimated to be displaced by slow-onset events will stand at ~22.5 million by 2030 and ~34.4 million by 2050 (*2*). A combi-

International Centre for Climate Change and Development (ICCCAD), Independent University, Dhaka, Bangladesh. Email: adeeba.nuraina@icccad.or

nation of sudden and slow-onset climate events, which affect all elements of the environment, becomes the main driver of environmental displacement.

Migration is an adaptation strategy. An estimated half a million people move to Dhaka, the capital city of Bangladesh, each year. Migration of this magnitude presents a challenge for Bangladesh given its small land area (147,570 km^2) and high population density (~1100/km^2). There is simply little space for retreat: Bangladesh's population is half that of the United States, living on ~1.5% of the land area of the United States.

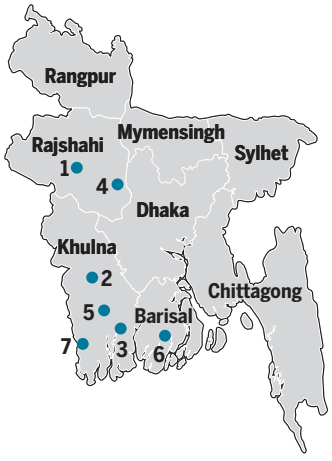
Usually, three pathways can be discerned with respect to how displaced people are settled: autonomous relocation by displaced individuals (without much government support), government-supported temporary settlement, and planned relocation. In Bangladesh, the first option overwhelms, followed by efforts for temporary settlement, until the government rehabilitates their former residences. Planned relocation or managed retreat in response to climate change (*3*) is not yet happening widely because of space and resource constraints.

Since the founding of Bangladesh in 1971, and even earlier in Pakistan, government-planned relocation of people displaced by riverbank erosion has fueled ethnic conflicts in the Chittagong Hill Tracts in the southeast part of the country, because the move was not backed by consultations with tribal communities. About 100,000 of more than a million Rohingya refugees in Bangladesh, fleeing persecution in Myanmar, are being relocated to Bhasan Char, an island in the Bay of Bengal. In land-hungry Bangladesh, most of the 30+ such Chars/mudflats in the bay are already inhabited at different degrees by people displaced by riverbank erosion and climate change.

Despite these odds, Bangladesh is a leader in economic growth among developing countries and in mainstreaming climate change into its development strategy. Partially in response to scientific findings, the National Strategy on the Management of Disaster and Climate Induced Internal Displacement (NSMDCIID) adopted in 2015 incorporated disaster risk reduction and rights-based approaches, so that vulnerable communities can enjoy their basic rights to livelihood, food, health, and housing. The Strategy is built on an integrated Displacement Management Framework, in line with the migration management cycle of the International Organization of Migration (IOM). This Framework elaborates responses during the three phases of mobility management: pre-displacement [disaster risk reduction (DRR)], displacement (emergency), a nd post-displacement (rehabilitation/relocation). Under the Strategy, the government has initiated support for livelihood opportunities, housing, and human development of displaced people in vulnerable hotspots. It is likely that the government-supported community mobilization and disaster management and DRR policies, both before and after adoption of this Strategy, were helpful in lessening the number of casualties from the supercyclone Amphan in May 2020.

Building migrant-friendly, climate-resilient cities in Bangladesh

The map shows some activities being undertaken to build migrant-friendly and climate-resilient cities in Bangladesh. Descriptions of activities are based on publicly available information about the programs, and on discussions with representatives of the NGO BRAC.



1 Rajshahi (city)
Provision for diversified livelihoods; training for skill development; access to health services and education.

2 Noapara
Investments in fish and shrimp export industries attract migrants; municipal authorities have improved water supply and drainage systems, reducing waterlogging and vulnerability.

3 Mongla
Investments in infrastructure projects attract migrants; government is building a vocational training institute; municipality and NGOs work to improve education and slum housing.

4 Sirajganj
Assessment of climate/disaster vulnerability; preparation and implementation of a risk reduction action plan at community and household levels.

5 Khulna (city)
Flood-resilient systems in slums; sanitation; training on climate-adaptable livelihood options; access to financial products in government-approved financial institutions and microcredits.

6 Barisal (city)
A baseline study is in preparation; a community-managed piped water supply network is arranged.

7 Satkhira
Community-led, low-cost, climate/disaster-resilient housing.

GRAPHIC: N. DESAI/*SCIENCE*

MIGRANT-FRIENDLY TOWNS

One way to address displacements under increasing urbanization across the world could be the establishment of peri-urban growth centers and transformation of cities and towns to be migrant-friendly. This option appears practicable for populous countries such as Bangladesh, having little space for retreat from vulnerable hotspots. To achieve this, institutional changes in a city need to be fostered by research, planning, design, and capacity building. Examples from cities such as Durban, Quito, Semarang, and Malé indicate that cities may need to develop general as well as sector-based strategies to manage effective climate change adaptation (*4*). This warrants the linking of adaptation planning and implementation to city priorities. Cities must have access to reliable information and opportunities to share experiences through local, regional, national, and international networks (*4*).

National and local governments should develop migrant-friendly plans along three lines: building of resilient hardware, such as low-cost housing, industries for employment generation, and other infrastructure; software, such as legal, policy, and institutional frameworks; and "heart-ware"—the promotion of awareness, reflecting values and ethics. The basic parameters for safe and orderly movement for migrants are to ensure employment, social protection, access to education, housing, health services, utilities, etc. Although government support is important, engagement of the private sector, nongovernmental organizations (NGOs), civil society, and university-led research can strengthen municipal adaptation efforts.

This is what the International Centre for Climate Change and Development (ICCCAD) in Bangladesh has been doing—to facilitate the transformation of smaller peripheral towns to be migrant-friendly as a climate adaptation strategy (see the figure). Our work has multiple purposes: to shift the tide of migration away from Dhaka and other large cities toward smaller towns, and to decentralize climate-resilient development and facilitate planning for basic services and amenities.

In Bangladesh, a majority of those displaced by climate change prefer non-migration from their ancestral roots (*5*) if they are provided support for improving their livelihood, housing, etc. Settlement of displaced people in a town nearer to their ancestral home allows them to maintain psychological kinship and cultural comforts. On the basis of such local context and needs, each migrant-friendly town needs its own development and adaptation plans to address climate risks and economic opportunities.

The NGO BRAC has initially identified about 20 towns and municipalities, considering their economic potential and climate stress, to determine whether they can absorb a sizeable number of displaced people. A number of satellite towns adjacent to economic hubs, such as relatively elevated sea and river ports and export processing zones (EPZs), can potentially employ millions of migrants. Investment in manufacturing and/or services is generating jobs through public, private, and community partnerships, such as private investments, government support, and microfinancing from BRAC and Grameen Bank. ICCCAD has formal agreements with many ministries and agencies including the Local Government Engineering Department (LGED), the agency for building and maintenance of rural infrastructure. ICCCAD has been working as an advisor and co-implementer of programs with all stakeholders, including mayors in two small towns in coastal Bangladesh, Mongla and Noapara (see the figure). It is helping town authorities in planning and implementing initiatives that are intended to be hospitable to incoming settlers, so that they can gradually be mainstreamed into citizenship (*6*). The process is based on a participatory, consultative process involving the municipal authorities, host community leaders, and settlers.

The Strategy (NSMDCIID) includes options such as supporting livelihood for new settlers and skill development, both in displacement hotspots and in new settlements. Although these towns do not yet have adaptation plans as such, the programs consider risk-informed and socially conducive adaptation measures. BRAC with its Climate Bridge Fund is also currently implementing different programs in five cities: Khulna, Rajshahi, Satkhira, Barisal, and Sirajgonj. For programs under implementation in these cities, the target groups are incoming migrants, who crowd the slums. The activities undertaken in these cities are similar, with some specific activities in each town (see the figure). Most of the new settlers have moved from rural areas rendered inhospitable as a consequence of slow and sudden-onset climate impacts. ICCCAD started facilitating this program 3 years ago with a strategy of learning by doing. Among the lessons learned: (i) Vibrant economic activities in these rapidly growing towns are absorbing increasing numbers of migrants from vulnerable hotspots, and (ii) migrants with energy and agency are engaging themselves in different small businesses, with government support and microcredits from Grameen Bank and BRAC.

The fact that an overwhelming share of those displaced by climate change around the world resettle internally indicates that adaptation in-country is the most viable option. The global community dealing with disaster displacement, including the United Nations Framework Convention on Climate Change (UNFCCC), primarily recommends this option. However, it requires adequate international support, which developed countries are obligated to deliver (with the language "shall provide") under the UNFCCC and the Paris Agreement. Unfortunately, adaptation finance continues to remain the "poor cousin" of mitigation, the ratio remaining 20:80 despite repeated pledges by developed countries and agencies. For domestic resource mobilization, some countries (for example, Fiji) have introduced an adaptation levy on all goods and services produced and consumed in the country.

SELECTIVE RELOCATION ABROAD

There are limits to relocation in-country; sudden and slow-onset events sometimes trigger cross-border movement of individuals seeking jobs and protection. The UN Commission on Human Rights argues for looking at such mobility from a human-rights perspective (i.e., the space for realizing the basic human rights of livelihood, health, housing, etc.). Currently, those displaced by climate change suffer an international protection deficit, not qualifying as "refugees" under the 1951 Geneva Convention. Consideration of those displaced by climate change began in 2008 under the UNFCCC, with research and advocacy. The Cancun Adaptation Framework (Decision 1./CP16, paragraph 14f) provides for different types of climate-induced human mobility (displacement, migration, and planned relocation), different scales of mobility (national, regional, and international), and different actions (research, cooperation, and coordination). This decision recognized migration as an adaptation strategy. The Nansen Initiative in 2011–2012 focused on promoting research and planned relocation. The Paris Agreement established a Task Force on Displacement under the Warsaw International Mechanism, with mandates to make recommendations for averting, minimizing, and addressing climate change–induced displacement. Finally, the Global Compact on Safe, Orderly, Regular and Responsive Migration was adopted in 2018 as the first multilateral framework to cooperate on migration, including in response to climate change.

Many major countries and think tanks started looking at climate displacement through a lens of national security, with its characterization as a "threat multiplier," and a number of nationally determined contributions under the Paris Agreement refer to those displaced by climate change as potentially fueling national and regional conflicts (*7*). However, climate security can be looked at either from a conflict perspective or from a lens of vulnerability-focused human and global security (*8*). The "conflict view" proponents call for closing the borders, but still the result of such a policy ends up being a humanitarian disaster, caused primarily by actions beyond the control of those being displaced or of their home countries. Should we see more of these displaced and disgruntled youth as victims in the hands of human traffickers? If not, we then argue—viewing this displacement in terms of vulnerability-focused human security—that planned relocation internationally can be an effective way forward under paragraph 14f of the Cancun agreement.

As multilateral processes are typically very cumbersome and painstakingly slow,

PHOTO: SUVRA KANTI DAS/ABACA/SIPA USA/AP IMAGES

Floods and erosion—such as along the Padma River in Louhajang, Munshiganj, on 23 July 2020—ravage communities, whose residents must retreat.

bilateral action can be more rapid and effective, and may then gradually feed into regional and global initiatives. For example, the Seasonal Migrant Worker Program in Australia and New Zealand, or New Zealand's Climate Visa Program (*9*), attract migrants from the Pacific Small Island States (although these initiatives are not solely meant for absorbing migrants displaced by climate change). Canada and the United States offered immigration opportunities to typhoon Haiyan victims, but these were based on kinship relations (*10*). Although the EU does not have a common policy, Finland and Sweden changed their earlier liberal policies on climate-induced displacement after the refugee influx from Syria (*11*). There are also provisions of circular migration, as between Spain and Colombia. The IOM continues recommending such migration between developed and developing countries as an adaptation response to climate-induced vulnerability. The Bangladesh Strategy recommends such options as well.

Many developed countries already suffer from demographic deficits, with negative growth, and increasingly aging cohorts. The rhetoric in many of these counties, which often is anti-immigrant, cannot change the reality that these countries will need more and more young and skilled labor. Using projected needs of specific skills, developed countries could thus enter into bilateral agreements with climate-vulnerable countries, where those displaced by climate change may be trained in jointly supported educational and training institutions, either for permanent or for circular migration. For example, under the "Triple Win" program, Germany recruits nurses from Serbia, Bosnia-Herzegovina, and the Philippines to meet their nursing shortage, while reducing unemployment and contributing to economic development in the countries of origin (*12*). It is only just and fair for developed-country emitters of greenhouse gases to take some responsibility under Article 3.1 of the UNFCCC for their disproportionate contributions to generating this increasing number of people displaced by climate change.

Lessons suggest that migration to rich countries can have strong positive impacts on labor market, GDP growth, and public revenue for host countries (*13*, *14*). Migration is also typically positive for countries of origin, through remittance, transfer of technology, skills, domestic consumption and GDP growth, housing, children's education, and more. In 2017, low- and middle-income countries received more than $466 billion in remittances, three times the amount of official aid (*15*). This presents an important indicator of the effects that bilateral agreements on migration of climate-displaced people may have on promoting many different Sustainable Development Goals. Such migration should be framed as a win-win option, not as climate humanitarianism (*10*).

The Bangladesh Strategy (NSMDCIID) argues for creating "opportunities for international labor migration by one or few members of families from the displacement hotspots" (p. 115). Older and underage family members and spouses can stay behind and rebuild their lives with remittance support. We believe this option of selective, not wholesale, relocation as a pragmatic policy can be scaled gradually, as warranted by projected demands of skills over time in developed countries. This relocation is based on bilateral planning and preparation, unlike the conventional, voluntary migration of skilled labor to industrial countries.

This option is challenging, though mutually rewarding. However, acceptance of this proposal by Western democracies depends on whether they are ready to embrace and enjoy more of "smart/pooled" sovereignty, with enlightened self-interests under climate-induced vulnerability interdependence, rather than holding on to a centuries-old "Westphalian" model of a zero-sum game in global cooperation. Many have argued that with the increasing number of global commons problems, we now live in a positive-sum world. But such a paradigm shift warrants a vigorous campaign to raise awareness among citizens in industrial countries about the "new normal" of increasing extreme and ever-growing slow-onset events. Those citizens and politicians must face the lead and obligatory responsibility their countries have assumed under the international climate regime to support adaptation in vulnerable countries. Such awareness must confront and overcome the xenophobia and anti-immigration sentiments that often surface in many countries, inhibiting the enjoyment of mutual dividends, which can contribute to real and sustainable global peace and security.

GERMINATE COORDINATION

Successful implementation of the two options raised above (migrant-friendly towns and bilateral agreements for international migration) could help to germinate coordinated implementation, as stipulated in the Cancun agreement, of global policy frameworks on climate change (UNFCCC), disaster risk reduction (Sendai Framework), and human migration (Global Compact for Migration). As many ideas and actions on planned internal or international relocation of climate change–induced displacement are relatively new in the national and global policy domains, continued research and science-policy interface are essential in order to determine the feasibility, efficacy, and scalability of these options. ■

REFERENCES AND NOTES

1. K. Rigaud *et al*., "Groundswell: Preparing for Internal Climate Migration" (World Bank, 2018).
2. H. Singh, J. Faleiro, T. Anderson, S. Vashist, "Costs of Climate Inaction Displacement and Distress Migration" (Actionaid, 2020).
3. J. Carmin, D. Roberts, I. Anguelovski, "Planning Climate Resilient Cities: Early Lessons from Early Adapters" (2011), pp. 5–8.
4. S. Weerasinghe *et al*., "Planned Relocation, Disasters and Climate Change: Consolidating Good Practices and Preparing for the Future" (UNHCR, 2014).
5. B. Mallick, K. G. Rogers, Z. Sultana, *Ambio* 10.1007/s13280-021-01552-8 (2021).
6. S. S. Alam, S. Huq, F. Islam, H. M. A. Hoque, "Building Climate-Resilient, Migrant-Friendly Cities and Towns" (International Centre for Climate Change and Development, 2018).
7. E. Wright, D. Tänzler, L. Rüttinger, "Migration, Environment and Climate Change: Responding via Climate Change Adaptation Policy" (German Environment Agency, 2020).
8. M. R. Khan, *Toward a Binding Climate Change Adaptation Regime: A Proposed Framework* (Routledge, 2014), chapter 6.
9. H. Dempster, "New Zealand's 'Climate Refugee' Visas: Lessons for the Rest of the World" (Centre for Global Development, Washington, DC, 2020).
10. D. M. S. Matias, *Clim. Change* **160**, 143 (2020).
11. A. Kraler, K. Caitlin, M. Wagner, "Climate Change and Migration: Legal and Policy Challenges and Responses to Environmentally-Induced Migration" (European Union, 2020).
12. German Development Agency, "Sustainable Recruitment of Nurses (Triple Win)" (2019); www.giz.de/en/worldwide/41533.html.
13. E.-j. Quak, "The effects of economic integration of migrants on the economy of host countries" (Institute of Development Studies, London, 2016).
14. V. Grossmann, "How Immigration Affects Investment and Productivity in Host and Home Countries" (IZA, 2016); https://wol.iza.org/articles/how-immigration-affects-investment-and-productivity-in-host-and-home-countries.
15. World Bank, "Record high remittances to low- and middle-income countries in 2017" (2018); www.worldbank.org/en/news/press-release/2018/04/23/record-high-remittances-to-low-and-middle-income-countries-in-2017.

10.1126/science.abi6364

REVIEW

Reframing strategic, managed retreat for transformative climate adaptation

Katharine J. Mach[1,2]* and A. R. Siders[3,4]

Human societies will transform to address climate change and other stressors. How they choose to transform will depend on what societal values they prioritize. Managed retreat can play a powerful role in expanding the range of possible futures that transformation could achieve and in articulating the values that shape those futures. Consideration of retreat raises tensions about what losses are unacceptable and what aspects of societies are maintained, purposefully altered, or allowed to change unaided. Here we integrate research on retreat, transformational adaptation, climate damages and losses, and design and decision support to chart a roadmap for strategic, managed retreat. At its core, this roadmap requires a fundamental reconceptualization of what it means for retreat to be strategic and managed. The questions raised are relevant to adaptation science and societies far beyond the remit of retreat.

Evolving social norms, technologies, and economies will create futures that fundamentally differ from our world today. Climate change constrains the range of possible futures and affects the level of transformation societies will experience. Extreme humid heat (*1*), limited freshwater availability (*2*), wildfire (*3*–*5*), sea level rise, and flooding (*6*–*10*), among other climate-exacerbated hazards, threaten relationships among people and places, the persistence of human settlements, and even human survivability in some locations. The extent of climate change, and how societies respond, will depend on which risks people decide not to tolerate and which values—and whose—are prioritized (*11*–*14*).

Numerous adaptation actions—often categorized as resistance, accommodation, avoidance, retreat, and advance (*7*, *15*)—can address climate risks. Resistance reduces exposure by armoring or keeping hazards at bay, whereas advance builds into hazardous areas (e.g., through land reclamation). Accommodation measures such as infrastructure elevation or warning systems reduce sensitivity to hazards. Avoidance limits new development in hazardous locations, and retreat removes people and assets from hazardous places after development has occurred. Each adaptation action represents a distinct value-laden decision about what to preserve, purposefully change, or allow to change unguided (*16*). Retreat has often been viewed as a failure to adapt or considered only when all other options are exhausted (*12*, *17*–*21*). But this conceptualization ignores lessons from numerous disciplines drawing on a long history of human movement and limits adaptation researchers and decision-makers in preparing for a broad range of futures.

The term "retreat" refers both to autonomous relocations and to coordinated movements supported by governments and organizations at multiple scales (*17*, *22*, *23*). Relatively modern social institutions, such as borders and policies designed for a climate assumed to be unchanging, now hinder rapid relocations (*14*, *24*–*26*), but throughout history, people have moved and retreated in response to climate (*27*, *28*). Retreat occurs today in the face of relatively moderate climate change (*17*, *23*, *29*, *30*). In the future, retreat will be a component of many climate-driven transformations that involve fundamental shifts in societies. Whether as a minor element or a substantial factor in portfolios of response, history suggests that movement will occur (Fig. 1).

Retreat can be more effective at reducing risk, more socially equitable, and more economically efficient if it is managed and strategic—i.e., if it is designed and executed in ways that promote broader societal goals (*30*, *31*). Serious consideration of retreat, even if subsequently rejected, can help communities articulate why remaining in place is a core value, what costs they are able to endure or what opportunities they can forego to remain in place, and what responses sustainably support individual and community priorities (*7*, *14*, *32*). Achieving this potential requires a reconceptualization of retreat, not as failure to adapt or last resort but as an adaptive option that can proactively support social values through a plurality of specific measures.

In this Review, we draw on empirical and theoretical literature to argue that societies will adapt more effectively and equitably—as evaluated against the values of individuals, communities, and societies—if they consider the potential role of strategic, managed retreat. We (i) outline ways in which strategic, managed retreat diverges from past practice; (ii) reframe two often-posed questions to demonstrate how a reconceptualization of retreat drives a value-informed and radically interdisciplinary adaptation science agenda; and (iii) discuss the role of retreat in creating adaptation visions and pathways that bridge the chasms between the present and long-term resilience.

Strategic, managed retreat differs from past practice

To date, managed retreat projects have been largely incremental, minor adjustments implemented using a handful of policy tools, guided by a limited set of social values, and small scale in their contributions to climate change adaptation (*15*, *23*, *33*). For example, in the United States, voluntary home buyouts have helped ~45,000 families move out of flood-prone homes over the past 30 years; this represents a tiny fraction of the millions at risk and is fewer than the number of homes experiencing repeat flood damage and the number of new homes built in floodplains. Buyout programs often affect just a handful of households within a flood-prone community, while simultaneously raising equity concerns about who is offered buyouts and how they are treated in the process of relocating (*19*). Around the world, retreat programs frequently involve relatively few people, compared with the number of people at risk, and are disconnected from larger strategies for climate adaptation (*17*).

Although many retreat programs have reduced risk exposure and economic losses, their effects on broader individual and social goals are contested (*7*, *14*, *21*, *29*). "Successful" retreat will be defined differently by diverse stakeholders and may change over time and across contexts (*11*, *13*, *34*). Nevertheless, the literature offers key areas for improvement, which we review in turn below: (i) proactively envisioning managed retreat as a feature of climate-driven transformations, (ii) engaging a diversity of approaches for strategic retreat, and (iii) coordinating across disciplinary and governance silos to connect societal priorities. We posit that these changes have not occurred because addressing them requires difficult systemic and structural adjustments and because retreat has not yet been adequately prioritized despite substantial efforts of some communities, practitioners, and researchers to overcome the barriers. Nevertheless, there are reasons to believe changes may yet occur.

Retreat as a feature of transformations

Retreat has figured into disaster risk management and adaptation portfolios largely as a deprioritized, politically perilous option (*7*, *15*, *20*, *32*, *33*), in part because the motivation in most cases has been to avoid transformation: to enable people to continue living where and how they have in the past. Yet as climate conditions shift outside the bounds of

[1]Rosenstiel School of Marine and Atmospheric Science, University of Miami, Miami, FL, USA. [2]Leonard and Jayne Abess Center for Ecosystem Science and Policy, University of Miami, Coral Gables, FL, USA. [3]Disaster Research Center, University of Delaware, Newark, DE, USA. [4]Biden School of Public Policy and Administration and Department of Geography and Spatial Sciences, University of Delaware, Newark, DE, USA.
*Corresponding author. Email: kmach@rsmas.miami.edu

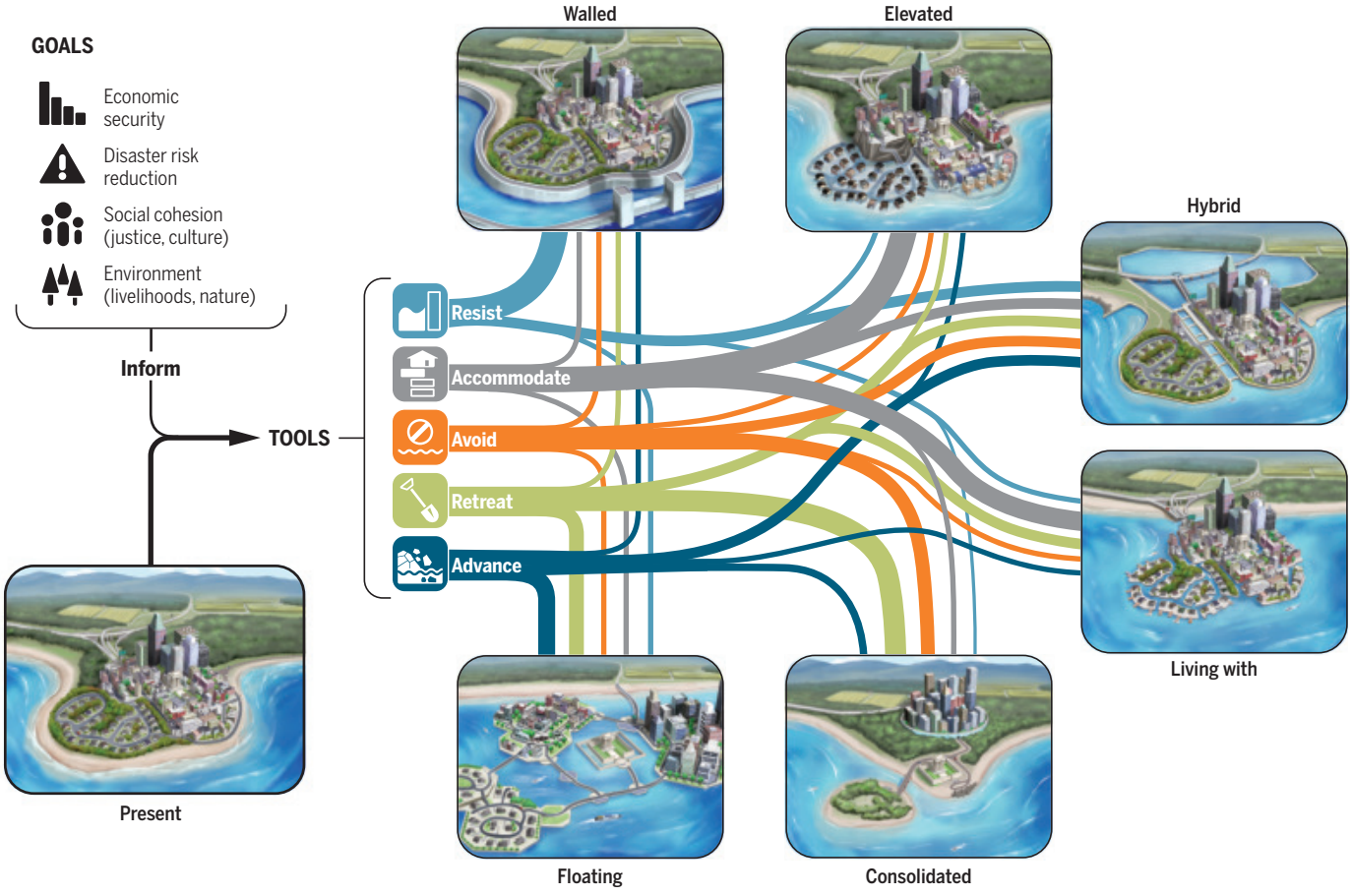


Fig. 1. Holistic pathways of climate transformation. (Bottom left) A hypothetical present-day settlement (dense city, suburban and periurban communities, and rural areas). (Right) Six possible futures. Adaptive pathways between the present day and the future will be shaped by climate risks, limits to adaptation, and societal goals. Strategic, managed retreat (green) will have some role in each future, along with other categories of response. The degree of retreat varies across scenarios (e.g., removal of a few structures to create space for retention ponds and pumping in the hybrid scenario or large-scale relocations in the consolidated or floating scenario). A decision not to engage in strategic, managed retreat complicates the pursuit of these futures and may eliminate certain futures as options.

historical human experience, the need for transformational adaptation that fundamentally alters systems will likely increase (*5*, *9*, *11*, *14*, *35*, *36*). Societies will shape how they transform and what tools they use, depending on the evolving values they prioritize (e.g., economic growth, social justice, environmental health, cultural heritage, and public safety).

Most climate-driven transformations will involve some degree of retreat in combination with measures to resist, accommodate, avoid, or advance (Fig. 1) (*19*, *22*, *35*). New or cheaper technologies may keep high waters out of coastal settlements for longer periods (*9*, *37*), enable inhabitants of wildland-urban interfaces to better live with fire threat (*38*), or support outdoor work in locations where extreme humid heat exceeds human physiological tolerances (*1*). However, even where technological solutions work and predominate—in creating floating settlements, encircling cities with storm or fire barriers, or enclosing workers in chilled suits—some retreat will occur to create room for technological adaptations or to increase their potential (Fig. 1). In other transformative futures, retreat is likely to become a core component rather than a supporting element. The extent and nature of retreat will depend on progress, breakthroughs, and constraints that shape myriad forms of adaptation (*7*).

The diversity of strategic retreat

Retreat to date has taken multiple forms, such as mandatory resettlements of communities and home buyouts post-disaster (*17*, *23*, *29*, *39*), but managed retreat can take even more forms (*14*, *22*, *28*, *30*) such as restrictions on rebuilding in hazardous locations; setbacks limiting development near areas prone to erosion, flooding, or wildfire; downzoning encouraging decreases in asset exposure over time; easements supporting ecological conservation and risk reduction; or community relocation (*40*, *41*). Empirical studies on managed retreat note the need to tailor programs to local contexts (*7*, *15*, *29*), and increasing the types of retreat drawn upon is essential for adaptation that effectively manages current, much less future, climate risks (*42*).

Practitioners are experimenting with retreat policies, and retreat in the future will likely combine social, financial, legal, engineering, and conservation innovations (Fig. 1) (*7*, *19*, *22*, *41*). Future retreat could involve more frequent or bigger applications of existing practices (*36*). For example, future strategic retreat could be extremely targeted to small, specific locations that create space for systems of retention ponds and pumps or downsize asset exposure in fire-prone locations (*7*, *14*, *38*). Practitioners and communities are already finding innovative ways to use small parcels of open land created through targeted buyouts (*43*). In some cases, strategic retreat might shift the core of a city, preserving the general location of the city while relocating neighborhoods. In others, such a rolling shift may be of limited utility, and strategic retreat might entail massive deployments, such as moving an entire city or converting roadways into lakes and canals, removing farms and shifting into aquaculture, or restricting power lines to a more limited set of fire-proofed enclosures (*36*). Managed retreat projects have helped people move as

ILLUSTRATION: E. HARTLEY

individuals by providing job training to facilitate migration, have relocated whole communities by constructing new towns nearby, and have developed new communities for people arriving from multiple locations (*30*, *44*). Programs have attempted to promote social justice by prioritizing voluntary choices and social support. They have aimed to maintain cultural heritage through relocation or documentation of cultural resources or to use newly open land effectively by establishing community gardens or endangered species habitats. Future managed retreat programs may engage these aims even more directly; for example, by addressing historic forced resettlements through reparations, preserving cultural sites, or facilitating ecosystem migration.

How retreat occurs will depend on the nature of the hazards to which it responds, their intensities, return periods, temporal and geographical trends, and the potential for other adaptation options to reduce risk (*1*, *4*, *6*, *38*). For example, wildfires affect communities through both fire and smoke, which threaten people in varying ways at differing geographic scales and with variable recurrence frequencies. Because most retreat to date has occurred in response to flooding, future retreat in response to a wider range of threats is likely to inspire novel forms and innovations (*9*, *14*, *38*, *45*). Future retreat may also increasingly result from slower-onset trends, such as continuing subsidence, recurrent high-tide flooding, permafrost melt, groundwater salinization, or desertification (*2*, *7*, *19*). Proactive retreat, planned before slow-onset changes severely threaten lives, livelihoods, and other things people value, is likely to be more effective and to reduce the psychological, sociocultural, and implementation burdens of retreat (*19*, *28*, *42*).

Coordination across silos

Retreat to date has focused on reducing risks and economic losses, responding to disaster impacts, or creating habitat—values defined by a narrow set of stakeholders and often disconnected from other societal objectives such as cultural or community cohesion, livelihoods, ecosystem health, and housing security (*21*, *29*). In the future, strategic, managed retreat must attend to a more extensive set of goals and stakeholders, from localities where retreat occurs to regional, national, and international communities involved in funding and implementation (*17*, *19*, *33*). In practice, achieving strategic retreat has been very difficult. Most managed retreat programs have lacked a holistic plan (*17*, *39*), and retreat efforts with strategic plans have been stymied by systemic implementation barriers and siloed governance systems (*23*). In rare cases such as the relocation of small towns away from flood-prone rivers or coastlines, retreat has promoted more holistic values, providing hope that strategic retreat could become more common.

Climate adaptation in unambiguous forms such as managed retreat raises multitudes of existing societal challenges, such as inadequate housing, energy access, or immigration law (*17*, *19*, *28*). For example, perhaps more than any other adaptation strategy, retreat has generated debate and discord related to equity (*7*, *14*, *21*, *28*). Categories of concerns include the locations where governments make (or do not make) managed retreat available; failures to recognize historic injustices shaping current conditions and stressors; lack of transparency, fairness, or voice; retreat processes or outcomes compromising the capabilities of people; or insufficient consideration of ecosystems (*17*, *29*, *39*, *46*, *47*). Equity considerations will be magnified as increasing climate change disproportionately affects low-income regions, marginalized populations, and future generations that have contributed little to cumulative greenhouse gas emissions and may less readily finance and pursue responses (*25*, *34*, *35*, *48*, *49*). Equitable approaches to retreat require considering social justice in both the outcomes and process of retreat: Transformations are inherently political, and without deliberate approaches, they risk exacerbating historic wrongs (*7*, *12*, *25*, *29*, *48*, *50*). Future strategic retreat could be a proactive process for addressing inequities; opening up opportunities; grappling with climate losses; and holistically supporting health, well-being, livelihoods, and security (*12*, *13*). For example, over decades and centuries, governments in many regions have displaced marginalized communities, including Indigenous peoples and racialized groups, and have failed to equitably serve them; where these communities now face immense climate risks, strategic retreat led by these communities could be a proactive process for recognition of historic injustices, reparations, and reconciliation that supports social healing. Achieving equitable, strategic retreat requires innovations, learning, and substantial deviations from the past.

Multiple levels and agencies of government, the private sector, and individuals all must reckon with the impacts and losses associated with climate change and adaptation (*7*, *13*, *24*, *42*, *45*, *51*). Retreat, as for much of adaptation, cannot be siloed (*14*, *21*, *25*), yet mainstreaming into ongoing decision-making involves pervasive, structural challenges. Sociopolitical and cultural barriers to transformative adaptation may be particularly salient for retreat because of its stark deviations from prevalent norms. Major efforts are underway to address these challenges for adaptation generally—by building capacity, increasing funding, raising awareness, and developing decision-making tools—and these efforts to facilitate equitable, transformational adaptation will make strategic, managed retreat easier (though not easy). Reforms are more likely to facilitate strategic, managed retreat if it is seen as a viable or even desirable adaptation option, rather than a fate to be avoided (*19*, *21*, *28*, *51*), but how communities and decision-makers might be inspired to reconceptualize retreat is an open question. In the following sections, we make two suggestions: (i) changing the way retreat is researched and evaluated to enable clearer comparisons to its alternatives and complements; and (ii) engaging in bold, long-term visioning of adaptation futures to help stakeholders identify which aspects of the present should be preserved and which should be actively changed, perhaps through retreat.

Adaptation science for strategic, managed retreat

Research on managed retreat has not adequately answered numerous questions that are critical in reshaping how and why retreat is considered by practitioners and communities. We argue that greater progress can be made by reframing existing questions, by asking a broader set of questions, and by engaging with a range of connected disciplines. To demonstrate, we reframe two questions often asked about retreat and identify several unanswered questions that need greater attention.

Where will climate retreat occur?

The question of where retreat will occur as a result of climate change is frequently posed. This underscores a crucial but often underappreciated distinction: Unmanaged retreat as a default that occurs when other options are absent is fundamentally different from managed retreat as multiple tools strategically deployed in pursuit of desired futures. There are at least two important reframings of the question: (i) Where will resources be allocated to prevent unmanaged retreat? and (ii) Where will strategic, managed climate retreat occur?

Adaptation resources are finite, and decisions about where to resist, accommodate, advance, and manage retreat are therefore also choices about where to allow unmanaged retreat (Fig. 1). Mapping where unmanaged retreat will occur requires numerous assumptions about the types and magnitudes of climate impacts that cause relocation, capacities to invest in alternative adaptations, future development patterns, and local goals. Estimates of sea level rise displacement by 2100 range from 88 million to 1.4 billion people, depending on whether the estimate assumes that all people in low-elevation coastal areas will be affected or only people in areas permanently inundated or historically flooded (*14*). These numbers could be substantially lower, depending on how much shoreline armoring or infrastructure elevation occurs and how effective

those strategies prove. One analysis estimated that, for 13% of the world's coastline, armoring is an economically robust adaptation strategy to minimize total costs of sea level rise, and for 65% of the global coastline, allowing unmanaged retreat minimizes total costs (*7*, *35*). The geographic disparities are notable, with shoreline armoring primarily in high-income countries and unmanaged retreat economically implied for small islands and rural coastlines. Notably, these estimates assume that retreat requires little or no public expenditure and occurs when resources are not available for other options.

Attempting to estimate where retreat will occur on the basis of where other adaptation strategies will not occur frames these approaches as alternatives when, in reality, retreat intermingles with armoring, accommodation, and advance measures (*15*, *39*). Even in places where armoring is the economically robust decision, retreat is likely to occur to create space for infrastructure or floodplains (Fig. 1). Moreover, individual and community decisions about whether to remain or retreat are not exclusively driven by risk exposure. People decide to relocate for many reasons: e.g., risk perceptions, place attachment, social norms and networks, livelihoods, institutional responses, resources and capabilities, and opportunities available elsewhere (*14*, *18*, *26*, *32*, *52*). To identify a place where retreat will or "should" occur purely on the basis of exposure to hazards or economic efficiency is to erase the lived experiences and factors central to relocation (*12*). In fact, identifying places where retreat "should" happen may inspire public resistance or investments to avoid retreat in those very places (*47*). Instead, asking where unmanaged retreat can be prevented creates agency: It highlights that unmanaged retreat can be limited through deployment of in situ adaptations and managed retreat.

The second way to reframe the question of where retreat will occur is even more difficult: "Where will strategic, managed retreat occur?" or, to be more precise, "Where and in what forms will managed retreat support futures desired by the populations involved?" Answers to this question require understanding of the full set of societal goals to be pursued and the transformative pathways that could promote them (Fig. 1). These determinants of strategic, managed retreat, both priorities and processes, cannot be readily mapped or deduced at large scales to provide probabilistic, geographically explicit estimates. Even if they could, the ways in which people relate to place, their aspirations, and the values they ascribe to mobility or stability will be dynamic through time and sometimes contradictory (*13*, *53*). Assessing future needs for managed retreat requires integration of compound-hazards climate science; studies of dynamic socioeconomic development and migration; techno-economic evaluation of engineering solutions; and analysis of sociocultural, psychological, political, institutional, and financial factors central to the viability—or lack thereof—of protecting societies in place as climate risks increase (*14*, *37*, *52*). Where retreat will occur is a beguilingly simple question that requires advances across adaptation-relevant disciplines and adaptation options.

Is retreat beneficial or harmful?

Climate-related transformations and adaptation are all about change, which involves losses and gains, sometimes for different groups (e.g., creating winners and losers) and sometimes both simultaneously for a given person, household, or community (*12*). In this regard, retreat is exceptional but not unique among adaptation approaches. Dialogues and debate about relocation, however, frequently hang on whether retreat is a benefit or a harm: whether or not it was a "success." The more powerful framing is, first, whether the benefits or harms of managed retreat are greater or lesser than for other actions and, second, how retreat can be managed to maximize benefits and minimize harms.

This reframing explicitly acknowledges that retreat intrinsically creates opportunities and losses simultaneously (*12*). Compassionate, proactive implementation of managed retreat must therefore recognize the incommensurability of some losses and the importance of engagement with them (*13*, *42*). Insights from the emerging science of climate change loss underscore the importance of recognizing the changing relationships among people and with places, the necessary and dynamic trade-offs among valued things, and the contested power relations that prioritize some people's losses over others' (*7*, *13*, *52*).

Crucially, transformation affects many intangible values—such as sense of belonging or identity, shared histories and heritage, or the intrinsic value of biodiversity—which are difficult, if not impossible, to measure and therefore generally overlooked (*52*). A situated approach to retreat reflects context and the narratives through which needs and nonsubstitutability are explored, and it is implemented by or with those who are affected (*28*, *44*, *48*, *50*). For example, through community partnerships and citizen science, threatened heritage sites can be monitored, records created, and losses acknowledged and memorialized (*54*). Intentionally loosening attachments with a given place and context, choosing elements to retain, and proactively forming new relationships in another location can reduce pain and enable people to work through loss and grief in adaptation and retreat (*12*, *13*, *51*).

The methods by which decisions to retreat are made, at individual or collective levels, are important. Having control over their destinies is essential in people's identities and relationships with place, yet further research is necessary to understand how retreat decisions are best made and how power dynamics affect how retreat conversations should be initiated—for example, with governments offering retreat or as grassroots efforts emerging from community requests (*14*, *19*, *33*, *46*, *51*). Voluntariness of retreat is often assumed to maximize benefits and minimize losses, but empirical work is needed to explore exactly what voluntariness entails. Is it enough that retreat is the best financial option available? Or does voluntariness require something more, and if more, what? Should a person suffering solastalgia [the pain of watching an environment change and degrade (*55*)] and unable to reclaim the old community be helped to move? Whether households and communities have the resources and capabilities to adapt in place, retreat on their own, or adjust to a new location may also influence perceptions and limits of voluntariness (*12*, *21*, *29*, *49*). Numerous communities are likely to require assistance in managing strategic retreat, so who should pay, and how should decision-makers prioritize allocation of resources? Research is also essential for enduring questions of how to address historical injustices that have shaped who is at risk from current and future climate-related hazards (*25*, *29*, *48*).

Concerns about retreat, as both beneficial and harmful, should be asked about every type of adaptation, especially transformational adaptations that fundamentally alter systems. Academic evaluations of managed retreat tend to consider a fuller range of harms and benefits—e.g., how retreat affects identity and connection to place, agency, equity, and property rights, leading to harms even where flood risk has been reduced (*7*, *14*, *15*, *21*, *28*, *29*, *37*)—than government agencies' evaluations of managed retreat and other adaptation options. Protective structures, such as levees, can alter place attachment when they fundamentally change a community's layout or reduce access to open spaces, but it is relatively uncommon to consider the implications of in situ adaptation for place attachment (*52*). Remaining in a place that experiences environmental degradation can cause painful emotions as relationships with that place are altered (*55*). Practitioners frequently ask how retreat can be financed, and the same question applies to how in situ adaptation approaches will be financed, who will pay, and whether the costs and benefits are equitably distributed. Comparing the answers for retreat with those for other adaptation approaches provides much more information than asking about retreat in isolation. Future research can help communities and decision-makers consider retreat more carefully by comparing opportunities and losses across the full portfolio of options, including resistance,

Fig. 2. Future visions for retreat in transformational adaptation. (Left) A floating city; (right) urban green space. These images illustrate the creativity that can be applied in imagining different relationships between people and places. We are not advocating for either of these visions, specifically. Rather, creatively imagining futures beyond the constraints of the present in equitable, participatory processes can enable new ideas, individually and collectively, about how to stay in place or how to preserve communal practices, cultural well-being, and other valued attributes even where retreat or other climate-driven transformations occur.

accommodation, advance, and avoidance as well as retreat (*15*).

Other unanswered questions

Retreat, as for most transformational adaptations, forces consideration of difficult questions that have far-reaching implications for numerous disciplines and areas of adaptation practice. How can agency and mental health be maintained under climate-driven transformations (*12*–*14*, *42*, *52*)? What engineering innovations will be needed to support infrastructure that is permanently inundated, sited on melting ground, or repeatedly exposed to wildfire? What are the economics of retreating versus staying in place, where mismatched incentives or risk perceptions lead to market failures? What are the legal implications of a country that loses its territory (*14*)? How might local jurisdictions be merged? Supporting societal deployments of strategic, managed retreat requires advances in adaptation science and should draw from the vanguard of political science, economics, migration studies, sociology, and cultural heritage studies, among other fields.

Implications for adaptation now

Most climate adaptation to date has been small scale and short term, involving minor modifications to standard practices such as planting crops earlier in the season or building floodwalls to address storms (*56*). The level of ambition and innovation in adaptation falls substantially short of the challenges posed by current, much less future, climate risks (*7*, *56*). Designing bold, innovative, and adequate adaptation strategies, including retreat and other transformational options, involves both motivating visions (i.e., future states that represent the goals and values of the people adapting) and pathways (i.e., series of steps that connect present actions with possible futures). Incorporating retreat into these visions and pathways will require widespread implementation of acknowledged best practices from adaptation science and innovations from other disciplines.

Adaptation visions

Adaptation science can tap bolder visions and pathways by implementing lessons learned from design-thinking and planning and by integrating insights from a range of social sciences and the arts. Methods such as scenario analysis and iterative, participatory deliberations geared toward decision-making under deep uncertainty (*7*, *19*, *45*, *57*) are frequently used in adaptation, although not consistently in accordance with acknowledged best practices. For example, planning activities that consider how the future may unfold too often constrain the range of futures being considered or do not represent the values of the participants. Approaches from architecture, design, environmental engineering, anthropology, archaeology, climate fiction, futurism, and security assessments, which are increasingly considering climate change, can offer rigorous, structured ways to question assumptions, generate creative ideas, and explore opportunities in future visions or scenarios (Fig. 2) (*58*–*62*). Disciplines central to social change can help people trust and use climate science in envisioning the futures they want to pursue through adaptation. Artists visualize the consequences of climate change through installations, such as lights representing future tidemarks (*60*). Climate fiction imagines dystopian and utopian visions of the future and pathways leading to beneficial change (*61*). Divides drawn between science and the arts blur when scientists consider future social conditions (e.g., the shared socioeconomic pathways) and related narratives (*62*).

Creating visions of the future to guide adaptation should not be a top-down process. Effective processes can reduce political polarization, involve marginalized groups, address historic wrongs, situate local concerns within broader contexts, highlight interactions and interdependencies within and across systems, and collaboratively construct future visions (*7*, *33*). Expanding the option set depends fundamentally on participant goals, and the process of eliciting and constructing these goals is inherently interactive. Listening across political differences or documenting a community's reality and valued attributes through photos or videos, as well as future storylines, heritage discussions, and participatory mapping, can be (and in some cases is being) used to create space for conversations about losses and damages, justice, culture, and curative and adaptive approaches for the future (*19*, *45*, *54*). Dialogue in such interactions supports open-ended consideration of what is valued and

CREDITS: (LEFT) OCEANIX/BIG-BJARKE INGELS GROUP; (RIGHT) STEFANO BOERI ARCHITETTI, VERTICAL FOREST NANJING RENDER

how creative options could preserve relationships under change and support the collective good (*13*). Iterative consideration of goals is crucial because visions for the future evolve through time, including through the process of adaptation itself (*53*).

Dynamic adaptation pathways

Articulating and pursuing visions can lead to harms as well as beneficial outcomes (*53*), and top-down or overly techno-optimistic visions, in particular, may continue colonialist, autocratic, or otherwise unjust traditions. Some of these concerns can be mitigated through participatory visioning processes. Other concerns, especially those about unintended harms, may be addressed by developing flexible pathways: steps bridging the present and future that allow for and even encourage revision (*7*, *45*, *57*). Current choices create path dependence and potential maladaptation, equally through conversations that are not happening right now as through those that are (*20*, *22*). For example, conversations about how to help shrinking settlements thrive or how to address historic injustices may be crucial for future adaptation responses, and these options may be limited because such discussions are currently inhibited by social or political risks. Key features of effective adaptation processes and decision support therefore involve: (i) flexible, diverse responses that can be adjusted through time; (ii) active monitoring and evaluation to guide necessary changes along the way; (iii) integration of local, Indigenous, scientific, and other scholarly knowledge to assess context-specific risks; and (iv) meaningful public deliberation to promote learning, consensus building, conflict resolution, and fair and just responses (*7*, *20*, *24*, *28*, *45*, *48*, *50*, *57*).

Retreat in adaptation visions and pathways

Retreat offers a valuable set of tools for both creating visions for adaptation (i.e., articulating goals and values) and navigating flexible pathways. Creative processes such as design often begin by considering wide ranges of possible options before narrowing down, encouraging participants to question assumptions and ignore limitations (*59*). Many large-scale adaptation strategies will involve managed retreat, as either a central feature or a minor but critical component, so refusing to consider managed retreat as a possibility limits the options available and the range of possible outcomes (Fig. 1). Thinking outside the box requires a willingness to embrace change, or at least to think critically about which elements of the current system should be maintained and which should be purposefully altered. Retreat—even discussion of the potential for retreat—engages some of these difficult conversations.

Our goal in authoring this article is not to suggest that managed retreat will be the optimal adaptation in any given place, but to encourage serious consideration of retreat in climate-related transformations: to better understand why remaining in place is a core value; which valued relationships should be maintained in adaptation portfolios; and what costs communities are able or willing to endure, or what opportunities they can forego, to remain in place. Early conversations about managed retreat—and where, when, and why its use could be considered acceptable (or not)—substantially increase the likelihood that future climate retreat will promote societal goals.

Adaptation visions have the potential to be bold, in pursuit of futures prepared for climate shocks that promote social justice, improve quality of life, and foster stronger relationships between peoples and between people and nature. Retreat is controversial because it challenges the status quo and has potential to cause major and inequitable losses. Conversely, retreat has enormous potential to inform adaptation precisely because it challenges the status quo, raises difficult questions about justice, forces people to confront the inevitability of change, and encourages people to make mindful choices about trade-offs and losses in the futures they pursue. Strategic, managed retreat may not be implemented in many places. Yet bringing it into adaptation conversations now, despite (or even because of) its complexities, creates better chances of long-term, sustainable well-being under intensifying climate risks.

REFERENCES AND NOTES

1. C. Raymond, T. Matthews, R. M. Horton, *Sci. Adv.* **6**, eaaw1838 (2020).
2. J. Elliott *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **111**, 3239–3244 (2014).
3. J. J. Sharples *et al*., *Clim. Change* **139**, 85–99 (2016).
4. T. Kitzberger, D. A. Falk, A. L. Westerling, T. W. Swetnam, *PLOS ONE* **12**, e0188486 (2017).
5. J. Dupuy *et al*., *Ann. For. Sci.* **77**, 35 (2020).
6. B. Neumann, A. T. Vafeidis, J. Zimmermann, R. J. Nicholls, *PLOS ONE* **10**, e0118571 (2015).
7. M. Oppenheimer *et al*., in *IPCC Special Report on the Ocean and Cryosphere in a Changing Climate*, H.-O. Pörtner *et al*., Eds. (Intergovernmental Panel on Climate Change, 2019), pp. 321–446; www.ipcc.ch/site/assets/uploads/sites/3/2019/11/08_SROCC_Ch04_FINAL.pdf.
8. M. E. Hauer, J. M. Evans, D. R. Mishra, *Nat. Clim. Chang.* **6**, 691–695 (2016).
9. C. D. Storlazzi *et al*., *Sci. Adv.* **4**, eaap9741 (2018).
10. F. V. Davenport, M. Burke, N. S. Diffenbaugh, *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2017524118 (2021).
11. K. Dow *et al*., *Nat. Clim. Chang.* **3**, 305–307 (2013).
12. K. E. McNamara, R. Bronen, N. Fernando, S. Klepp, *Clim. Policy* **18**, 111–117 (2018).
13. P. Tschakert *et al*., *WIREs Clim. Change* **8**, e476 (2017).
14. M. E. Hauer *et al*., *Nat. Rev. Earth Environ.* **1**, 28–39 (2020).
15. B. Doberstein, A. Tadgell, A. Rutledge, *J. Environ. Manage.* **253**, 109753 (2020).
16. G. Schuurman *et al*., "Resist-accept-direct (RAD)—a framework for the 21st-century natural resource manager" (Natural Resource Report NPS/NRSS/CCRP/NRR—2020/2213, National Park Service, 2020); doi:10.36967/nrr-2283597.
17. S. Greiving, J. Du, W. Puntub, *J. Extreme Events* **05**, 1850011 (2018).
18. R. McLeman, *Popul. Environ.* **39**, 319–338 (2018).
19. A. R. Siders, *One Earth* **1**, 216–225 (2019).
20. R. Anderson, K. Patsch, C. Lester, G. Griggs, *Shore Beach* **88**, 13–33 (2020).
21. C. Farbotko, O. Dun, F. Thornton, K. E. McNamara, C. McMichael, *Nat. Clim. Chang.* **10**, 702–704 (2020).
22. C. S. Dyckman, C. St. John, J. B. London, *Ocean Coast. Manage.* **102**, 212–223 (2014).
23. M. Hino, C. B. Field, K. J. Mach, *Nat. Clim. Chang.* **7**, 364–370 (2017).
24. R. K. Craig, *Harvard Environ. Law Rev.* **34**, 9–74 (2010).
25. M. Burkett, *Harvard Civil Rights-Civil Lib. Law Rev.* **53**, 445–493 (2018).
26. J. McAdam, *Am. J. Int. Law* **114**, 708–725 (2020).
27. A. Timmermann, T. Friedrich, *Nature* **538**, 92–95 (2016).
28. J. Lawrence *et al*., *Curr. Clim. Change Rep.* **6**, 66–80 (2020).
29. I. Ajibade, *Clim. Change* **157**, 299–317 (2019).
30. A. R. Siders, M. Hino, K. J. Mach, *Science* **365**, 761–763 (2019).
31. C. Kousky, *Clim. Change* **124**, 9–20 (2014).
32. W. H. Butler, R. E. Deyle, C. Mutnansky, *J. Plann. Educ. Res.* **36**, 319–332 (2016).
33. B. Burson, W. Kälin, J. McAdam, S. Weerasinghe, *Refug. Surv. Q.* **37**, 379–407 (2018).
34. J. Blythe *et al*., *Antipode* **50**, 1206–1223 (2018).
35. D. Lincke, J. Hinkel, *Glob. Environ. Change* **51**, 67–73 (2018).
36. R. W. Kates, W. R. Travis, T. J. Wilbanks, *Proc. Natl. Acad. Sci. U.S.A.* **109**, 7156–7161 (2012).
37. J. Hinkel *et al*., *Nat. Clim. Chang.* **8**, 570–578 (2018).
38. M. A. Moritz *et al*., *Nature* **515**, 58–66 (2014).
39. K. J. Mach *et al*., *Sci. Adv.* **5**, eaax8995 (2019).
40. M. Scott *et al*., *Plann. Theory Pract.* **21**, 125–154 (2020).
41. A. R. Siders, "Managed Coastal Retreat: A Legal Handbook on Shifting Development Away from Vulnerable Areas" (SSRN Scholarly Paper ID 2349461, Social Science Research Network, 2013), doi:10.2139/ssrn.2349461.
42. J. Barnett, P. Tschakert, L. Head, W. N. Adger, *Nat. Clim. Chang.* **6**, 976–978 (2016).
43. E. Zavar, R. R. Hagelman III, *Disaster Prev. Manag.* **25**, 360–374 (2016).
44. A. Forsyth, R. Peiser, *Landsc. Urban Plan.* **205**, 103957 (2021).
45. C. Raymond *et al*., *Nat. Clim. Chang.* **10**, 611–621 (2020).
46. C. Mortreux *et al*., *Glob. Environ. Change* **50**, 123–132 (2018).
47. Z. Lamb, in *Louisiana's Response to Extreme Weather: A Coastal State's Adaptation Challenges and Successes*, S. Laska, Ed. (Extreme Weather and Society Series, Springer, 2020), pp. 65–91; https://doi.org/10.1007/978-3-030-27205-0_3.
48. J. T. Malloy, C. M. Ashcraft, *Clim. Change* **160**, 1–14 (2020).
49. J. Martinich, J. Neumann, L. Ludwig, L. Jantarasami, *Mitig. Adapt. Strateg. Glob. Change* **18**, 169–185 (2013).
50. I. Scoones *et al*., *Curr. Opin. Environ. Sustain.* **42**, 65–75 (2020).
51. J. McAdam, E. Ferris, *Camb. Int. Law J.* **4**, 137–166 (2015).
52. D. Clarke, C. Murphy, I. Lorenzoni, *J. Environ. Psychol.* **55**, 81–89 (2018).
53. B. Taebi, J. H. Kwakkel, C. Kermisch, *WIREs Clim. Change* **11**, e666 (2020).
54. T. Dawson, J. Hambly, A. Kelley, W. Lees, S. Miller, *Proc. Natl. Acad. Sci. U.S.A.* **117**, 8280–8286 (2020).
55. G. A. Albrecht, *Am. Imago* **77**, 9–30 (2020).
56. United Nations Environment Programme, "Adaptation Gap Report 2020" (United Nations Environment Programme, 2021); www.unenvironment.org/resources/adaptation-gap-report-2020.
57. V. A. W. J. Marchau, W. E. Walker, P. J. T. M. Bloemen, S. W. Popper, Eds., *Decision Making Under Deep Uncertainty: From Theory to Practice* (Springer, 2019); https://doi.org/10.1007/978-3-030-05252-2_4.
58. J. G. Carter *et al*., *Prog. Plann.* **95**, 1–66 (2015).
59. R. Razzouk, V. Shute, *Rev. Educ. Res.* **82**, 330–348 (2012).
60. P. Niittyvirta, Lines (57° 59′ N, 7° 16′W). *Pekka Niittyvirta*; https://niittyvirta.com/lines-57-59-n-7-16w/.
61. M. Schneider-Mayerson, *Environ. Humanit.* **10**, 473–500 (2018).
62. J. Hinkel, D. Mangalagiu, A. Bisaro, J. D. Tàbara, *Clim. Change* **160**, 495–506 (2020).

ACKNOWLEDGMENTS

We thank E. Hartley for graphical illustration and J. Niemann for formatting of references. **Funding:** The University of Miami Rosenstiel School of Marine and Atmospheric Science and Leonard and Jayne Abess Center for Ecosystem Science and Policy provided funding for this work. **Competing interests:** The authors declare no conflicts of interest.

10.1126/science.abh1894

RESEARCH

IN *SCIENCE* JOURNALS

Edited by **Michael Funk**

OPTOMECHANICS

Really cool mirrors

Cooling objects to low temperature can increase the sensitivity of sensors and the operational performance of most devices. Removing most of the thermal vibrations—or phonons—such that the object reaches its motional quantum ground state has been achieved but typically with tiny, nanoscale objects. Using the suspended mirrors of the Laser Interferometer Gravitational-Wave Observatory (LIGO) that form a 10-kg optomechanical oscillator, Whittle *et al*. demonstrate the ability to cool such a large-scale object to nearly the motional ground state. An upgrade to LIGO with such a modification could increase its sensitivity and range to gravitational waves but also extend studies of quantum mechanics to large-scale objects. —ISO

Science, abh2634, this issue p. 1333

A technician's image is reflected while inspecting a set of mirrors at LIGO.

PHOTO: LIGO/CALTECH/MIT

SIGNAL TRANSDUCTION

Timing cues epigenomic reprogramming

Different temporal dynamics of activation of the transcription factor nuclear factor κB (NF-κB) can influence the inflammatory response of activated macrophages. Cheng *et al*. report a mechanism by which oscillatory and sustained NF-κB signaling may produce distinct transcriptional responses (see the Perspective by Nandagopal *et al*.). Oscillatory activation of NF-κB activated poised enhancers to transcribe inflammatory genes in mouse macrophages. However, sustained activation of NF-κB produced in cells activated by other stimuli acted on the epigenome. These stimuli relieved chromatin silencing at enhancers and enabled regulation of additional genes. —LBR

Science, abc0269, this issue p. 1349; see also abj2040, p. 1263

SURFACE CHEMISTRY

Imaging reactions across facets

Metal nanoparticles used in heterogeneous catalysis can bear different facets with different reaction kinetics. Suchorski *et al*. used field electron microscopy with high spatial (~2 nanometers) and time (~2 milliseconds) resolution to study hydrogen oxidation on a curved rhodium crystal that displayed individual nanofacets. They also performed field ion microscopy of the water products. Periodic formation and depletion of subsurface oxygen blocked or allowed hydrogen adsorption, respectively, and led to oscillatory kinetics that could frequency lock between facets but at different frequencies. Surface reconstructions could also induce collapse of spatial coupling of oscillations. —PDS

Science, abf8107, this issue p. 1314

EPIDEMIOLOGY

Improved disease tracking with AI

Powerful machine-learning tools have the potential to improve real-time estimates of disease activity, such as for COVID-19 or influenza. Aiken *et al*. developed such a method that produces dynamic weekly and city-level forecasts of disease cases and is more accurate than existing techniques, including those that use real-time data from web searches. The researchers' neural network integrates disease activity data across cities and time to produce highly

accurate forecasts of activity up to 8 weeks in the future. These findings illustrate the potential of modern artificial intelligence (AI) to better track and predict local epidemic dynamics up to several weeks ahead of current health care–based surveillance systems, which would support better public health policy and individual decisions. —AJC

Sci. Adv. 10.1126/sciadv.abb1237 (2021).

GENE THERAPY

Gene correction for SCD stem cells

Sickle cell disease (SCD) is an autosomal-recessive disease resulting from a point mutation in the β-globin gene that leads to sickle-shaped red blood cells, pain crises, and decreased life span. Lattanzi *et al.* studied ex vivo β-globin gene correction in autologous patient-derived hematopoietic stem and progenitor cells (HSPCs) as a potential cure for SCD. The authors demonstrated 20% gene correction after transplantation of corrected HSPCs into immunodeficient mice, with no evidence of genotoxicity or tumorigenicity. The gene-corrected HSPCs could also be reliably produced at scale. These studies lay the groundwork for a clinical trial of this gene correction strategy in patients with SCD. —MN

Sci. Transl. Med. **13**, eabf2444 (2021).

CORONAVIRUS

Shifting frames to make more proteins

Severe acute respiratory syndrome coronavirus 2 critically depends on the ribosomal frameshifting that occurs between two large open reading frames in its genomic RNA for expression of viral replicase. Programmed frameshifting occurs during translation, when the ribosome encounters a stimulatory pseudoknot RNA fold. Using a combination of cryo–electron microscopy and biochemistry, Bhatt *et al*. revealed that the pseudoknot resists unfolding as it lodges at the entry of the ribosomal messenger RNA channel. This causes back slippage of the viral RNA, resulting in a minus-1 shift of the reading frame of translation. A partially folded nascent viral polyprotein forms specific interactions inside the ribosomal tunnel that can influence the efficiency of frameshifting. —DJ

Science, abf3546, this issue p. 1306

QUANTUM SIMULATION

A spinning quantum gas

Ultracold atomic gases are very good at simulating electrons in solids but lack one essential party trick: charge. Their neutrality makes it challenging to simulate phenomena such as the quantum Hall effect, which, in the case of charged electrons, is easily induced by an external magnetic field. One way to produce a similar effect in a neutral system is to rotate it, but achieving the equivalent of strong magnetic fields remains difficult. Fletcher *et al*. rotated a gas of trapped sodium atoms, reaching a state in which the gas could be described by a single lowest Landau-level wavefunction. The system is expected to be a testbed for studying the behavior of strongly interacting many-body states. —JS

Science, aba7202, this issue p. 1318

MAGNETISM

Capturing exotic magnetism

Ferromagnetism is associated with the breaking of time-reversal symmetry, most frequently by the spin degree of freedom. Although the orbital motion of electrons can also contribute to ferromagnetism, in most materials, it is small relative to the spin contribution. Tschirhart *et al*. showed that the reverse is true in an unusual magnetic state hosted by twisted bilayer graphene. Their scanning magnetometry measurements were consistent with ferromagnetism of predominantly orbital origin. —JS

Science, abd3190, this issue p. 1323

IN OTHER JOURNALS

Edited by **Caroline Ash** *and* **Jesse Smith**

ECOLOGY

Persistent seeds predict invasiveness

Understanding the biological traits that predict the potential of an organism to colonize and occupy new territory is key to ameliorating and combating the problem of alien species invasion, a problem that has been exacerbated by global networks of travel and trade. In an analysis of seed trait data from 2350 angiosperm species, Gioria *et al*. found that the ability of seeds to form a persistent seed bank in the soil (where persistence was defined as >1 year) was strongly associated with potential to become naturalized or invasive. Adding seed persistence to the suite of plant traits that predict invasiveness may be beneficial to efforts to prevent or eradicate new introductions before they are beyond control. —AMS

Ecol. Lett. 10.1111/ele.13783 (2021).

A germinating sunflower seed

PHOTO: NIGEL CATTLIN/SCIENCE SOURCE

VACCINES

Cytokine vaccine for asthma

The cytokines interleukin-4 (IL-4) and IL-13 play important, nonredundant roles in the pathogenesis of asthma. Monoclonal antibodies against IL-4 receptor α (IL-4Rα), which inhibit both IL-4 and IL-13 signaling, can effectively ameliorate severe asthmatic exacerbations. However, monoclonal antibody therapy is expensive and requires regular injections. Conde *et al.* developed conjugate vaccines ("kinoids") against IL-4 and IL-13 by coupling these cytokines with a nontoxic mutant of diptheria toxin. These kinoids were effective as both prophylactics and therapeutics in mouse models of allergic airway inflammation, durably reducing hallmarks of disease including serum immunoglobulin E, airway hyperresponsiveness, and mucus hypersecretion. Similar results were obtained with transgenic mice expressing human IL-4, IL-13, and IL-4Rα. Thus, dual IL-4/IL-13 vaccination may be a long-lasting and economical approach to the treatment of asthma and other allergic diseases. —STS

Nat. Commun. **12**, 2574 (2021).

FOREST ECOLOGY

Losing trees despite the forest

When we think of human impacts on forests, we usually think of whole-scale destruction from logging or other habitat destruction. However, loss of individual trees within a persistent forest can have unforeseen effects. Such losses occur through drivers such as disease and drought and can occur within a specific tree species or across species. Fleming *et al.* conducted a large meta-analysis of mostly North American and Australian studies and found both positive and negative impacts on forest animal abundance. Species that rely on specific living tree resources such as nectar and pollen declined, whereas those that rely on resources such as tree cavities, open canopies, and ground resources increased. Impacts also shifted over time, with individual tree deaths leading to increases that eventually shifted to overall decline. Although sometimes occurring in opposite directions, species responses were substantial enough that increased tree loss could lead to altered forest communities. —SNV *Biol. Rev. Camb. Philos. Soc.* 10.1111/brv.12725 (2021).

A fallen dead and rotted tree in the Białowieża Forest

SOLAR CELLS

Efficient organic solar cells

There is currently considerable interest in the development of organic solar cells (OSCs) because of the emergence of so-called non-fullerene acceptors, which has pushed the power conversion efficiencies of OSCs close to those of inorganic and hybrid semiconductor solar cells. Chen *et al.* combined experimental and theoretical study of more than 30 non-fullerene OSC devices to demonstrate the key electronic structure parameters of the donor-acceptor blends, establish the relationships between these parameters and the nonradiative voltage losses, and show that these losses can be reduced without sacrificing the charge-generation efficiencies. This work opens the way for the future rational design of high-efficiency OSC materials. —YS

Nat. Energy 10.1038/s41560-021-00843-4 (2021).

METABOLISM

Innervating the gut

Sensory neurons that detect signals in the gut and relay them to the brain are an important feature of regulating food intake and metabolic responses to food. When such gut-brain signaling is impaired, overeating, weight gain, and metabolic dysfunction can occur. Borgmann *et al.* investigated feeding and glucose regulation by different sensory neurons that innervate the gut of mice. Their findings enabled reconstruction of peripheral and central sensory neuron projections and revealed differential innervation of the mucosal and muscular layers of the intestine and different regions of the gastrointestinal tract, likely reflecting distinct responses to specific signals in the gut. These subsets of neurons may provide targets for controlling metabolic diseases in the future. —GKA

Cell Metab. 10.1016/j.cmet.2021.05.002 (2021).

BIOMATERIALS

Magnetic guidance for nerve repair

After an injury, peripheral nerve regeneration is possible but often fails to fully restore function, in part because of the need for slow-moving axons to traverse long distances. Although adult stem cell therapies have shown some promise, there is a challenge in getting enough cells to the injury. Soto *et al.* enhanced adipose-derived mesenchymal stem cell delivery by loading the cells with citric acid–coated superparamagnetic iron oxide nanoparticles. Tests were done in rats using a Wallerian degeneration model of the sciatic nerve. By magnetically guiding the cells to the injury site, they were able to improve recovery, with partial conservation of the nerve structure and indications of remyelination. —MSL

Acta Biomater. 10.1016/j.actbio.2021.05.050 (2021).

COSMOLOGY

Sounding out the Universe

Propagation of sound waves in the early Universe imprinted characteristic density fluctuations known as baryon acoustic oscillation (BAO). Galaxies preferentially form in higher-density regions, so BAO can be measured using galaxy redshift surveys. Alam *et al.* report the final cosmological parameters from the extended Baryon Oscillation Spectroscopic Survey (eBOSS) both independently of other datasets and in combination with alternative cosmological techniques. The authors found strong support for standard cosmology with cold dark matter, a flat Universe, and dark energy described by a cosmological constant. The only notable inconsistency was the well-known tension in measurements of the Hubble constant, which persists in the new data. —KTS

Phys. Rev. D **103**, 083533 (2021).

PHOTO: ISTOCK.COM/APEYRON

RESEARCH ARTICLE SUMMARY

IMMUNOLOGY

Neutrophils self-limit swarming to contain bacterial growth in vivo

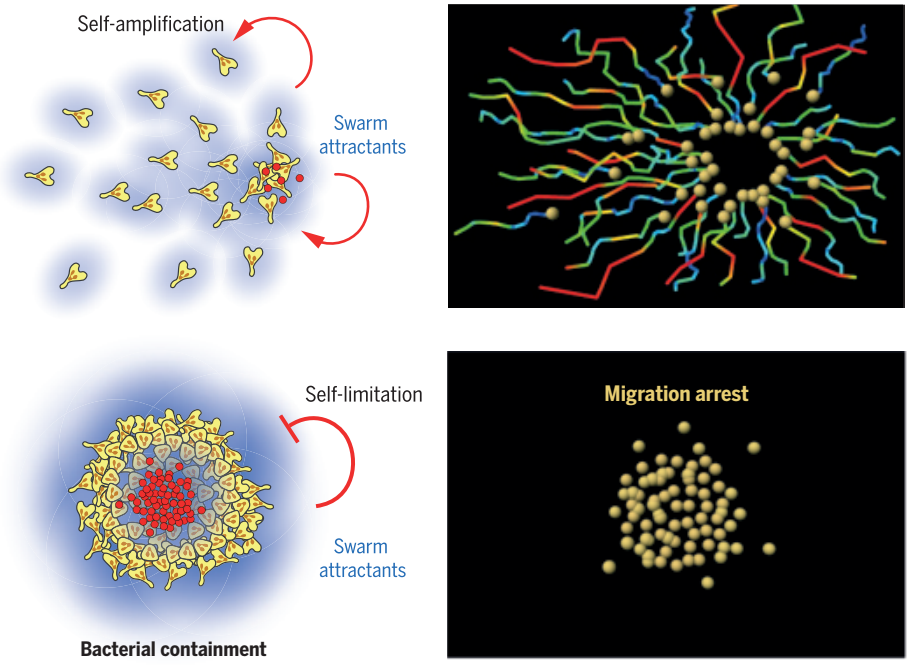
Korbinian Kienle, Katharina M. Glaser, Sarah Eickhoff, Michael Mihlan, Konrad Knöpper, Eduardo Reátegui, Maximilian W. Epple, Matthias Gunzer, Ralf Baumeister, Teresa K. Tarrant, Ronald N. Germain, Daniel Irimia, Wolfgang Kastenmüller, Tim Lämmermann*

INTRODUCTION: The collective behavior of cells and insects often relies on self-organizing processes. By releasing attractant signals, a few individuals can initiate the accumulation and aggregation of a whole population. Neutrophils, key players in the innate immune response, infiltrate inflamed and infected tissues in large numbers. These cells make use of such positive feedback amplification to find and kill bacteria in tissues. By secreting attractants that act through cell surface–expressed G protein–coupled receptors (GPCRs) on neighboring cells, neutrophils use this form of intercellular communication and coordinate their hunt for pathogens as a swarm. How this swarming response is terminated to avoid uncontrolled neutrophil accumulations and prevent excessive inflammation is currently unknown.

RATIONALE: The stop signals for neutrophil swarming in mammalian tissues have not yet been defined. They may be derived from cells of the surrounding inflammatory environment or from neutrophils themselves. We reasoned that the attractants released by neutrophils may become highly concentrated at sites where these cells cluster in larger numbers. It is well established that high chemoattractant concentrations can attenuate cellular responses by a process termed GPCR desensitization. We hypothesized a self-limiting mechanism for swarming: The local accumulation of the same neutrophil-expressed attractants that amplify swarming during early stages would cause desensitization of their respective GPCRs at later stages of neutrophil clustering. This led us to investigate the role of GPCR desensitization in neutrophil tissue navigation and host defense.

RESULTS: We generated mouse strains whose neutrophils were deficient in GPCR kinases (GRKs), critical enzymes for mediating the GPCR desensitization process. Of the four GRK isoforms tested, in vitro experiments identified GRK2 as the kinase necessary to desensitize GPCRs activated by swarm-released attractants (LTB4 and CXCL2). When neutrophils sense high concentrations of swarm attractants in vitro, GRK2 desensitizes the corresponding receptors to induce migration arrest. Two-photon intravital imaging of injured skin and infected lymph nodes of mice showed that GRK2 and GPCR desensitization play critical roles during neutrophil swarming in physiological tissue. At sites where swarming neutrophils accumulate and self-generate local fields of high swarm attractant concentration, GPCR desensitization was crucial to stop neutrophil migration arrest. Desensitization-resistant neutrophils moved faster and explored larger areas of lymph node tissue infected with the bacterium *Pseudomonas aeruginosa*. Such behavior suggested more effective bacterial sampling throughout the infected organ. Surprisingly, mice with GRK2-deficient neutrophils showed impaired rather than improved bacterial clearance. This finding could not be explained by altered antibacterial effector functions. In vitro assays for the detailed analysis of swarming behavior and bacterial growth revealed that GPCR desensitization to swarm attractants is required to induce neutrophil arrest for optimal bacterial phagocytosis and containment in swarm clusters.

CONCLUSION: We describe a cell-intrinsic stop mechanism for the self-organization of neutrophil collectives in infected tissues, which is based on sensing the local accumulation of the same cell-secreted attractants that amplify swarming during early stages. GPCR desensitization acts as a negative feedback control mechanism to stop neutrophil migration in swarm aggregates. This navigation mechanism allows neutrophils to self-limit their dynamics within forming swarms and ensures optimal elimination of bacteria. Desensitization to a self-produced activation signal as a principle of self-organization is important for immune host defense against bacteria, and likely informs other categories of collective behavior in cells and insects. ■



Self-organization of neutrophil swarms. Top: Swarming neutrophils self-amplify their highly chemotactic recruitment toward sites of tissue injury or bacterial invasion by releasing attractants that act on neighboring neutrophils. Neutrophils are displayed as spheres with migration tracks (right). Bottom: The local accumulation of the same cell-secreted attractants stops neutrophils when they accumulate and form clusters, a process important for the containment of bacteria in infected tissues.

The list of author affiliations is available in the full article online.
*Corresponding author. Email: laemmermann@ie-freiburg.mpg.de
Cite this article as K. Kienle *et al.*, *Science* **372**, eabe7729 (2021). DOI: 10.1126/science.abe7729

RESEARCH ARTICLE SUMMARY

NEUROSCIENCE

Common genetic variation influencing human white matter microstructure

Bingxin Zhao†, Tengfei Li†, Yue Yang, Xifeng Wang, Tianyou Luo, Yue Shan, Ziliang Zhu, Di Xiong, Mads E. Hauberg, Jaroslav Bendl, John F. Fullard, Panagiotis Roussos, Yun Li, Jason L. Stein, Hongtu Zhu*

INTRODUCTION: White matter in the human brain serves a critical role in organizing distributed neural networks. Diffusion magnetic resonance imaging (dMRI) has enabled the study of white matter in vivo, showing that interindividual variations in white matter microstructure are associated with a wide variety of clinical outcomes. Although white matter differences in general population cohorts are known to be heritable, few common genetic variants influencing white matter microstructure have been identified.

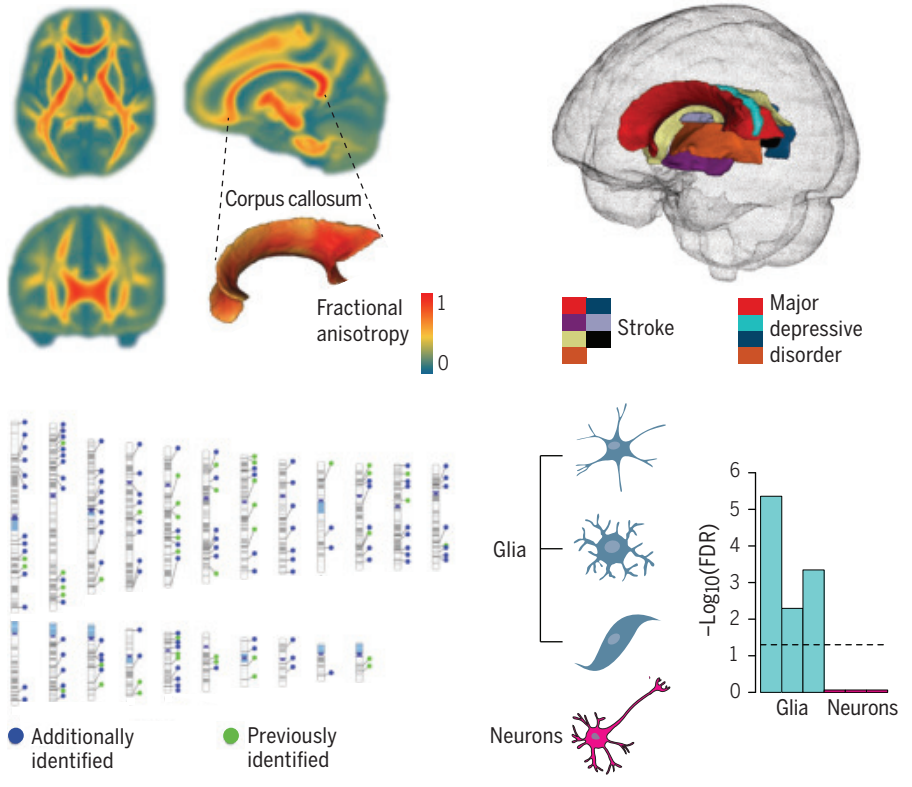
RATIONALE: To identify genetic variants influencing white matter microstructure, we conducted a genome-wide association study (GWAS) of dMRI data from 43,802 individuals across five data resources. We analyzed five major diffusion tensor imaging (DTI) model–derived parameters along 21 cerebral white matter tracts.

RESULTS: In the discovery GWAS with 34,024 individuals of British ancestry, we replicated 42 of the 44 genomic regions discovered in the largest previous GWAS and identified 109 additional regions associated with white matter microstructure ($P < 2.3 \times 10^{-10}$, adjusted for the number of phenotypes studied). These results indicate strong polygenic influences on white matter microstructure. Of the 151 regions, 52 passed the Bonferroni significance level ($P < 5 \times 10^{-5}$) in our analysis of nine independent validation datasets, including four with subjects of non-European ancestry.

On average, common genetic variants explained 41% (standard error = 2%) of the variation in white matter microstructure. The 151 identified genomic regions can explain 32.3% of heritability for white matter microstructure, whereas the 44 previously identified genomic regions can only explain 11.7% of heritability. As a biological validation of our GWAS findings, we observed heritability enrichment within regulatory elements active in oligodendrocytes and other glia, whereas no enrichment was observed in neurons. These results are expected and suggest that genetic variation leads to changes in white matter microstructure by affecting gene regulation in glia.

We observed genetic correlations and colocalizations of white matter microstructure with a wide range of brain-related complex traits and diseases, such as cognitive functions, cardiovascular risk factors, as well as various neurological and psychiatric diseases. For example, of the 25 reported genetic risk regions of glioma, 11 were also associated with white matter microstructure, which illustrates the close genetic relationship between glioma and white matter integrity. Additionally, we found that 14 white matter microstructure–associated genes ($P < 1.2 \times 10^{-8}$) were targets for 79 commonly used nervous system drugs, such as antipsychotics, antidepressants, anticonvulsants, and drugs for Parkinson's disease and dementia.

CONCLUSION: This large-scale study of dMRI scans from 43,802 subjects improved our understanding of the highly polygenic genetic architecture of human brain white matter tracts. We identified 151 genomic regions associated with white matter microstructure. The GWAS findings were supported by enrichments within cell types that make up white matter microstructure. Moreover, we uncovered genetic relationships between white matter and various clinical endpoints, such as stroke, major depressive disorder, schizophrenia, and attention deficit hyperactivity disorder. The targets of many drugs commonly used for disabling cognitive disorders have genetic associations with white matter, which suggests that the neuropharmacology of many disorders can potentially be improved by studying how these medications work in the brain white matter. ■



Identifying genetic variants influencing human brain white matter microstructure. (Top left) Quantifying the microstructure in white matter tracts using DTI models. (Bottom left) Genomic locations of common genetic variants associated with white matter microstructure. (Top right) Selected genetic correlations between white matter microstructure and brain disorders (stroke and major depressive disorder). (Bottom right) Partitioned heritability enrichment analysis in brain cell types. FDR, false discovery rate.

The list of author affiliations is available in the full article online.
*Corresponding author. Email: htzhu@email.unc.edu
†These authors contributed equally to this work.
Cite this article as B. Zhao *et al.*, *Science* **372**, eabf3736 (2021). DOI: 10.1126/science.abf3736

READ THE FULL ARTICLE AT
https://doi.org/10.1126/science.abf3736

RESEARCH ARTICLE SUMMARY

ECHOLOCATION

Echolocation in soft-furred tree mice

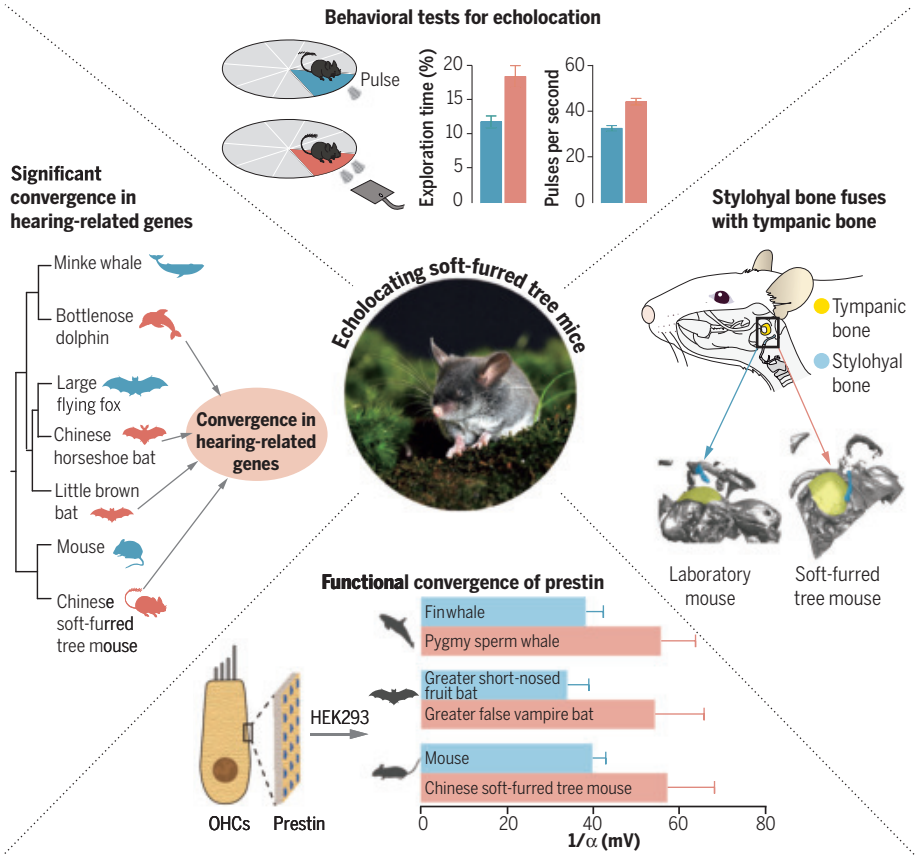
Kai He†, Qi Liu†, Dong-Ming Xu†, Fei-Yan Qi†, Jing Bai, Shui-Wang He, Peng Chen, Xin Zhou, Wan-Zhi Cai, Zhong-Zheng Chen, Zhen Liu*, Xue-Long Jiang*, Peng Shi*

INTRODUCTION: Echolocation is a form of orientation behavior in which some animals can assess environments in which vision is ineffective. Echolocating mammals have been recognized for decades, including microbats and toothed whales. Recently, the Chapa soft-furred tree mouse (*Typhlomys chapensis*) of the rodent family Platacanthomyidae was suggested to echolocate but this was not practically evidenced. There are three other recognized species in the soft-furred tree mice genus (*Typhlomys*) that share similar ecological and morphological traits, suggesting that echolocation may be a general trait within this genus. In this study, we performed behavioral, morphological, genomic, and functional analyses to test whether echolocation is present across the four species of soft-furred tree mice.

RATIONALE: Echolocation occurs when emitted sonic signals are compared with received signals to facilitate orientation and object identification. This adaptive trait is expected to be reflected at the behavioral, morphological, and molecular levels of the organism. To test whether soft-furred tree mice generally have evolved echolocation, we conducted multiple behavioral experiments to assess the performance of different species in detecting and avoiding obstacles dependent on hearing and examined the anatomical structures of their vocal and hearing apparatus. We further investigated genome-wide convergence in hearing-related genes and the functional convergence of a well-documented echolocation-related gene, *prestin*, between the soft-furred tree mouse and other known echolocating mammals.

RESULTS: All four recognized soft-furred tree mouse species were capable of emitting regular ultrasonic vocalizations (USVs) with a peak frequency of ~98 kHz. When these mice explored in a cluttered environment and approached an obstacle, they produced USVs with a significantly larger pulse rate. In tests for echolocation, the soft-furred tree mice spent a longer time exploring and emitted more sonic pulses in the sector of the central disk over the escaping platform, consequently dropping to the platform. When their ears were plugged, they lost the preference for the over-platform sector, as shown by exploration time and emitting pluses, and could not land on the platform. The above preference was regained when the earplugs were removed or a plastic tube was inserted into the ear canals. The stylohyal bone fuses with the tympanic bone in soft-furred tree mice, which is an anatomic characteristic previously exclusively observed in laryngeally echolocating bats. By sequencing a high-quality genome of the soft-furred tree mouse, we found a significant genome-wide convergence in hearing-related genes with other echolocating mammals, including the well-documented echolocation-related gene *prestin*. In vitro experimental analyses also showed a functional convergence of prestin between the soft-furred tree mouse and other echolocating mammals, which was largely contributed by the identified convergent amino acids.

CONCLUSION: Our findings from behavioral experiments, anatomical structures, evolutionary genomics, and gene functional analyses provide strong evidence that soft-furred tree mice are a new echolocating lineage within mammals. The discovery of this echolocating rodent genus suggests that echolocation may be an underappreciated trait in mammals. The genome-wide convergent evolution in hearing-related genes implicates a similar molecular mechanism underlying the origination or elaboration of this complex adaptive phenotype. ■



Multiple lines of evidence for echolocation in soft-furred tree mice (*Typhlomys*). When the soft-furred tree mice explored on a central disk, they spent more time exploring and emitted ultrasonic pulses at a higher rate in the sector over the escaping platform. The stylohyal bone fuses with the tympanic bone in soft-furred tree mice, which was previously observed exclusively in laryngeally echolocating bats. Significant molecular convergence in hearing-related genes and functional convergence of the well-documented echolocation-related gene *prestin* were found between the soft-furred tree mouse and known echolocating mammals.

The list of author affiliations is available in the full article online.
*Corresponding author. Email: ship@mail.kiz.ac.cn (P.S.); jiangxl@mail.kiz.ac.cn (X.-L.J.); zhenliu@mail.kiz.ac.cn (Z.L.)
†These authors contributed equally to this work.
Cite this article as K. He *et al.*, *Science* 372, eaay1513 (2021).
DOI: 10.1126/science.aay1513

READ THE FULL ARTICLE AT
https://doi.org/10.1126/science.aay1513

RESEARCH ARTICLE

CORONAVIRUS

Structural basis of ribosomal frameshifting during translation of the SARS-CoV-2 RNA genome

Pramod R. Bhatt[1,2,3]†, Alain Scaiola[1]†, Gary Loughran[2], Marc Leibundgut[1], Annika Kratzel[4,5,6], Romane Meurs[7], René Dreos[7], Kate M. O'Connor[2], Angus McMillan[8], Jeffrey W. Bode[8], Volker Thiel[4,5], David Gatfield[7], John F. Atkins[2,3,9]*, Nenad Ban[1]*

Programmed ribosomal frameshifting is a key event during translation of the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) RNA genome that allows synthesis of the viral RNA-dependent RNA polymerase and downstream proteins. Here, we present the cryo–electron microscopy structure of a translating mammalian ribosome primed for frameshifting on the viral RNA. The viral RNA adopts a pseudoknot structure that lodges at the entry to the ribosomal messenger RNA (mRNA) channel to generate tension in the mRNA and promote frameshifting, whereas the nascent viral polyprotein forms distinct interactions with the ribosomal tunnel. Biochemical experiments validate the structural observations and reveal mechanistic and regulatory features that influence frameshifting efficiency. Finally, we compare compounds previously shown to reduce frameshifting with respect to their ability to inhibit SARS-CoV-2 replication, establishing coronavirus frameshifting as a target for antiviral intervention.

Ribosomal frameshifting, a process during which the reading frame of translation is changed at the junction between open reading frames (ORFs) 1a and 1b, is one of the key events during translation of the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) positive-sense single-stranded RNA genome. This programmed –1 translational frameshifting is conserved in all coronaviruses and is necessary for the synthesis of viral RNA-dependent RNA polymerase (RdRp or Nsp12) and downstream viral nonstructural proteins that encode core enzymatic functions involved in capping of viral RNA, RNA modification and processing, and RNA proofreading (*1*). Although the translational machinery typically prevents frameshifting as a potential source of one of the most disruptive errors in translation (*2*, *3*), many viruses rely on programmed ribosomal frameshifting to expand and fine-tune the repertoire and stoichiometry of expressed proteins (*4*).

[1]Department of Biology, Institute of Molecular Biology and Biophysics, ETH Zurich, Zurich, Switzerland. [2]School of Biochemistry and Cell Biology, University College Cork, Cork T12 XF62, Ireland. [3]School of Microbiology, University College Cork, Cork T12 K8AF, Ireland. [4]Institute of Virology and Immunology, University of Bern, Bern, Switzerland. [5]Department of Infectious Diseases and Pathobiology, Vetsuisse Faculty, University of Bern, Bern, Switzerland. [6]Graduate School for Cellular and Biomedical Sciences, University of Bern, Bern, Switzerland. [7]Center for Integrative Genomics, Génopode, University of Lausanne, 1015 Lausanne, Switzerland. [8]Laboratorium für Organische Chemie, Department of Chemistry and Applied Biosciences, ETH Zurich, Zurich, Switzerland. [9]MRC Laboratory of Molecular Biology, Cambridge CB2 0QH, UK.
***Corresponding author. Email: j.atkins@ucc.ie (J.F.A.); ban@mol.biol.ethz.ch (N.B.)**
†These authors contributed equally to this work.

Programmed –1 frameshifting in SARS-related coronaviruses occurs at the slippery sequence U_UUA_AAC in the context of a 3′ stimulatory RNA sequence that was predicted to form a three-stemmed pseudoknot structure (*5*) and, in parallel, was independently tested by our lab and others (*6*–*8*). The frameshifting occurs with high efficiency (25 to 75%), depending on the system used (*6*, *7*, *9*–*11*), and changes the reading frame to UUU_AAA_C (*12*) (Fig. 1A). Consequently, two viral polyproteins are synthesized: one encoded by ORF1a when frameshifting does not take place, and ORF1ab, which is expressed as a result of frameshifting. Translation of ORF1a produces polyprotein 1a, which ends with Nsp10 followed by the short Nsp11. Conversely, when the frameshift occurs, the polyprotein 1ab is generated, which contains almost 2700 additional amino acids and in which the viral RdRp, Nsp12, is produced after Nsp10 as a consequence of translation in the –1 frame. A putative secondary structure element in the viral RNA that forms a loop upstream of the shift site has been proposed to play an attenuating role in frameshifting and is referred to as the 5′ attenuator loop (*8*). Maintaining the precise level of coronavirus frameshifting efficiency is crucial for viral infectivity, as evidenced by the fact that mutation of a single nucleotide in the frameshifting region of the SARS-CoV-1 RNA results in a concomitant abrogation of viral replication (*13*). Therefore, the importance of three-stemmed pseudoknot-dependent –1 ribosomal frameshifting for the propagation of SARS-related coronaviruses, a process that has not been seen to occur on any endogenous human transcript in human cells, presents itself as an opportune drug target with minimal tolerance for drug-resistant mutations.

Because of its importance in the life cycle of many important viruses and coronaviruses in particular, programmed frameshifting has been extensively studied using a range of structural and functional approaches (*4*). The structure of a 3′ stimulatory pseudoknot in isolation or in context of the viral genome has been proposed recently by various groups using techniques that include molecular dynamics, nuclease mapping, in vivo selective 2′-hydroxyl acylation analyzed by primer extension (SHAPE), nuclear magnetic resonance (NMR), and cryo–electron microscopy (cryo-EM) (*7*, *14*–*17*). Furthermore, a ribosomal complex with a frameshift stimulatory pseudoknot from the avian infectious bronchitis virus was reported at low resolution (*18*). Here, to provide a structural and mechanistic description of the events during ribosomal frameshifting, we investigated mammalian ribosomes captured in distinct functional states during translation of a region of SARS-CoV-2 genomic RNA where –1 programmed frameshifting occurs.

Structure determination of a frameshifting-primed ribosomal complex

We captured a 0 frame, preframeshift ribosomal complex by introducing a stop codon in place of the second codon of the slippery site (U_UUA_AAC to U_UUA_UAA) (Fig. 1A) and adding mutant eukaryotic release factor 1 [eRF1 (AAQ)] that is unable to release the nascent polypeptide. Translating complexes were prepared in an in vitro translation reaction using an in-house–generated rabbit reticulocyte lysate (RRL) system that supported efficient frameshifting in the previously reported range of around 50% (*19*) according to dual luciferase experiments (see methods). The ribosomes were programmed with mRNA encoding an affinity tag and harboring a region of the SARS-CoV-2 genome that encodes proteins Nsp10 (C terminus), Nsp11, and most of Nsp12. Western blotting showed that when using the wild-type (WT) RNA template, frameshifting was efficient, whereas the stop codon mutation prevented frameshifting and led to ribosome pausing. This effect was further enhanced when eRF1 (AAQ) was present in excess over endogenous WT eRF1 (Fig. 1B).

The cryo-EM three-dimensional (3D) reconstruction of ribosome–nascent chain complexes affinity-purified from the reactions supplemented with eRF1 (AAQ) revealed two distinct ribosomal complexes captured in the process of translating the slippery sequence (figs. S1 and S2). One represented a termination complex that contained the ATP-binding cassette

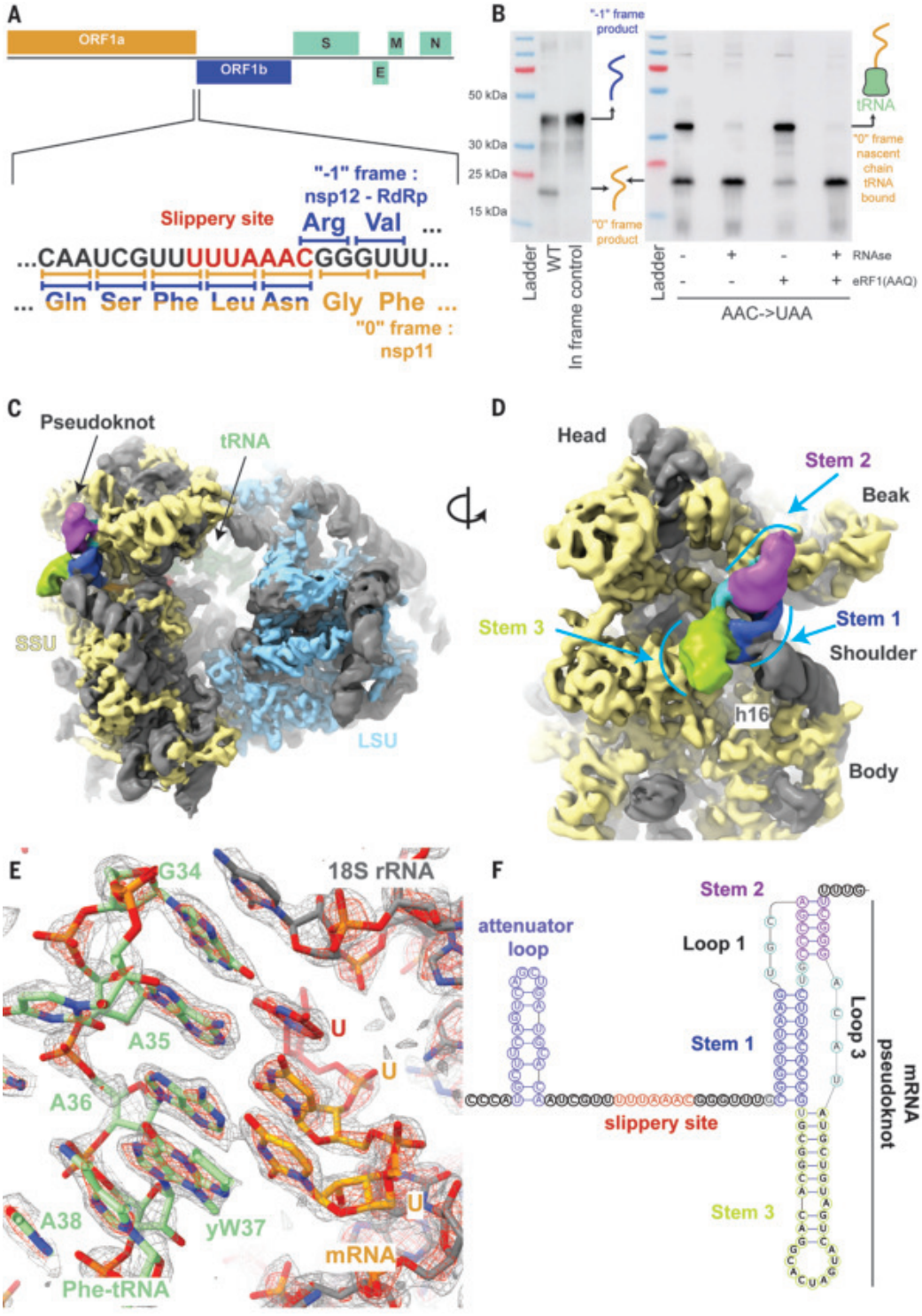


Fig. 1. The SARS-CoV-2 pseudoknot interacts with the ribosome and pauses translation upstream of the slippery site. (**A**) Schematic of the SARS-CoV-2 main ORF. In the close-up view of the frameshift event, codons and corresponding amino acids are shown. During −1 frameshifting, the slippery site codons UUA (Leu) and AAC (Asn) are the last codons decoded in the 0 frame. Upon −1 frameshifting of the AAC codon to AAA, translation resumes at the CGG (Arg) triplet, where elongation proceeds uninterrupted to produce full-length Nsp12. (**B**) In vitro translation reaction depicting pausing at the frameshift site, as shown with Western blotting. Efficient frameshifting is observed for the WT template, consistent with our dual luciferase assays (see methods). Samples for cryo-EM originally intended to be trapped by dominant negative eRF1 (AAQ) show a tRNA-bound pause in proximity of the frameshift site. The tRNA-associated band is lost upon RNase treatment. Reactions without added eRF1 (AAQ) produce a similarly paused product. (**C**) Overview of the density low-pass filtered to 6 Å with the pseudoknot found close to the entry of the mRNA channel on the small subunit (SSU). The SSU proteins are colored in yellow, the large subunit (LSU) proteins in blue, and the rRNA in gray. The pseudoknot is colored according to its secondary structure as in (F), and the P-site tRNA is colored in green. (**D**) Close-up view of the pseudoknot from the solvent-exposed side of the SSU. Helix h16 of the 18*S* rRNA interacts with the base of Stem 1. Unpaired loop-forming nucleotides are colored in cyan. (**E**) P-site codon-anticodon interactions reveal a Phe (UUU) codon interacting with tRNA(Phe). yW37, wybutosine at position 37. (**F**) Schematic of the revised secondary structure elements in the pseudoknot necessary for −1 programmed ribosomal frameshifting, with different functional regions labeled and colored accordingly.

transporter 1 (ABCE1), which is known to be involved in termination and recycling together with mutant eRF1 interacting with the stop codon (fig. S3). The second reconstruction resolved translating 80*S* ribosomes containing bound P- and E-site tRNAs (fig. S2). This reconstruction at 2.2-Å resolution allowed us to build the most accurate structure of a mammalian 80*S* ribosome so far and directly visualize many protein and virtually all rRNA modifications identified for the human ribosome based on quantitative mass spectrometry and as interpreted in a recent human ribosome structure (*20*, *21*), consistent with the complete conservation of all modified residues between rabbit and human ribosomal RNAs (rRNAs) (figs. S4 and S5; and tables S1 to S3). Importantly, this reconstruction also featured additional density at the entrance to the mRNA channel suggestive of a structured RNA, which, after focused classification, revealed a prominent density for

a complete 3′ frameshifting stimulatory pseudoknot at the entry of the mRNA channel on the 40*S* subunit (Fig. 1, C and D). The resolution of this reconstruction ranged from 2.4 Å at the core of the ribosome to ~7 Å at the periphery, where the most flexible regions of the pseudoknot are located (figs. S2 and S6). Based on the high-resolution maps that allowed visualization of the codon-anticodon interactions and modifications in the tRNA (Fig. 1E and fig. S6, A and B), we could unequivocally determine that a Phe-tRNA(Phe) was bound at

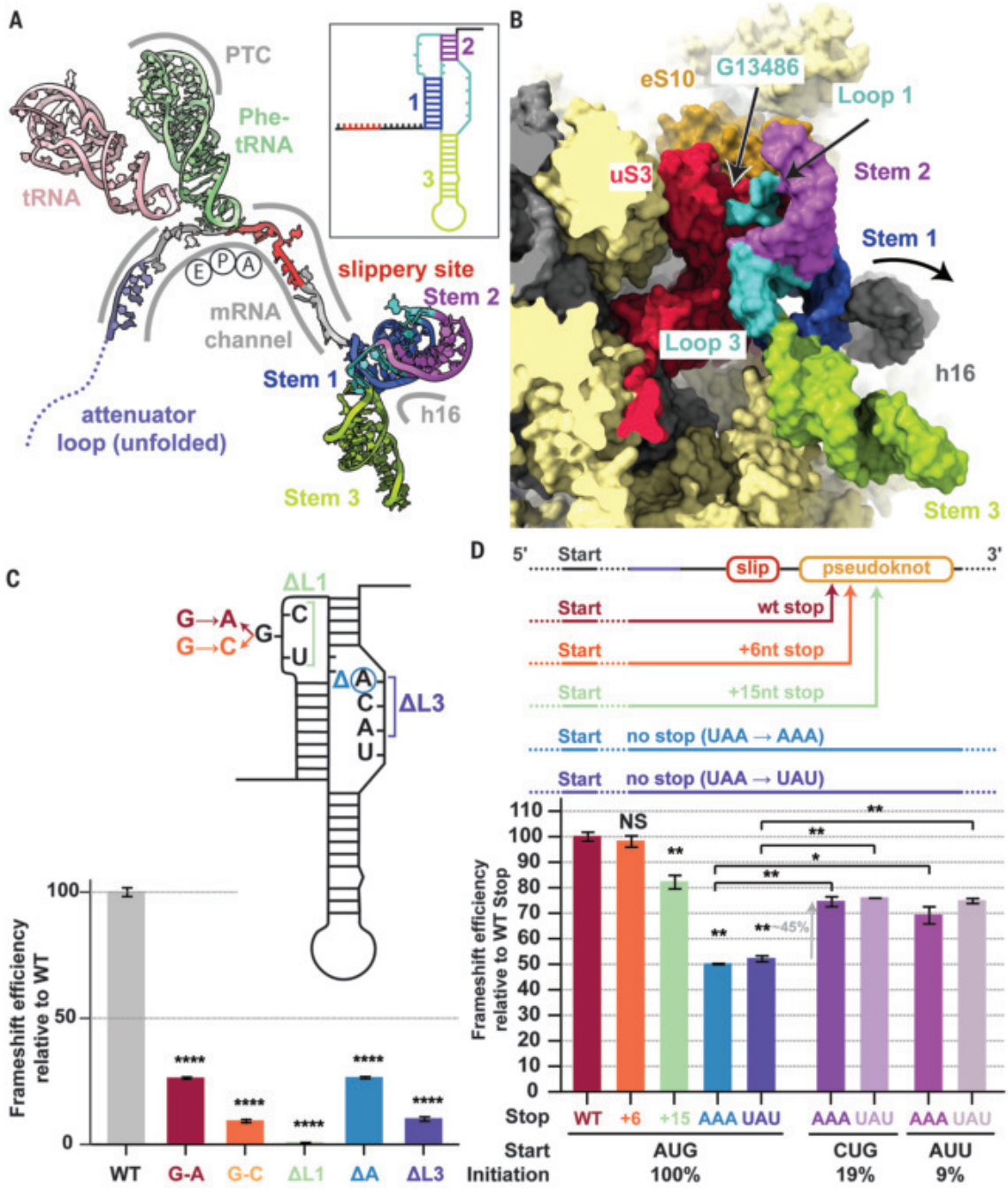


Fig. 2. Critical features of the ribosome-bound pseudoknot. (**A**) Overview of the frameshift-primed state. The stimulatory pseudoknot pauses the ribosome at the penultimate codon (UUU) of the slippery site (red), with P-sites (green) and E-sites (pink) occupied by tRNAs and an empty A-site awaiting decoding in the nonrotated state. The length of the spacer region (gray) is critical for exact positioning of the pseudoknot as the spacer exerts tension at the entry of the mRNA channel (fig. S6C). The inset shows a secondary structure depiction of the frameshift-stimulating pseudoknot colored accordingly. PTC, peptidyl transferase center. (**B**) The backbone of Loop 1 (UGC) (cyan) of the pseudoknot interacts with the N-terminal domain of uS3 (red) and the C-terminal tail of eS10 (orange). mRNA residue G13486 is flipped out and interacts with uS3 (fig. S6D). (**C**) Mutagenesis experiments using dual luciferase assays in HEK293T cells indicate that the G13486 interaction is specific. Mutation of G13486 to other residues leads to a marked reduction in frameshifting efficiency, and deletion of Loop 1 (ΔL1) completely abolishes frameshifting. Similarly, deletion of a single nucleotide (A13537) in Loop 2 reduces frameshifting, whereas deletion of the entire loop (ΔL2) abolishes frameshifting. Normalized (Firefly-Renilla) luciferase activities were calculated for each construct as a percentage of their individual normalized in-frame controls. Data are presented as mean values ± standard deviations of three biological replicates (sets of translation reactions) averaged after three measurements, with error bars representing standard deviations. ****P < 0.0001 by Student's two-tailed t test. (**D**) Mutagenesis experiments using dual luciferase reporter assays in HEK293T cells show that the position of the 0 frame stop codon influences frameshifting. When leaving the pseudoknot unaltered, an incremental increase in the distance of the 0 frame stop codon from the frameshift site leads to a concomitant decrease in frameshifting levels. Loss of the stop codon in the 0 frame leads to a sharp decline in frameshifting levels. This reduction is rescued by ~45% upon decreasing ribosome loading levels by implementing weaker initiation codons. The graph is normalized relative to the WT frameshifting of 25%. Mutations and complementary mutations are shown in fig. S8. Error bars represent standard deviation. NS, not significant; *P < 0.1; and **P < 0.01 by Student's two-tailed t test.

the P-site (*22*). The mRNA does not adopt any unusual structure in the A-site of the ribosome as was observed for the HIV-1 frameshifting sequence visualized on the bacterial ribosome (*23*). This implied that the ribosome is paused by the downstream pseudoknot located at the entrance to the mRNA channel such that the P-site tRNA interacts with the UUU codon just prior to the first codon, UUA, of the slippery site (Fig. 2A).

The pseudoknot causes ribosomal pausing prior to −1 frameshifting

The observation that the pseudoknot acts as an obstacle to slow down translation as the ribosome approaches the slippery site is mechanistically reasonable. Because the pseudoknot is a stable structural element in the mRNA, it will resist unfolding and consequently generate a back-pull on the viral RNA, resulting in an increased chance of −1 frameshifting as the tRNAs are translocated. A pause in translocation at a codon that precedes the slippery site, characterized by a >10 times longer occupancy prior to the slippage event, was observed in an analogous case of heptanucleotide −1 frameshifting on the bacterial *dnaX* gene using single-molecule experiments (*24*). According to this model, it would be anticipated that a further round of translocation results in unwinding of Stem 1 of the downstream stimulatory

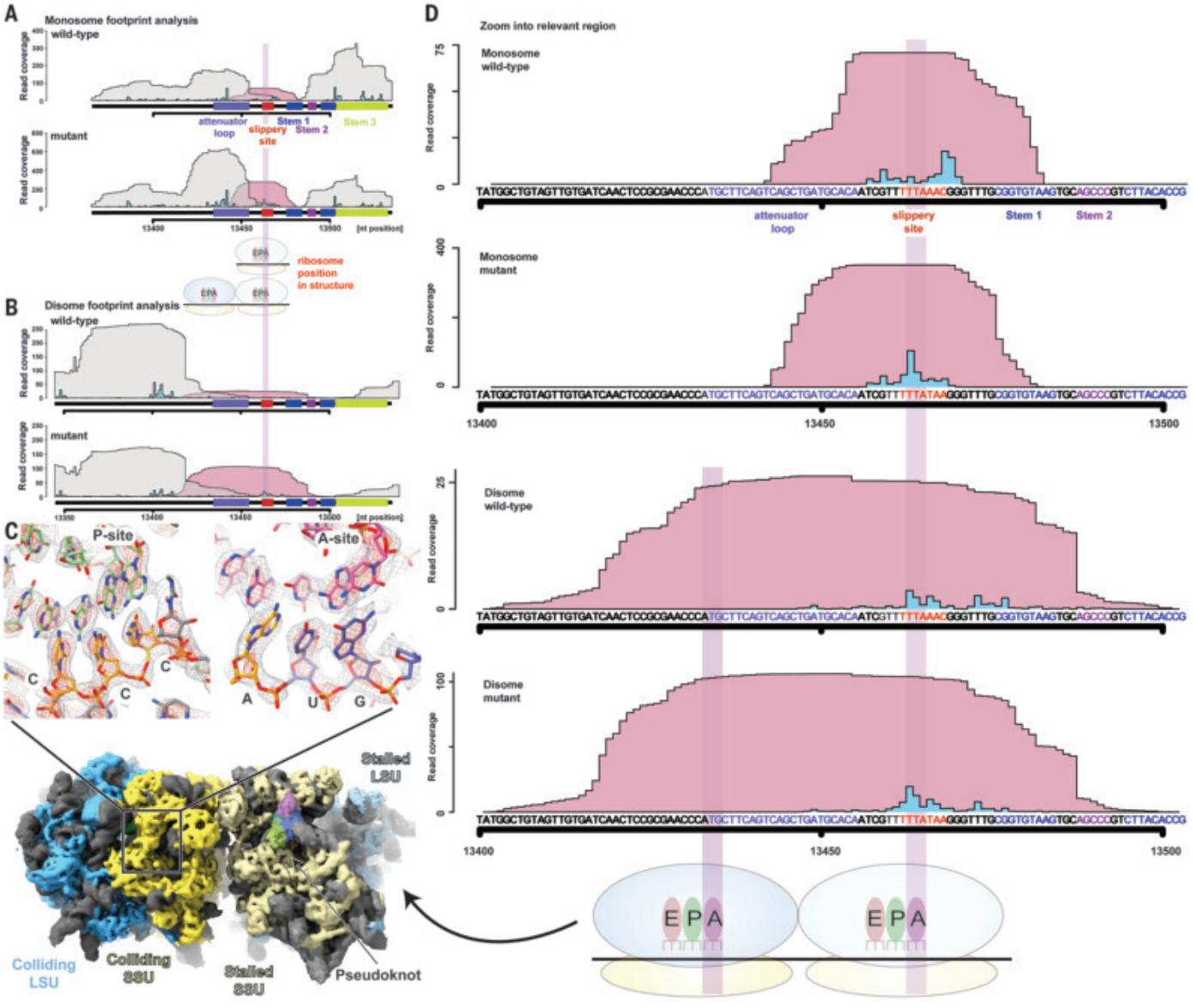


Fig. 3. Pseudoknot-mediated pause occurs prior to the −1 frameshifting event. (**A** and **B**) Footprint coverage for WT and mutant constructs determined by monosome-selective (A) and disome-selective (B) ribosome profiling. Pileup of reads from the indicated areas are plotted separately for reads that overlap (pink) or do not overlap (gray) the frameshift site (indicated by red bar below the *x* axis). The predicted A-sites of the ribosomes that give rise to the footprints are depicted as blue peaks. A-site predictions were carried out as described in the supplementary materials. (**C**) In high-resolution cryo-EM reconstructions of disomes at the frameshift site, the P- and A-sites of the trailing ribosome show occupancy of CCC and AUG codons, respectively, corresponding to the positions estimated by disome profiling. Disome maps were calculated by separately refining the orientational parameters for each ribosome. (**D**) Magnification of the frameshift region from (A) and (B) reveals that monosome profiles show transient occupancy in the vicinity of the frameshift site, whereas disome profiles, which are indicative of strong pause sites, show a similarly enhanced occupancy at the first codon (UUA) of the frameshift site in both WT and mutant constructs. A-site codons of the leading and trailing ribosome are highlighted with a translucent bar and correspond to those seen in the disome structure in (C).

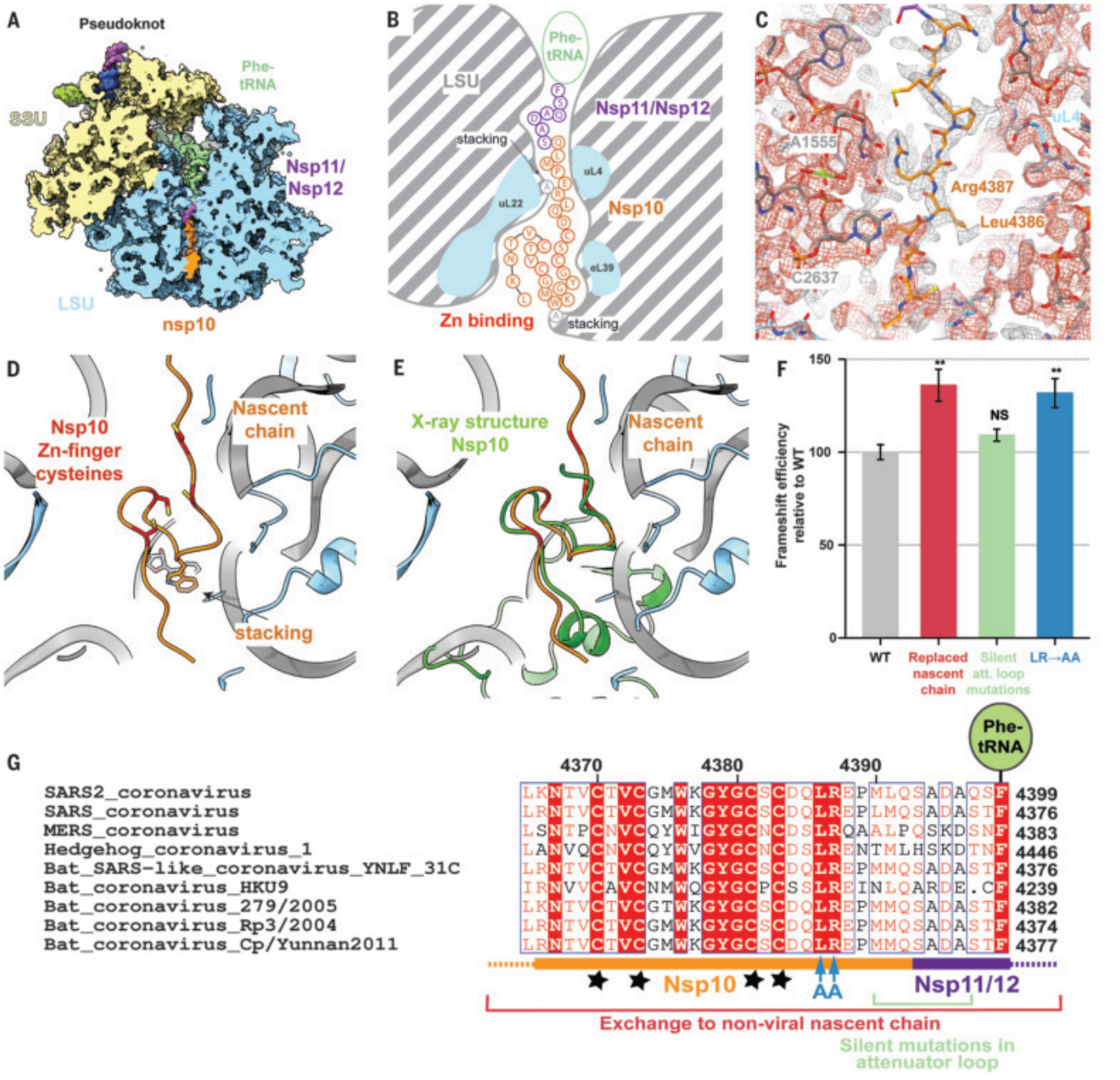


Fig. 4. The nascent viral polypeptide cotranslationally folds and specifically interacts with the ribosomal tunnel. (**A**) Cross-section of the pseudoknot-paused ribosome structure showing the exit tunnel. The nascent C terminus of Nsp10 (orange) and the N terminus of Nsp11/Nsp12 (purple) are visible from the PTC to the periphery of the ribosome exit tunnel (LSU in blue). (**B**) Schematic representation of the path of nascent peptide along the exit tunnel. Arg^{4387} stacks with 28*S* rRNA residue A^{1555} at the constriction site. Further down, where the tunnel widens, the C-terminal zinc finger domain of Nsp10 folds cotranslationally, with Trp^{4376} stacking on A^{2261} of 28*S* rRNA. (**C**) A well-ordered density is visible for Arg^{4387} of Nsp10 as it stacks onto A^{1555} of 28*S* rRNA at the constriction site and is stabilized by Leu^{4386}. The structure is shown within the cryo-EM map contoured at two different levels (gray and red). (**D** and **E**) The overlay of the cotranslationally folded zinc finger domain with the crystal structure of Nsp10 [green, PDB 2FYG (*37*)] reveals the structural similarity. (**F**) Probing the role of nascent chain interactions with the ribosome exit tunnel using an RRL in vitro system. Mutations of the interacting residues were tested for their effect on frameshifting shown in comparison to the WT frameshifting (41% frameshifting was normalized to 100%). Replacement of the entire nascent chain with an unrelated sequence leads to a 35% relative increase in frameshifting, which is only in part due to the loss of the 5′ attenuator loop. Interactions around the constriction site likely serve to attenuate frameshifting, because replacement of the interacting Arg^{4387} and stabilizing Leu^{4386} (LR) with Ala (AA) increases frameshifting by 30%. Error bars represent standard deviation. NS, not significant, and $^{**}P < 0.01$ by Student's two-tailed *t* test. (**G**) Alignment of SARS2 with closely related sequences of other coronaviruses highlighting the conservation of the mutated residues [colored as in (F)]. The shown sequence stretch encompasses the C-terminal zinc finger domain of Nsp10 (orange) and parts of Nsp11/Nsp12 (purple) visible in our reconstruction. Nascent-chain residues Leu^{4386} and Arg^{4387} that interact with the ribosomal exit tunnel are strictly conserved, whereas the conservation of neighboring residues is lower. Stars represent the four cysteines of the Nsp10 zinc finger. Single-letter abbreviations for the amino acid residues are as follows: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; and Y, Tyr.

pseudoknot structure. Consistently, in our structure of the eRF1 (AAQ)–bound ribosome that advanced one codon further along the mRNA, no clear secondary structure is visible at the entrance to the mRNA channel because the mRNA now becomes disordered at this position (figs. S1 and S3, A and B).

To investigate the slowdown of translation on the WT slippery sequence, we performed disome footprint profiling, a method that identifies translational pause sites through the analysis of transitory ribosome collisions (*25*–*27*) (see methods). Notably, recent studies using conventional ribosome profiling methodology reported a lack in monosome footprint coverage across the frameshifting region on the SARS-CoV-2 RNA (*11*, *28*), possibly because ribosomes in this area became trapped in temporary collisions. Moreover, the highly structured pseudoknot at the entry to the mRNA channel would likely preclude efficient trimming by ribonuclease I (RNase I), the enzyme used for footprint generation, further reducing efficient monosome footprint capture. Using a modified nuclease treatment protocol (see methods) that recovered monosome footprints from the frameshift region (Fig. 3, A and C), our experiments revealed that ribosome collisions occur as a result of ribosomal pausing at the same position that is observed in the structure of the pseudoknot-engaged ribosome (Fig. 3, B and D). Apparently, although the base substitutions creating a stop codon in the 3′ adjacent slippery site did not change the features of pausing, they increased the dwell time of the ribosomes at the pause site sufficiently to allow visualization in the cryo-EM experiment.

The results of our disome profiling experiments prompted us to structurally investigate disomes by cryo-EM. We were able to visualize the pseudoknot-paused ribosome followed by a closely trailing ribosome. Upon focused refinement, we obtained a high-resolution (3.1 Å) structure of the trailing ribosome in a rotated state (fig. S1). In congruence with our estimated positioning of the ribosomes in disome profiling (Fig. 3D), the purine-pyrimidine pattern of codon-anticodon pairs in the structure of the colliding ribosome revealed that the pause occurs with CCC and AUG triplets in the P- and A-sites, respectively (Fig. 3C).

The SARS-CoV-2 RNA pseudoknot specifically interacts with ribosomal proteins and 18*S* rRNA

The intermediate local resolution (5 to 7 Å) of the cryo-EM map in the area of the pseudoknot allowed us to visualize the overall fold of the RNA and readjust its previously predicted secondary structure (*14*–*17*, *19*) (Fig. 1, C, D, and F). The stimulatory pseudoknot forms an H-type pseudoknot with Stem 1 and Stem 2 coaxially stacked on top of each other to form a quasi-continuous helix, whereas Stem 3 stands out almost perpendicular to this plane (Figs. 1D and 2B). This corkscrew-like formation provides a bulky and well-structured obstacle wedged at the mRNA entry channel, which has the potential to resist unwinding by the helicase activity of the ribosome and generate tension on the upstream mRNA up to the decoding center. Stem 1 of the pseudoknot forms a 9–base pair helix that is GC rich at the bottom (Fig. 1F). The penultimate nucleotides of the "spacer region" before Stem 1 are located at the mRNA entry tunnel, where they interact with several basic residues in the C-terminal domain of uS3 on one side and are supported by uS5 from the other, with an additional weak contact contributed by the C-terminal end of eS30. uS3 and eS30 are primary components of the ribosome helicase, and uS5 has been proposed to be a component of the ribosomal helicase processivity clamp at the mRNA entry site (*29*, *30*). The observed distance between the P-site UUU codon and Stem 1 of the pseudoknot underscores the critical dependence of the frameshifting efficiency on the length of the spacer region (*31*). Translocation to the next codon would place the frameshifting codon UUA into the P-site, with a simultaneous increase in the tension of the mRNA and unwinding of the GC-rich base of Stem 1 upon entering the mRNA entry channel, comparable to the situation when the ribosome proceeds to the engineered stop codon, as observed in our eRF1 (AAQ)–stalled structure (fig. S3).

The pseudoknot structure also reveals a hitherto unobserved and possibly unappreciated role for the distal site of the mRNA entrance channel in helicase activity. Although mRNA unwinding studies outside the mRNA entrance channel have so far implicated only a helix in the C-terminal domain of uS3 (*32*), we noticed that Loop 1 of the pseudoknot contacts the N-terminal domain of uS3 as well as the C-terminal tail of eS10 (Fig. 2B and fig. S6D), whereas the flipped-out base G13486 in this loop forms specific interactions (Fig. 2B). Furthermore, because the pseudoknot is located at the entry to the mRNA channel, helix h16 of the 18*S* rRNA is noticeably pushed outward owing to a direct contact with the minor groove of Stem 1 (Fig. 2B and fig. S7A). Because the pseudoknot wedges between the head and the body of the small ribosomal subunit, it would restrict their relative motions that need to take place during translocation. This is consistent with the studies on dynamics of coronavirus frameshifting, which revealed that the mechanism of −1 frameshifting involves restriction of small subunit head motion (*33*).

The structure also reveals another key aspect of the architecture of the pseudoknot as the ribosome encounters it. The start of the pseudoknot is shifted relative to the predicted secondary structure (*14*–*17*, *19*) by two nucleotides. The two opposed nucleotides, which were assumed to base pair with Stem 1, are actually forming the start of Stem 3 by pairing with bases predicted to be in the single-stranded linker 2 (Fig. 1F and fig. S7, B and C). Our cryo-EM density reveals that Loop 3 accommodates a total of four nucleotides, three of which were originally attributed to Stem 2. Thus, we observe that Loop 3 is shifted and expanded relative to the initially predicted secondary structures (*14*–*17*, *19*).

To functionally support our structural findings and confirm the nature and specificity of the pseudoknot interactions, we performed structure-guided mutagenesis experiments using dual luciferase reporter assays in human embryonic kidney (HEK) 293T cells (see methods) and monitored the frameshifting efficiency relative to the WT (Fig. 2C). Mutation of G13486 of Loop 1 to another purine reduced the frameshifting efficiency to 30% of the WT level, and mutation of this base to a pyrimidine further reduced frameshifting to 15%. As expected from our structural data, deletions of the nucleotides of the spacer regions also had a deteriorating effect on frameshifting. Loss of Loop 1 entirely abolished frameshifting. Deletion of a single nucleotide of Loop 3 in agreement with its proposed role in forming the base-pairing interactions diminished the frameshifting rate to 25% of the WT level. Loss of the entire Loop 3 reduced frameshifting to 10% of WT levels.

Frameshifting efficiency depends on the position of the 0 frame stop codon

In SARS-CoV-2, the 0 frame stop codon is located five codons downstream of the frameshift site and is a constituent of Stem 1. The placement of the stop codon in such proximity to the frameshift site is a common feature in coronaviruses, and its presence in a critical region of the stimulatory pseudoknot prompted us to probe the effect of the distance of the 0 frame stop codon on frameshifting. To this end, knowledge of the 3D structure of the pseudoknot helped us to confidently manipulate the stop codon without hampering pseudoknot formation. We introduced mutations to incrementally extend the stop codon from the WT position and to completely remove the occurrence of a stop codon in the 0 frame (Fig. 2D and fig. S8). Whereas introducing a stop codon six nucleotides downstream of the WT position only marginally decreased the frameshifting rate (98% of WT), a stronger attenuation was observed when the distance of the stop codon was increased to 15 nucleotides from the WT stop (80% of WT). Finally, removal of the stop codon by two different point mutations led to a reduction of frameshifting efficiency to 50% of WT levels. To test whether reduced ribosomal loading rescues the effect

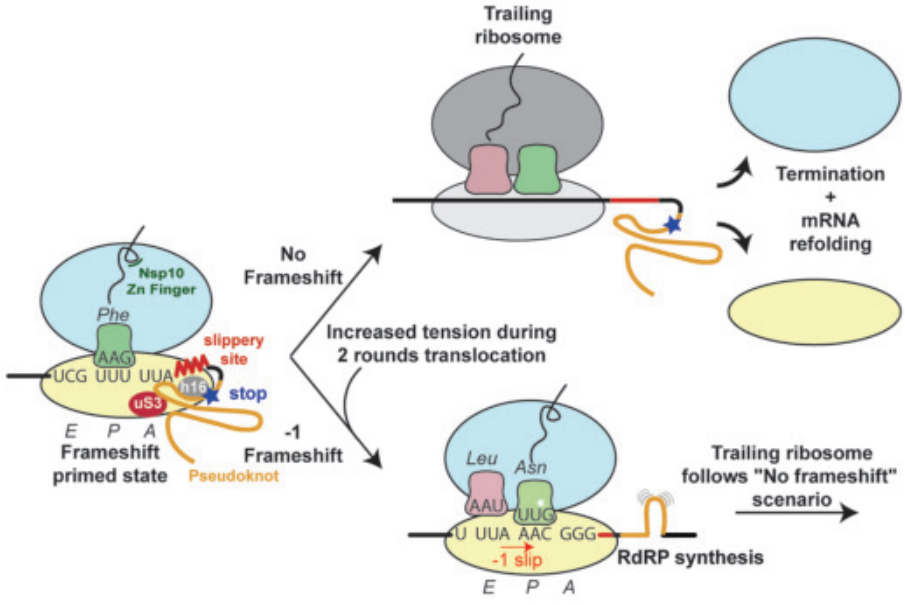


Fig. 5. Structure-based model for −1 programmed frameshifting in coronaviruses and its regulation. The observed interactions between the pseudoknot and the ribosome prime the system for frameshifting. The features of the pseudoknot and the interactions between the nascent chain and the ribosomal tunnel play a role in the efficiency of frameshifting. The efficiency of frameshifting is increased by the presence of a stop codon near the frameshifting site. Ribosomes that progress beyond the frameshifting site in the 0 frame quickly terminate and disassemble, thereby increasing the chances that the pseudoknot will refold before it is encountered by the closely trailing ribosome. The trailing ribosome in turn encounters the pseudoknot, which increases the possibility of undergoing −1 frameshifting.

of stop codon removal, we analyzed the frameshifting efficiency in the context of weaker initiation codons such as CUG and AUU (Fig. 2D). These constructs led to a 45% rescue of the reduction in frameshifting compared with stop codon mutants initiating at an AUG start.

Taken together, these observations suggest that the stop codon position plays an important role in maintaining optimum frameshift efficiency. We propose that the stop codon serves to prevent the closely trailing ribosome from encountering a viral RNA that was unfolded by the leading ribosome. In this case, upon encountering a stop codon, termination and subunit disassembly will occur, which will provide an opportunity for the pseudoknot to refold without the constraints of the mRNA channel (see Conclusions). According to this model, although the WT stop codon will make the frameshifting efficiency less sensitive to ribosome loading in the "no-frameshifting" scenario, the frameshifting events that occur after a −1 frameshift will nevertheless be more likely when the ribosomes are spaced further apart. Our measurements of the efficiency of frameshifting for the WT sequence in the context of different rates of translation initiation are in agreement with this hypothesis (fig. S9). This mechanism, consistent with our biochemical data, increases the efficiency of frameshifting to the levels required by SARS-CoV-2 and may be used by viruses in general when high-efficiency frameshifting is required.

Nascent chain forms specific interactions with the ribosomal exit tunnel

Notably, in the reconstruction of the paused translating ribosome, the nascent chain that corresponds to the viral polyprotein was visible along the entire length of the ribosomal exit tunnel (Fig. 4A). The density corresponds to the C-terminal region of Nsp10, which is the activator of the viral proofreading exonuclease and N7-methyltransferase Nsp14 (*34*, *35*), and then (depending on the frameshifting event) continues as either the viral RNA-dependent RNA polymerase Nsp12 (*6*) or as protein Nsp11, whose function is still unknown (Figs. 1A and 4B). The nascent chain makes several specific interactions with the ribosomal tunnel, one of which is at the constriction site where Arg^{4387} of Nsp10 interacts with A^{1555} of the 28*S* rRNA [corresponding to A^{1600} in humans, numbering according to PDB 6EK0 (*36*)] and is stabilized by the preceding Leu^{4386} (Fig. 4C). Notably, these two amino acids are highly conserved across multiple coronaviruses (Fig. 4G), although they are located in the unstructured C-terminal region of Nsp10 and therefore considered not to be important for the fold of the protein (*37*).

Further down the tunnel, the C-terminal end of Nsp10 adopts a partially folded zinc finger motif (Fig. 4, D and E), which, upon superposition, reveals similarity with the corresponding fully folded C-terminal domain previously observed in the crystal structure of SARS-CoV-1 Nsp10 (*37*). Trp^{4376}, which is located between the two pairs of cysteines that form the zinc finger, stacks with A^{2261} (A^{2418}), an interaction that might serve to promote the change of nascent chain direction and facilitate folding of the zinc finger at the end of the exit tunnel. Cotranslational events, such as insertion of a transmembrane domain at the exit of the ribosomal tunnel, were shown to promote −1 ribosomal frameshifting in alphaviruses (*38*).

To investigate whether the observed contacts between the nascent chain and the ribosomal tunnel are specific and whether these interactions and cotranslational folding of Nsp10 might play a role in modulating the frameshifting process, we used our dual luciferase reporter assay to measure the frameshifting efficiency of WT and mutant nascent chain sequence constructs. Because our measurements in HEK293T cells did not reveal an appreciable change of frameshift efficiency, we carried out the same experiments in vitro using RRL to monitor the effects in a single mRNA setup. Replacement of the entire nascent chain with an unrelated sequence leads to a 35% increase in frameshifting (Fig. 4F). Importantly, this effect was provoked by the change in peptide sequence and not simply by the loss of the 5′ attenuator loop, given that a reporter containing silent attenuator loop mutations resulted in only a slight increase in frameshifting (Fig. 4F). Mutation of the Leu^{4386} and Arg^{4387} to alanine led to a considerable (30%) increase in frameshifting (Fig. 4, F and G), implying that these nascent chain interactions with the ribosomal exit tunnel play an important role in regulating frameshifting levels, possibly mechanistically akin to the well-studied SecM stalling system in bacteria (*39*), where it was shown that cotranslational folding and the translocon-induced mechanical force can rescue the stall induced by interactions between the nascent chain and the ribosomal tunnel (*40*). These observations also suggest that any cellular nascent chain factors (*41*, *42*) might influence the rate of frameshifting.

Inhibition of viral replication by a compound that targets the SARS-CoV-2 pseudoknot

The sensitivity of the coronavirus to the finely controlled frameshifting levels (*13*) may present an opportunity to develop compounds that interfere with the frameshifting process and thus inhibit replication of the virus. Using computational modeling and reporter assays, compounds that have been predicted to bind the pseudoknot and inhibit SARS-CoV-2 frameshifting were described (*19*, *43*) but never tested with respect to their ability to inhibit viral replication. Furthermore, the fluoroquinolone compound merafloxacin was recently reported to also inhibit −1 frameshifting efficiency of SARS-CoV-2 and other betacoronaviruses (*44*). To demonstrate that the inhibition of frameshifting is a plausible strategy for drug development, we compared two of the previously described compounds with respect to their ability to reduce viral levels in infected African green monkey VeroE6 cells (fig. S10 and methods). Our experiments demonstrate that merafloxacin is a better candidate compound because it showed a concentration-dependent inhibition of frameshifting, whereas, contrary to earlier reports (*19*, *43*), the small-molecule ligand MTDB did not specifically inhibit frameshifting under our experimental conditions (fig. S10). The two compounds showed no cellular toxicity and resulted in a three to four orders of magnitude reduction of SARS-CoV-2 titer, with a half-maximal inhibitory concentration (IC_{50}) of 48 μM for MTDB and an order of magnitude higher efficacy for merafloxacin, with an IC_{50} of 4.3 μM (fig. S10). Because MTDB did not appear to affect frameshifting in our reporter construct experiments in vitro and in vivo, it is possible that it inhibits SARS-CoV-2 replication by a different mechanism. Although the potency range for these compounds is not what would be expected from potential drug candidates, it nevertheless provides a starting point for high-throughput screening and establishes that frameshifting is a viable target for therapeutic intervention against SARS-CoV-2.

Conclusions

Our results provide a mechanistic description of frameshifting that occurs during translation of the SARS-CoV-2 genome and reveal the features that may be exploited by the virus to finely control the stoichiometry of viral proteins at different stages of infection (Fig. 5). Interfering with the frameshifting process at the level of nascent chain interactions with the ribosomal tunnel, at the level of RNA folding that leads to the formation of the frameshift stimulatory pseudoknot, or to perturb the interactions between the pseudoknot and the mRNA channel represent viable strategies in our search for new drugs against SARS-CoV-2, the virus that is currently causing the global COVID-19 pandemic. Our results will also be useful for understanding the mechanism of programmed ribosomal −1 frameshifting (*4*), including that used by many other medically important viruses.

REFERENCES AND NOTES

1. P. V'kovski, A. Kratzel, S. Steiner, H. Stalder, V. Thiel, *Nat. Rev. Microbiol.* **19**, 155–170 (2021).
2. J. Parker, *Microbiol. Rev.* **53**, 273–298 (1989).
3. J. M. Ogle, A. P. Carter, V. Ramakrishnan, *Trends Biochem. Sci.* **28**, 259–266 (2003).
4. J. F. Atkins, G. Loughran, P. R. Bhatt, A. E. Firth, P. V. Baranov, *Nucleic Acids Res.* **44**, 7007–7078 (2016).
5. F. Dos Ramos, M. Carrasco, T. Doyle, I. Brierley, *Biochem. Soc. Trans.* **32**, 1081–1083 (2004).
6. P. V. Baranov *et al*., *Virology* **332**, 498–510 (2005).
7. E. P. Plant *et al*., *PLOS Biol.* **3**, e172 (2005).
8. M. C. Su, C. T. Chang, C. H. Chu, C. H. Tsai, K. Y. Chang, *Nucleic Acids Res.* **33**, 4265–4275 (2005).
9. I. Brierley, P. Digard, S. C. Inglis, *Cell* **57**, 537–547 (1989).
10. N. Irigoyen *et al*., *PLOS Pathog.* **12**, e1005473 (2016).
11. Y. Finkel *et al*., *Nature* **589**, 125–130 (2021).
12. V. Thiel *et al*., *J. Gen. Virol.* **84**, 2305–2315 (2003).
13. E. P. Plant, R. Rakauskaite, D. R. Taylor, J. D. Dinman, *J. Virol.* **84**, 4330–4340 (2010).
14. R. Rangan *et al*., *RNA* **26**, 937–959 (2020).
15. K. Zhang *et al*., Cryo-electron microscopy and exploratory antisense targeting of the 28-kDa frameshift stimulation element from the SARS-CoV-2 RNA genome. bioRxiv 2020.07.18.209270 [Preprint]. 20 July 2020; https://doi.org/10.1101/2020.07.18.209270.
16. T. C. T. Lan *et al*., Insights into the secondary structural ensembles of the full SARS-CoV-2 RNA genome in infected cells. bioRxiv 2020.06.29.178343 [Preprint]. 19 February 2021; https://doi.org/10.1101/2020.06.29.178343.
17. N. C. Huston *et al*., *Mol. Cell* **81**, 584–598.e5 (2021).
18. O. Namy, S. J. Moran, D. I. Stuart, R. J. C. Gilbert, I. Brierley, *Nature* **441**, 244–247 (2006).
19. J. A. Kelly *et al*., *J. Biol. Chem.* **295**, 10741–10748 (2020).
20. M. Taoka *et al*., *Nucleic Acids Res.* **46**, 9289–9298 (2018).
21. W. Li, S. T. L. Chang, F. R. Ward, J. H. D. Cate, *Nat. Commun.* **11**, 4941 (2020).
22. G. Keith, G. Dirheimer, The primary structure of rabbit, calf and bovine liver tRNAPhe*Biochim. Biophys. Acta* **517**, 133–149 (1978).
23. C. Bao *et al*., *eLife* **9**, e55799 (2020).
24. J. Choi, S. O'Loughlin, J. F. Atkins, J. D. Puglisi, *Sci. Adv.* **6**, eaax6969 (2020).
25. P. Han *et al*., *Cell Rep.* **31**, 107610 (2020).
26. A. B. Arpat *et al*., *Genome Res.* **30**, 985–999 (2020).
27. S. Meydan, N. R. Guydosh, *Mol. Cell* **79**, 588–602.e6 (2020).
28. M. Puray-Chavez *et al*., The translational landscape of SARS-CoV-2 and infected cells. bioRxiv 2020.11.03.367516 [Preprint]. 16 November 2020; https://doi.org/10.1101/2020.11.03.367516.
29. J. Rabl, M. Leibundgut, S. F. Ataide, A. Haag, N. Ban, *Science* **331**, 730–736 (2011).
30. S. Takyar, R. P. Hickerson, H. F. Noller, *Cell* **120**, 49–58 (2005).
31. Z. Lin, R. J. C. Gilbert, I. Brierley, *Nucleic Acids Res.* **40**, 8674–8689 (2012).
32. H. Amiri, H. F. Noller, *RNA* **25**, 364–375 (2019).
33. N. Caliskan, V. I. Katunin, R. Belardinelli, F. Peske, M. V. Rodnina, *Cell* **157**, 1619–1631 (2014).
34. M. Bouvet *et al*., *J. Biol. Chem.* **289**, 25783–25796 (2014).
35. E. C. Smith *et al*., *J. Virol.* **89**, 6418–6426 (2015).
36. S. K. Natchiar, A. G. Myasnikov, H. Kratzat, I. Hazemann, B. P. Klaholz, *Nature* **551**, 472–477 (2017).
37. J. S. Joseph *et al*., *J. Virol.* **80**, 7894–7901 (2006).
38. H. R. Harrington *et al*., *J. Biol. Chem.* **295**, 6798–6808 (2020).
39. H. Nakatogawa, K. Ito, *Cell* **108**, 629–636 (2002).
40. D. H. Goldman *et al*., *Science* **348**, 457–460 (2015).
41. G. Kramer, D. Boehringer, N. Ban, B. Bukau, *Nat. Struct. Mol. Biol.* **16**, 589–597 (2009).
42. K. Döring *et al*., *Cell* **170**, 298–311.e20 (2017).
43. S. J. Park, Y. G. Kim, H. J. Park, *J. Am. Chem. Soc.* **133**, 10094–10100 (2011).

ACKNOWLEDGMENTS

We thank A. Jomaa for advice on cryo-EM data processing, N. Casey for support with technical resources, and A. Piceno ni for help with the grid preparation. We are indebted to the ETH scientific center for optical and electron microscopy (ScopeM) for access to electron microscopes and, in particular, to M. Peterek and D. Boehringer. We thank J. Guo for helpful discussions and for providing the merafloxacin compound. **Funding:** This work was supported by the Swiss National Science Foundation (SNSF) to N.B. and V.T. (grants 31003A_182341 and 310030_173085) through the National Center of Excellence in RNA and Disease (project funding 182880) to N.B., V.T., and D.G.; by the Federal Ministry of Education and Research, Germany (BMBF; grant RAPID, #01KI1723A) to A.K. and V.T.; and by the Irish Research Council Advanced Laureate (IRCLA/2019/74) to J.F.A., who thanks G. F. Fitzgerald for funds from Carbery Group Ltd. **Author contributions:** P.R.B. and N.B. initiated the project and, together with J.F.A., designed the experiments. P.R.B. carried out biochemical experiments and sample preparation. A.S. and P.R.B. prepared grids. A.S. carried out data collection and calculated EM maps. A.S. and M.L. performed molecular model building and refinement. A.S., M.L., P.R.B., and N.B. interpreted the structure. P.R.B., A.S., M.L., and N.B. drafted the manuscript. A.S. prepared the figures. G.L. and K.M.O. performed dual luciferase reporter assays. A.M. prepared and purified the drug-like ligand MTDB. A.K. and V.T. formulated experiments to test viral infected cells. A.K. carried out testing of ligand on viral-infected cells. R.M., R.D., and D.G. designed, carried out, analyzed, and interpreted ribosome profiling experiments. All authors contributed to the final version of the manuscript. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** The structure and cryo-EM map for the high-resolution reconstruction is available in the Protein Data Bank (PDB) as PDB ID 7O7Y and in the Electron Microscopy Data Bank (EMDB) as EMD-12756, respectively. The structure refined into the further classified particle set reconstruction and the corresponding maps are available as PDB-7O7Z and EMD-12757, respectively. The structure of the ribosome bound with eRF1 (AAQ) and ABCE1 reconstruction and the corresponding maps are available as PDB-7O80 and EMD-12758, respectively. The structure of the colliding ribosome reconstruction and the corresponding maps are available as PDB-7O81 and EMD-12759, respectively. The structure of the composite map of the disome is available in the EMDB as EMD-12760. Ribosome profiling data have been deposited at GSE167421.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/372/6548/1306/suppl/DC1
Materials and Methods
Figs. S1 to S11
Tables S1 to S4
References (*44*–*78*)
MDAR Reproducibility Checklist

20 October 2020; resubmitted 24 February 2021
Accepted 7 May 2021
Published online 13 May 2021
10.1126/science.abf3546

REPORTS

SURFACE CHEMISTRY

Resolving multifrequential oscillations and nanoscale interfacet communication in single-particle catalysis

Y. Suchorski[1], J. Zeininger[1], S. Buhr[1], M. Raab[1], M. Stöger-Pollach[2], J. Bernardi[2], H. Grönbeck[3], G. Rupprechter[1]*

In heterogeneous catalysis research, the reactivity of individual nanofacets of single particles is typically not resolved. We applied in situ field electron microscopy to the apex of a curved rhodium crystal (radius of 650 nanometers), providing high spatial (~2 nanometers) and time resolution (~2 milliseconds) of oscillatory catalytic hydrogen oxidation, to image adsorbed species and reaction fronts on the individual facets. Using ionized water as the imaging species, the active sites were directly imaged with field ion microscopy. The catalytic behavior of differently structured nanofacets and the extent of coupling between them were monitored individually. We observed limited interfacet coupling, entrainment, frequency locking, and reconstruction-induced collapse of spatial coupling. The experimental results are backed up by microkinetic modeling of time-dependent oxygen species coverages and oscillation frequencies.

The overall reactivity of a technological catalyst results from the superposition of the activity of a huge number of oxide-supported metal nanoparticles (NPs) that can vary in size, shape, and type of exposed facets. Furthermore, the properties of each single NP result from the superposition of its individual facets and their interfaces. In the course of a reaction, the different metal NPs and facets may even respond differently to the reaction conditions, such as through restructuring (*1*), oxidation (*2*), coking (*3*), or sintering (*4*). All of these effects limit the validity of ensemble-derived structure-performance relations.

Several methods have been used to study the catalytic properties of single-catalyst particles, including electron microscopy (*5*–*7*), nuclear magnetic resonance (*8*), Raman, infrared and nano-infrared (*9*), x-ray scattering and nanotomography (*10*, *11*), fluorescence (*12*), and plasmonic nanospectroscopy (*13*). Our recent studies of individual palladium (Pd) NPs by means of photoemission electron microscopy (PEEM) revealed, using the "kinetics by imaging" approach, a long-ranging communication effect of the metal-oxide interface in carbon monoxide (CO) oxidation (*14*). In this approach, single-particle kinetic measurements were based on the local PEEM image intensity variation during the reaction. However, the individual facets of the Pd particles could not be resolved because of the resolution limit of the PEEM used.

[1]Institute of Materials Chemistry, TU Wien, Getreidemarkt 9, 1060 Vienna, Austria. [2]University Service Center for Transmission Electron Microscopy, TU Wien, Wiedner Hauptstraße 8-10, 1040 Vienna, Austria. [3]Department of Physics and Competence Centre for Catalysis, Chalmers University of Technology, 41296 Göteborg, Sweden.
*Corresponding author. Email: guenther.rupprechter@tuwien.ac.at

In the present work, this approach is refined by resolving even the individual (hkl) nanofacets of a single catalytic particle. We applied in situ field electron microscopy (FEM), which provided high spatial (~2 nm) and time resolution (~2 ms) and allowed imaging of adsorbed species and reaction fronts, to catalytic hydrogen oxidation on a hemispherical apex of a curved rhodium (Rh) crystal (radius of 650 nm). The FEM studies were complemented by directly visualizing the active sites using the ionized reaction product (H_2O molecules) as imaging species in the field ion microscopy (FIM).

Altogether, we obtained a comprehensive picture of the ongoing oscillatory hydrogen oxidation, visualizing the working surface, reactants, and even the products. Dozens of nanofacets could be individually resolved, revealing their reactivity (active sites) and the nanoscale communication between them during the reaction in real time. Furthermore, we detected how just a few atomic rows on the surface could influence the local activity, leading to frequency locking and entrainment effects and eventually to a collapse of spatial coupling. The experimental results were supported by microkinetic modeling of time-dependent oxygen coverages and oscillation frequencies.

Like a smaller multifaceted catalytic NP, the curved Rh crystal exhibited several nanometer-sized facets (10 to 100 nm in diameter) of different crystallographic orientations [Fig. 1A, scanning electron microscopy (SEM) image]. The surface of the apex of the Rh specimen could be imaged with FEM (Fig. 1B), allowing monitoring of the facet-resolved kinetics, transfer of species, and reaction coupling between facets. In FEM, the electric field applied to the sample deforms the potential barrier on the surface, leading to tunneling of electrons into vacuum. The emitted electrons are accelerated toward a screen, where they form a magnified projection image of the surface with ~2-nm resolution.

The image contrast in a FEM results from the local work function convoluted with the local electrostatic field. As a result, FEM allows direct identification of adsorbed reactants on the sample surface, provided the work function differences are large enough for detectable image-contrast. In addition to "static" imaging, FEM can also monitor dynamic surface processes in real time, such as diffusion (*15*, *16*) or catalytic surface reactions (*17*–*20*).

The specimens used in previous FEM and FIM studies (*17*–*21*) were typically nanotips with a radius of curvature one magnitude smaller (25 to 50 nm) than the curved crystal of the present study, so that the effects presented here could not be observed. Catalytic H_2 oxidation is a well-suited reaction for FEM imaging because there is a sufficient work function difference between the catalytically inactive, oxygen-covered Rh surface (high O_{ads} coverage, high work function, and dark contrast) and the active surface (low H_{ads} and O_{ads}

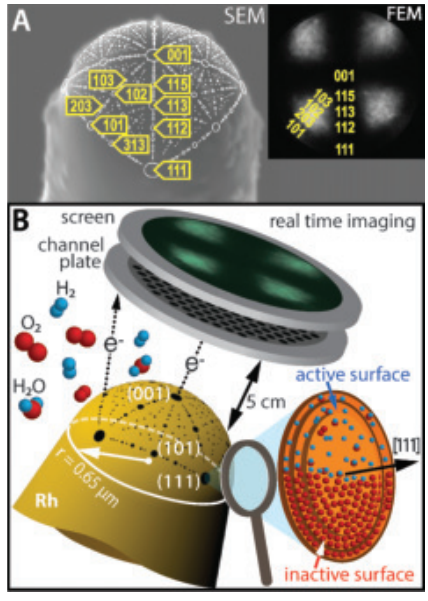


Fig. 1. Apex of a curved Rh crystal as a single catalytic particle. (**A**) SEM micrograph of the Rh specimen with a crystallographic net as overlayer (radius of the hemispherical apex, 650 nm). (Inset) The corresponding FEM micrograph with nanofacets indicated. (**B**) Schematics of the experiment. Electrons, field-emitted from the sample surface, create a projection image on the screen. The FEM chamber is used as a flow reactor, and the FEM image brightness reflects the local reaction rate. The locally magnified schematic shows a Rh(111) nanofacet with a reaction front between the catalytically active and inactive surface.

coverage; low work function, and bright contrast) (*20*). Unlike scanning techniques, FEM provides parallel imaging; the electrons forming the image are emitted from each atomic site simultaneously.

The experimental approach is illustrated in Fig. 1B: The sample was characterized by SEM (Fig. 1A), and the surface crystallography was deduced from the stereographic projection by combining SEM and FEM images (Fig. 1A, inset; supplementary text S1; and fig. S1) (*22*). The FEM chamber was operated as flow reactor under reactant pressures in the 10^{-5} to 10^{-6} mbar range (Fig. 1B). The local reaction kinetics of individual nanosized Rh(hkl) domains of the surface sample were extracted from real-time FEM video files following the recently developed approach (*23*).

Catalytic hydrogen oxidation follows the Langmuir-Hinshelwood mechanism: The reactants hydrogen and oxygen both first dissociatively adsorb on the catalyst surface before they react to water, which desorbs immediately at the given reaction temperatures (*24*). Under specific conditions (both in vacuum and atmospheric pressure, temperatures above 300 K), the H_2 oxidation reaction rate on platinum (Pt) and Rh may oscillate between its catalytically active and inactive states in a self-sustained way despite constant external parameters, such as temperature and pressure of reactants (*25*–*27*). Different feedback mechanisms governing the oscillations in surface reactions—such as surface reconstruction, subsurface oxygen, and oxide formation—were discussed in the literature (*25*–*29*). In the present case, fine tuning of experimental conditions plays a crucial role, allowing subsurface oxygen formation but preventing oxidation of the specimen surface. Apart from fundamental science, oscillating modes of catalytic reactions may also have technological impact: Periodic operation regimes can improve catalytic reactor performance (*30*) or enable tuning of surface activity (*31*).

We studied the oscillating mode of H_2 oxidation on the facetted Rh particle at constant partial pressures of oxygen (PO_2) = 4.4 × 10^{-6} mbar and hydrogen (PH_2) = 4.8 × 10^{-6} mbar and different constant temperatures of 413, 433, and 453 K. At 413 K, oscillations occurred in a synchronized way on almost all particle facets, apart from the central non-oscillating Rh(001) facet (Fig. 2A). This is evident from the FEM intensity curves for regions of interest (ROIs) placed on the Rh(135), Rh(111), Rh(115), and Rh(001) nanofacets. The parallel imaging principle of FEM allowed the construction of a frequency map (Fig. 2A, bottom) that showed synchronized oscillations on different individual facets. Details on the construction of the frequency map are presented in supplementary text S2 and fig. S2 (*22*).

Previous nanometer-scale FIM studies of the oscillating hydrogen oxidation on sharp Pt and Rh nanotips revealed that kinetic transitions between the catalytically active and inactive steady states occurred always in a spatially synchronized way over the entire sample surface (*26*, *32*, *33*). High-vacuum conditions excluded gas-phase reaction coupling, so the spatial coupling was provided by surface diffusion of hydrogen (*34*, *35*). Similar

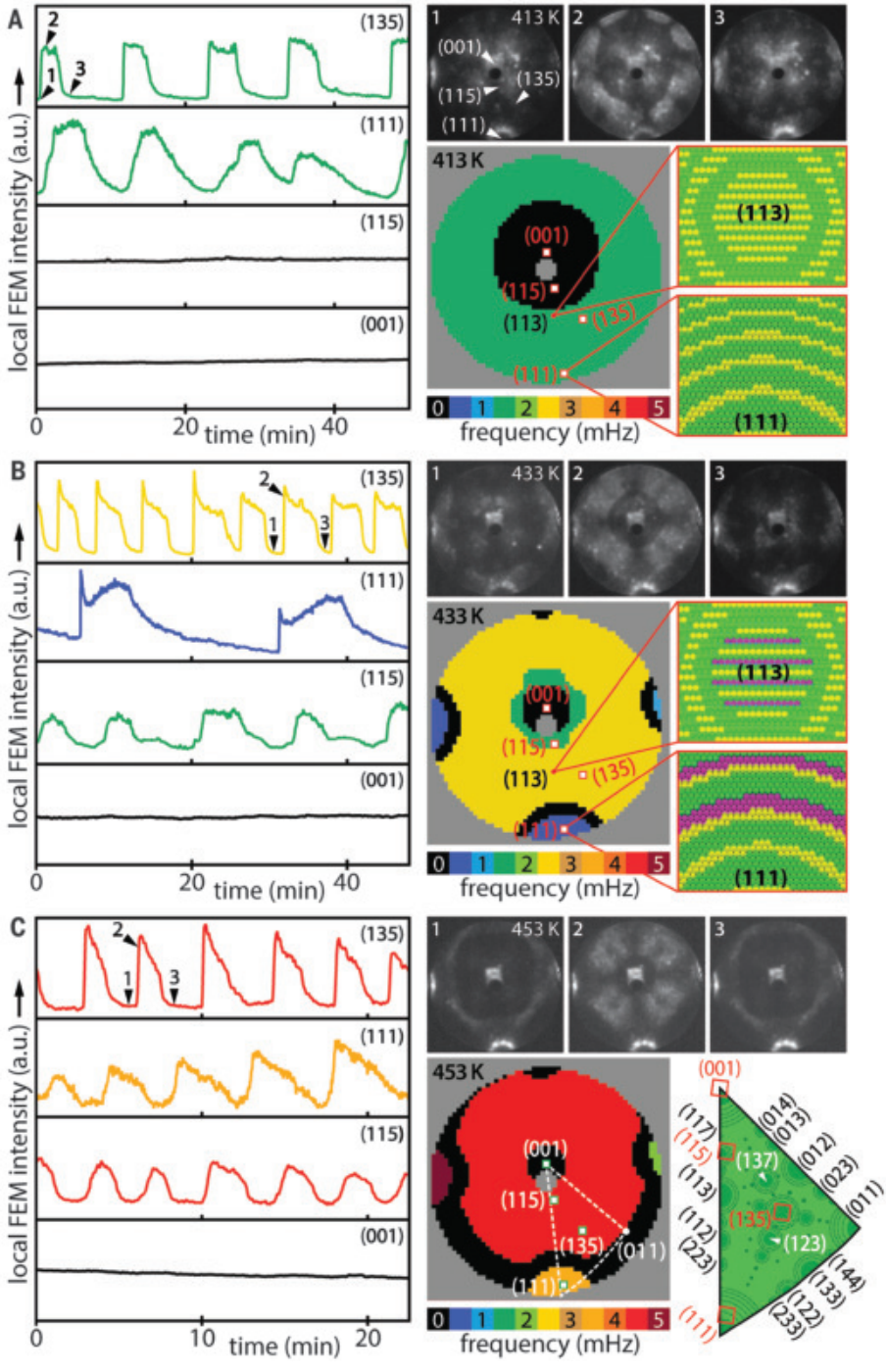


Fig. 2. Oscillations in the catalytic H_2 oxidation reaction. Imaging of the curved Rh crystal shown in Fig. 1A, observed with FEM at PO_2 = 4.4 × 10^{-6} mbar, PH_2 = 4.8 × 10^{-6} mbar, and at constant temperatures (413, 433, and 453 K). (Left) FEM intensity curves. (Top right) FEM video frames (1.1 by 1.1 μm each). (Bottom right) Frequency maps and crystallography. (**A**) Temperature (*T*) = 413 K. ROIs (40 by 40 nm) for the FEM intensity curves are placed in the (135), (111), (115), and (001) regions (compare with Fig. 1). (**B**) *T* = 433 K. All ROIs are the same as in (A). (**C**) *T* = 453 K. ROIs correspond to those in (A) and (B). (Insets at bottom right) (A) Unreconstructed and (B) reconstructed stepped (113) and (111) facets as well as (C) arrangements of Rh(hkl) facets within a segment indicated with the white dotted line, with positions of ROIs indicated with red squares.

global synchronization on the nanometer-scale was observed in CO oxidation and nitric oxide (NO) reduction, suggesting that catalytic reactions (and therein occurring self-sustaining oscillations) in nanosystems are spatially synchronized over the entire particle surface (*21*, *33*–*37*). However, as shown below, this does not hold true under certain conditions.

When the sample temperature was increased from 413 to 433 K, the reaction exhibited local multifrequential oscillations (Fig. 2B), in which the peripheral {111} regions (Fig. 2B, frequency map, blue) oscillated independently from the surrounding extended (Fig. 2B, frequency map, yellow) sample region, despite being separated only by tens of nanometers. The independence of oscillations on different facets demonstrates the intraparticle heterogeneity. The frequencies differed up to 3.8-fold, as illustrated by the FEM intensity curves in Fig. 2B, middle. The oscillations from the extended area also partially permeated into the central region, so that the nonoscillating area shrank. The nascent oscillating part of the previously nonoscillating central region (Fig. 2B, frequency map, green) showed a double-period–like feature, presumably resulting from the interaction of perturbing reaction fronts (infiltrating from the surrounding yellow region in Fig. 2B) and the natural period of {115} regions, which is governed by the local feedback mechanism involving the formation of subsurface oxygen (*27*, *38*).

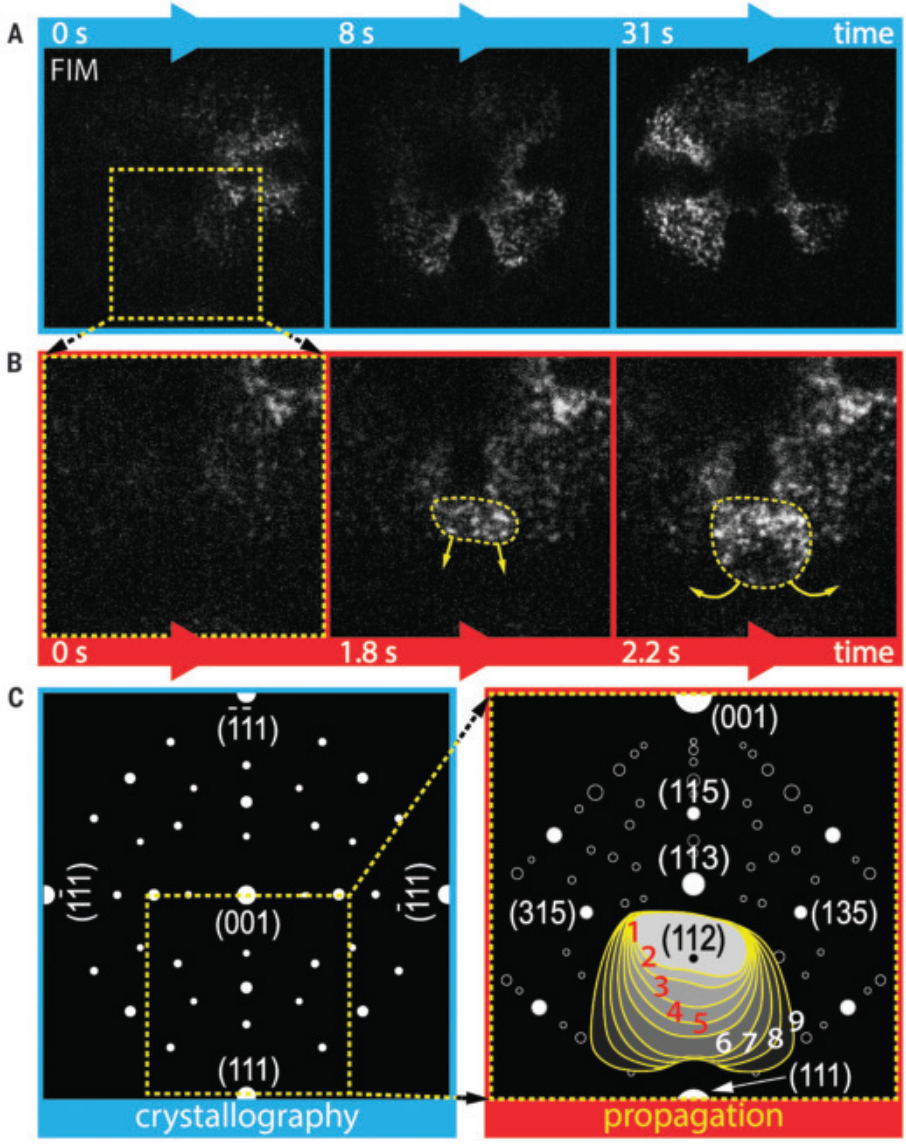


Fig. 3. FIM mapping of active sites on the curved Rh crystal during hydrogen oxidation by using ionized product (water) species as imaging ions. (**A**) FIM images taken during the kinetic transition from the inactive to the active state (frame size, 1.1 by 1.1 μm). The evolution of water production over the whole Rh curved crystal surface is shown. (**B**) Slow-motion snapshots of the magnified yellow square region (580 by 580 nm) indicated in (A) taken at 0, 1.8, and 2.2 s. (**C**) (Left) Crystallographic map according to the FIM images in (A). (Right) Magnified region from (B) with contours of the reaction front, demonstrating the predominant directions of the reaction front propagation. The numbered front contours 1 to 9 correspond to 1.8, 1.9, 2.0, 2.1, 2.2, 2.4, 2.7, 3.0, and 3.6 s, respectively.

At 453 K, the region with the double-period became synchronized with the ring-shaped surrounding (Fig. 2C, red region), which in turn contracted and increased in its frequency from 2.7 to 4.3 mHz. The peripheric {111} regions continued to oscillate fully independently from the rest of the sample and each other, but with increased frequencies (supplementary text S3 and fig. S3) (*22*).

The observed behavior was unexpected because the surface diffusion length of hydrogen (in the submicrometer range) greatly exceeded the distance between any two sample facets and would have synchronized the oscillations on the different nanofacets. Apparently, the spatial coupling collapsed in the present case, even though obstacles that can cause such collapse, such as grain boundaries between micrometer-sized domains of a polycrystal (*39*), were not present on the curved Rh particle. The nanofacets were separated solely by nano-ridges and nanometer-sized regions of stepped surfaces.

As demonstrated recently (*27*, *38*), the oscillations in H_2 oxidation on Rh are generated by the periodic formation and depletion of subsurface oxygen, so the observed behavior may be related to local peculiarities of periodic subsurface oxygen formation and depletion and to the coupling of this process with neighboring regions. Under the pressure conditions used, qualitative changes in oscillations occurred between 413 and 433 K, whereas the changes observed between 433 and 453 K were mainly quantitative, so we focused on processes on the Rh surface occurring between 413 and 433 K. In this temperature interval, the (1×2) missing row reconstruction on Rh{011} and Rh{113} facets (*40*, *41*) and step-doubling reconstruction on the (111)-oriented terraces of vicinal Rh surfaces (*42*) take place. Thus, at the two different temperatures (Fig. 2, A and B), the reaction proceeded on crystallographically differing surfaces (Fig. 2, corresponding ball models).

These changes in surface crystallography and temperature altered the propagation of reaction fronts and impeded their spreading to certain areas. Because the oscillation frequencies are transported with reaction fronts, the synchronized regions in the frequency map corresponded to the areas of front circulation (supplementary text S4 and figs. S4 and S5) (*22*). The protruding atomic rows created by the reconstruction effectively hindered the expansion of oscillating fronts from or toward the {111} regions and decoupled these regions

from the central oscillating area. The protruding atomic rows functioned as nanosized frequency transformers and allowed different nanofacets to support different oscillatory regimes.

The increased local roughness of the particle surface also modified the local kinetics of subsurface oxygen formation and changed the frequency of oscillations (compare Fig. 2, A and B). This reaction does not oscillate on a Rh(111) single crystal surface nor on the (111)-oriented domains of a Rh foil (*38*). The unexpected present oscillations of the Rh{111} regions can be explained by the regular steps on these regions caused by the hemispherical particle shape (as marked in Fig. 1), whereas sporadic steps on the planar Rh{111} surfaces cannot generate oscillations. The sufficient density of steps facilitating subsurface oxygen formation was shown to be a requirement for oscillations in H_2 oxidation on Rh surfaces (*38*). Under the present conditions, subsurface oxide formation does not result in surface oxide formation owing to kinetic limitations of surface oxide formation (*43*). Because the reaction does not oscillate on the oxidized Rh surface, balancing the narrow range between the subsurface oxygen formation and Rh surface oxidation is crucial for the present observations.

To reveal the reactivity and decoupling function of the highly corrugated regions between the {111} facets and the central part of the sample, the catalytically active sites were directly imaged. For the visualization of catalytically active sites, the ionized reaction product water, forming H_3O^+ and a small amount of H_2O^+ ions (*44*), was used as imaging species in the FIM mode of our microscope. Corresponding water-FIM images obtained in situ during the kinetic transition from the inactive (O covered) to the active state (Fig. 3) at 393 K not only illustrate the transition itself but also revealed the decoupling function of the highly corrugated regions that separate the {111} facets and the central area of the sample. The water production (bright regions) started in the vicinity of {112} facets, which appeared to be the most active, and then spread over almost the entire surface.

The "slow-motion" frames (Fig. 3B) and a scheme showing the front propagation (Fig. 3C, numbered front contours) demonstrate that the reaction front did not cross the (113) facet (because of the presence of protruding atomic rows) and hence did not proceed toward the central (001) plane. Rather, the front propagated toward the (111) facet (Fig. 3C, front contours 1 to 5) and was eventually diverted around the (111) facet because of the double-step reconstructions surrounding the (111) vicinals (Fig. 3C, front contours 6 to 9).

To deepen the understanding of the observed effects, we performed microkinetic simulations within a model describing oscil-

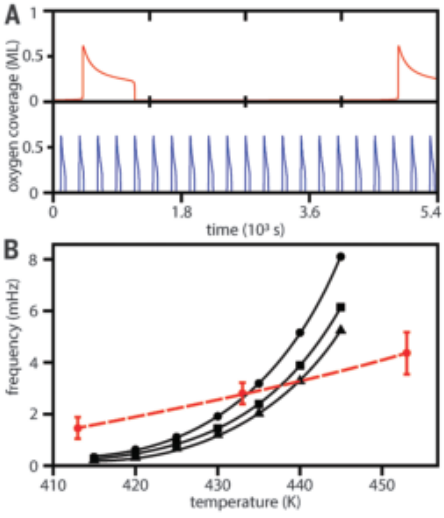


Fig. 4. Simulations of the kinetic oscillations at O_2 and H_2 pressures of 4.4×10^{-6} and 4.8×10^{-6} mbar, respectively. (**A**) Coverage of oxygen as a function of time with activation energy of subsurface oxygen formation E_{ox} = 1.100 eV at (top) 415 K and (bottom) 440 K. (**B**) Frequency of the oscillations as a function of temperature for E_{ox} = 1.085 eV (circles), E_{ox} = 1.100 eV (squares), and E_{ox} = 1.108 eV (triangles). Red symbols indicate experimental frequency data obtained at the {135} nanofacets from Fig. 2.

lations in H_2 oxidation on Rh caused by field-enhanced surface oxidation of Rh and observed with FIM (*37*). In the present FEM studies, such effects that result from the field-induced electron-density redistribution cannot occur because the field strengths were considerably lower and, even more notably, because of the opposite direction of the field vector (unlike FIM, the sample is negatively charged in FEM). The electrostatic field of ≤4 V/nm applied in the FEM did not influence the catalytic H_2 oxidation (supplementary text S5) [(*20*) and discussion therein], so we modified the model for the present quasi field-free case (supplementary text S6).

In this subsurface oxygen feedback model, two steady states (high activity ↔ low activity) periodically switch through a coverage-dependent sticking coefficient of oxygen. The high initial sticking coefficient of oxygen on metallic Rh leads to a high coverage of adsorbed oxygen O_{ads} (inactive state) and subsequent formation of subsurface oxygen O_{sub} (*38*). The presence of O_{sub} reduces the sticking of oxygen on the surface, and hydrogen competitively coadsorbs with oxygen and reduces the coverage of O_{ads} and O_{sub} to their initial values to start the next cycle (details of the cycle and of the model calculations are provided in supplementary text S6 and fig. S6) (*22*). The calculated oscillations in oxygen coverage on the Rh surface for the present conditions are shown in Fig. 4A. Such oscillations of the oxygen coverage caused oscillating variations of the surface work function, which are captured in the experiments as variations of the FEM image intensity. To represent different facets, the temperature dependence of the oscillation frequency is shown in Fig. 4B for three different activation energies of subsurface oxygen formation E_{ox} in comparison with the experimental data from Fig. 2.

Solitary atomic-size surface features were shown to govern the catalytic behavior of an entire submicrometer particle through interfacet communication. Choosing a surface containing "pacemaker nanofacets" large enough to undergo surface reconstruction, but still with a curvature radius that enabled FEM imaging, was crucial for revealing the new effects. The interfacet communication ensured that oscillations remained single frequency within the unicolor areas in the frequency maps (Fig. 2B, compare the yellow, blue, and green regions), although differently oriented nanofacets within these areas should in principle exhibit different frequencies (as the different domains of a polycrystal do) (*27*, *38*). The "natural" frequencies would be governed by the rate of subsurface oxygen formation and depletion, which is structure dependent (*38*). In the present case, many differently oriented facets of the Rh particle oscillate with a "forced" rather than their "natural" frequency. The heterogeneous surface thus behaves as an extended system of diffusively coupled oscillators, exhibiting effects such as entrainment and frequency locking that are characteristic for these systems (*45*, *46*).

REFERENCES AND NOTES

1. P. Nolte *et al.*, *Science* **321**, 1654–1658 (2008).
2. M. E. Grass *et al.*, *Angew. Chem. Int. Ed.* **47**, 8893–8896 (2008).
3. D. Neagu *et al.*, *Nat. Commun.* **6**, 8120 (2015).
4. T. W. Hansen, A. T. Delariva, S. R. Challa, A. K. Datye, *Acc. Chem. Res.* **46**, 1720–1730 (2013).
5. T. W. Hansen *et al.*, *Science* **294**, 1508–1510 (2001).
6. M. Plodinec, H. C. Nerl, F. Girgsdies, R. Schlögl, T. Lunkenbein, *ACS Catal.* **10**, 3183–3193 (2020).
7. R. Reske, H. Mistry, F. Behafarid, B. Roldan Cuenya, P. Strasser , *J. Am. Chem. Soc.* **136**, 6978–6986 (2014).
8. Y. Li *et al.*, *Nat. Commun.* **8**, 581 (2017).
9. S. Dery, E. Amit, E. Gross, *Top. Catal.* **61**, 923–939 (2018).
10. A. Lassenberger *et al.*, *Chem. Mater.* **29**, 4511–4522 (2017).
11. Y. Liu, F. Meirer, C. M. Krest, S. Webb, B. M. Weckhuysen, *Nat. Commun.* **7**, 12634 (2016).
12. Y. Ebenstein, T. Mokari, U. Banin, *Appl. Phys. Lett.* **80**, 4033–4035 (2002).
13. P. Tabib Zadeh Adibi, T. Pingel, E. Olsson, H. Grönbeck, C. Langhammer, *ACS Nano* **10**, 5063–5069 (2016).
14. Y. Suchorski *et al.*, *Nat. Mater.* **17**, 519–522 (2018).
15. R. Gomer, *Rep. Prog. Phys.* **53**, 917–1002 (1990).
16. J. Beben, Y. Suchorski, *Prog. Surf. Sci.* **74**, 3–24 (2003).
17. V. V. Gorodetskii, V. I. Elokhin, J. W. Bakker, B. E. Nieuwenhuys, *Catal. Today* **105**, 183–205 (2005).
18. Y. Suchorski, W. Drachsel, *Top. Catal.* **46**, 201–215 (2007).
19. C. Barroo, T. V. de Bocarmé, Y. De Decker, N. Kruse, "Surface reactions investigated at the nanoscale by field emission techniques: nonlinear dynamics of the catalytic hydrogenation of NO and NO_2 over platinum crystallites" in *Model Studies in Heterogeneous Catalysis*, K. Wandelt, Ed. (Elsevier, ed. 1, 2018), vol. 2.1 of *Encyclopedia of Interfacial Chemistry*, chap. 25, pp. 251–260.

20. M. Datler *et al.*, *Catal. Lett.* **146**, 1867–1874 (2016).
21. T. Visart de Bocarmé, G. Beketov, N. Kruse, *Surf. Interface Anal.* **36**, 522–527 (2004).
22. Materials and methods are available as supplementary materials.
23. Y. Suchorski, G. Rupprechter, *Surf. Sci.* **643**, 52–58 (2016).
24. L. K. Verheij, M. B. Hugenschmidt, *Surf. Sci.* **416**, 37–58 (1998).
25. R. Imbihl, G. Ertl, *Chem. Rev.* **95**, 697–733 (1995).
26. J.-S. McEwen, P. Gaspard, T. V. de Bocarmé, N. Kruse, *Proc. Natl. Acad. Sci. U.S.A.* **106**, 3006–3010 (2009).
27. Y. Suchorski *et al.*, *Nat. Commun.* **9**, 600 (2018).
28. Y. Suchorski, W. Drachsel, V. Gorodetskii, V. Medvedev, H. Weiss, *Surf. Sci.* **600**, 1579–1585 (2006).
29. G. Ertl, *Angew. Chem. Int. Ed.* **47**, 3524–3535 (2008).
30. F. J. M. Horn, R. C. Lin, *Ind. Eng. Chem. Process Des. Dev.* **6**, 21–30 (1967).
31. J. Wolff, A. G. Papathanasiou, I. G. Kevrekidis, H. H. Rotermund, G. Ertl, *Science* **294**, 134–137 (2001).
32. V. Gorodetskii, J. H. Block, W. Drachsel, *Appl. Surf. Sci.* **76–77**, 129–135 (1994).
33. T. Visart de Bocarmé, T.-D. Chau, N. Kruse, *Top. Catal.* **39**, 111–120 (2006).
34. C. Barroo, Y. De Decker, T. Visart de Bocarmé, N. Kruse, *Phys. Rev. Lett.* **117**, 144501 (2016).
35. V. Gorodetskii, J. Lauterbach, H. H. Rotermund, J. H. Block, G. Ertl, *Nature* **370**, 276–279 (1994).
36. V. Medvedev, Y. Suchorski, *Surf. Sci.* **364**, L540–L546 (1996).
37. J.-S. McEwen, P. Gaspard, T. Visart de Bocarmé, N. Kruse, *J. Phys. Chem. C Nanomater. Interfaces* **113**, 17045–17058 (2009).
38. Y. Suchorski *et al.*, *J. Phys. Chem. C Nanomater. Interfaces* **123**, 4217–4227 (2019).
39. C. Spiel, D. Vogel, R. Schlögl, G. Rupprechter, Y. Suchorski, *Ultramicroscopy* **159**, 178–183 (2015).
40. P. W. Murray *et al.*, *Phys. Rev. B Condens. Matter* **47**, 12976–12979 (1993).
41. C. Voss, A. Gaussmann, N. Kruse, *Appl. Surf. Sci.* **67**, 142–146 (1993).
42. G. Hoogers, D. A. King, *Surf. Sci.* **286**, 306–316 (1993).
43. P. Winkler *et al.*, *Nat. Commun.* **12**, 69 (2021).
44. N. Ernst, G. Bozdech, V. Gorodetskii, J. H. Block, *Int. J. Mass Spectrom.* **152**, 185–191 (1996).
45. D. Krefting, P. Kaira, H. H. Rotermund, *Phys. Rev. Lett.* **102**, 178301 (2009).
46. P. S. Bodega *et al.*, *New J. Phys.* **9**, 61 (2007).
47. Y. Suchorski, J. Zeininger, S. Buhr, M. Raab, M. Stöger-Pollach, J. Bernardi, H. Grönbeck, G. Rupprechter, Resolving multifrequential oscillations and nanoscale interfacet communication in single particle catalysis - Database S1. Zenodo (2021).

ACKNOWLEDGMENTS

Funding: This work was supported by the Austrian Science Fund (FWF) (P 32772-N and SFB F45 FOXSI) and by the Swedish Research Council (2020-05191). **Author contributions:** J.Z. and S.B. performed the FEM/FIM experiments, J.Z. and M.R. carried out processing of the FEM/FIM video-files, and M.S.-P. and J.B. made the SEM characterization of the sample. H.G. performed the microkinetic modeling. Y.S. and G.R. supervised the experimental work and were involved in the analysis of the experimental data. Y.S., G.R., and H.G. prepared the manuscript. All authors contributed to the discussion and approved the manuscript and supporting information. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** The FEM and FIM video data (database S1) that support the findings of this study are available as zipped HDF5 files in a Zenodo repository under https://doi.org/10.5281/zenodo.4709843 (*47*). The data used for modeling is detailed in supplementary text S6 and table S1.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/372/6548/1314/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S6
Table S1
References (*48–60*)
Movies S1 to S3

20 November 2020; resubmitted 17 February 2021
Accepted 5 May 2021
10.1126/science.abf8107

QUANTUM SIMULATION

Geometric squeezing into the lowest Landau level

Richard J. Fletcher*, Airlia Shaffer, Cedric C. Wilson, Parth B. Patel, Zhenjie Yan, Valentin Crépel, Biswaroop Mukherjee, Martin W. Zwierlein

The equivalence between particles under rotation and charged particles in a magnetic field relates phenomena as diverse as spinning atomic nuclei, weather patterns, and the quantum Hall effect. For such systems, quantum mechanics dictates that translations along different directions do not commute, implying a Heisenberg uncertainty relation between spatial coordinates. We implement squeezing of this geometric quantum uncertainty, resulting in a rotating Bose-Einstein condensate occupying a single Landau gauge wave function. We resolve the extent of zero-point cyclotron orbits and demonstrate geometric squeezing of the orbits' centers 7 decibels below the standard quantum limit. The condensate attains an angular momentum exceeding 1000 quanta per particle and an interatomic distance comparable to the cyclotron orbit. This offers an alternative route toward strongly correlated bosonic fluids.

In 1851, Foucault directly demonstrated the rotation of Earth via the precession of a pendulum's oscillation axis. This occurs because in the rotating frame, counter- and corotating motions no longer oscillate at the pendulum's natural frequency ω. Instead, their frequencies are increased and decreased, respectively, by Earth's rotation frequency Ω, which leads to the bob performing epicycles (Fig. 1A). In Foucault's experiment, for which $\Omega \ll \omega$, this manifests as an apparent precession of the oscillation axis. If we imagine instead that $\Omega = \omega$, the centrifugal force exactly cancels the restoring force. The pendulum can still perform cyclotron orbits against the frame's rotation, but the motion of the orbit's guiding center is free. In a quantum mechanical description, the energy spectrum is analogous to that of charged particles in a magnetic field and forms discrete Landau levels spaced by $2\hbar\omega$, where $\hbar$ is the reduced Planck constant. The levels correspond to different states of cyclotron motion, each with a large degeneracy arising from the possible guiding-center positions.

An intrinsic characteristic of both neutral particles under rotation and charged particles in a magnetic field is the noncommutativity of space. This can be seen from the quantized Hamiltonian of a pendulum of mass m viewed in the rotating frame,

$$\hat{H} = \frac{\hat{p}_x^2 + \hat{p}_y^2}{2m} + \frac{1}{2}m\omega^2\left(\hat{x}^2 + \hat{y}^2\right) - \Omega\hat{L}_z \quad (1)$$

where $\hat{p}_{x,y}$ are the canonical momenta along x and y, and $\hat{L}_z$ is the axial angular momentum. The rotational term $\Omega\hat{L}_z$ mixes spatial and momentum coordinates into new normal modes, and one decouples Eq. 1 by transforming into cyclotron coordinates $\xi = \frac{1}{2}[x - (p_y/m\omega)]$ and $\eta = \frac{1}{2}[y + (p_x/m\omega)]$, and guiding-center coordinates $X = \frac{1}{2}[x + (p_y/m\omega)]$ and $Y = \frac{1}{2}[y - (p_x/m\omega)]$, yielding

$$\hat{H} = m\omega(\omega + \Omega)\left(\hat{\xi}^2 + \hat{\eta}^2\right) + m\omega(\omega - \Omega)\left(\hat{X}^2 + \hat{Y}^2\right) \quad (2)$$

(*1*). Because $\hat{x} = \hat{X} + \hat{\xi}$ and $\hat{y} = \hat{Y} + \hat{\eta}$, the particle's motion is the sum of a fast cyclotron motion and a slow drift of the guiding center (Fig. 1A). Crucially, while the absolute spatial coordinates $\hat{x}$ and $\hat{y}$ always commute, the two pairs of cyclotron and guiding-center coordinates separately do not. Each pair spans the phase space of a one-dimensional harmonic oscillator, and consequently

$$\left[\hat{\xi}, \hat{\eta}\right] = -\left[\hat{X}, \hat{Y}\right] = i\ell_B^2 \quad (3)$$

where $\ell_B = \sqrt{\hbar/(2m\omega)}$ is the rotational analog of the magnetic length. If an applied potential $\hat{V}(\hat{x}, \hat{y})$ varies little over this length scale, it cannot resolve the cyclotron motion and only couples to the guiding centers. In this case, $\hat{V}(\hat{x}, \hat{y}) \to \hat{V}\left(\hat{X}, \hat{Y}\right)$, and the resulting dynamics occurs within a noncommutative space (*2*).

This noncommutativity of guiding-center motion lies at the heart of the Hall effect. Each spatial variable generates translations in the orthogonal direction, meaning that a force along X effects motion along Y. Particles therefore drift along isopotentials of V with a velocity $\mathbf{v}_d = \boldsymbol{\Omega} \times \nabla V/(2m\Omega\omega)$ in analogy to the $\mathbf{E} \times \mathbf{B}$ drift of electromagnetism. This flow is divergence-free, reflecting the incompressibility of phase-space distributions (*3*), and defines a one-to-one mapping between a particle's initial position and its final position. Time evolution therefore always results in a purely geometric, equiareal transformation of the guiding-center distribution.

MIT-Harvard Center for Ultracold Atoms, Research Laboratory of Electronics, and Department of Physics, Massachusetts Institute of Technology, Cambridge, MA 02139, USA.
***Corresponding author. Email: rfletch@mit.edu**

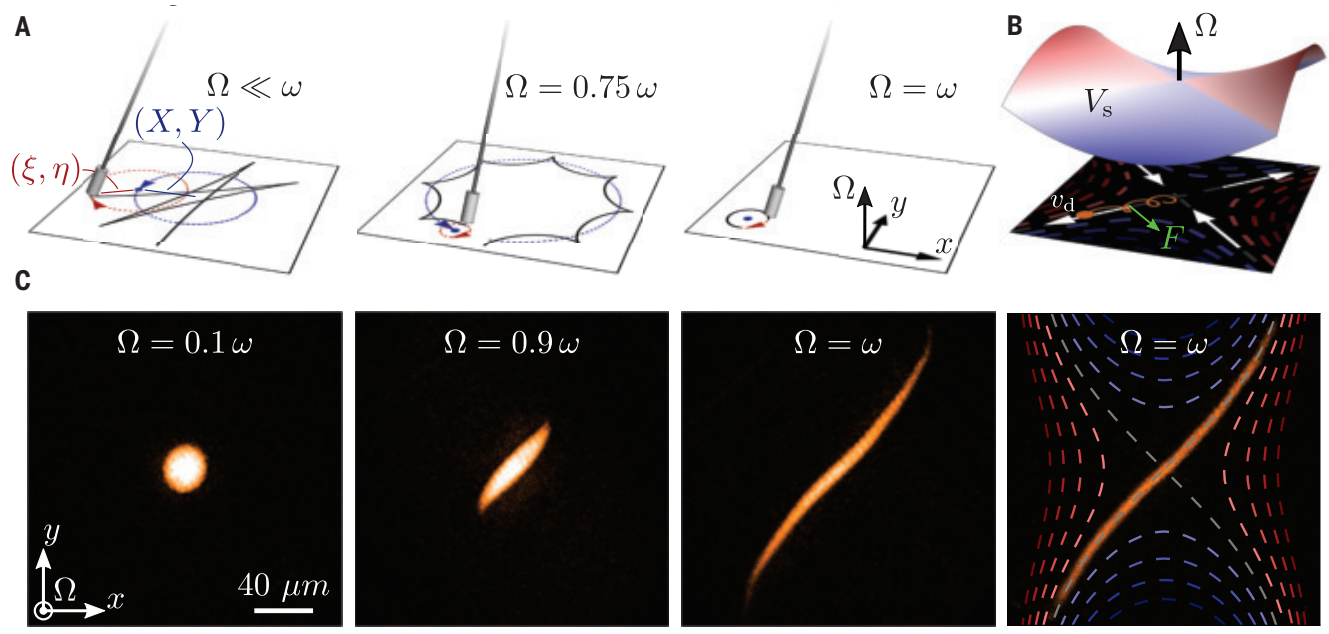


Fig. 1. Geometric squeezing of a rotating Bose-Einstein condensate. (**A**) Viewed in a frame rotating at Ω, the motion of a Foucault pendulum with natural frequency ω separates into a slow corotating drift of the guiding center (*X*, *Y*), in blue, and fast counterrotating cyclotron orbits with relative coordinates (ξ, η), in red. For Ω < ω, the pendulum performs skipping orbits, whereas if Ω = ω, the guiding-center motion is free. (**B**) Atoms in an elliptical harmonic trap rotating at Ω = ω evolve under both a vector potential and a scalar saddle potential V_s, whose isopotentials are shown as red (V_s > 0) and blue (V_s < 0) dashed lines. Particles perform cyclotron orbits whose guiding centers drift along isopotentials with a velocity $\mathbf{v}_d$ (white arrows) orthogonal to the local force $\mathbf{F} = -\nabla V_s$ (green arrow). (**C**) In situ images of the condensate in the rotating frame. During the hold time at Ω = ω, the atoms flow out along one diagonal and in along the other, mediating squeezing of the distribution in guiding-center phase space. The final image is overlaid with the isopotentials of V_s.

These concepts are relevant to atomic nuclei (*4*–*6*), astrophysical phenomena (*7*, *8*), quantum Hall systems (*9*), and ultracold atomic quantum gases, which offer a highly versatile experimental arena for studying rotating quantum fluids (*10*). In Bose-Einstein condensates rotating close to the trap frequency, signatures of the gas approaching the lowest Landau level (LLL) were seen in a softening of the vortex lattice (*11*, *12*). A principal goal is to address the quantum Hall regime, but the exacting requirements on the trap isotropy and rotation speed present a major challenge. Synthetic magnetic fields (*13*–*15*) have also been engineered by other methods, such as spin-orbit coupling (*16*, *17*), and by direct phase-imprinting in both optical lattices (*18*–*21*) and synthetic dimensions (*22*). Experiments have shown a transverse Hall response in lattice transport (*23*) and superfluid collective modes (*24*), as well as chiral edge states in synthetic dimensions (*25*, *26*).

Here, we directly exploited the noncommutativity of guiding-center motion to realize geometric squeezing, cleanly distilling a single Landau gauge wave function in the lowest Landau level (*1*). In comparison to previous work in azimuthally symmetric condensates (*11*), this obviates delicate fine-tuning of trapping and rotation parameters, and offers a complementary "Landau gauge" starting point from which to investigate interaction-driven physics in quantum Hall systems. To begin our experiment, we prepared a condensate of N_{Tot} = 8.1 (±0.1) × 10^5 atoms of ^{23}Na in an elliptical time-orbiting-potential (TOP) trap (*27*), with trap frequencies $(\omega_x, \omega_y, \omega_z) = (\sqrt{1+\varepsilon}, \sqrt{1-\varepsilon}, \sqrt{8})\omega$. Here, ω = 2π × 88.6 (±0.1) Hz and the trap ellipticity is ε = 0.125 ± 0.004. We smoothly ramped the trap's rotation frequency from zero to ω, waited for a variable time *t*, and then obtained an absorption image of the in situ density distribution. Our imaging resolution was sufficient to observe

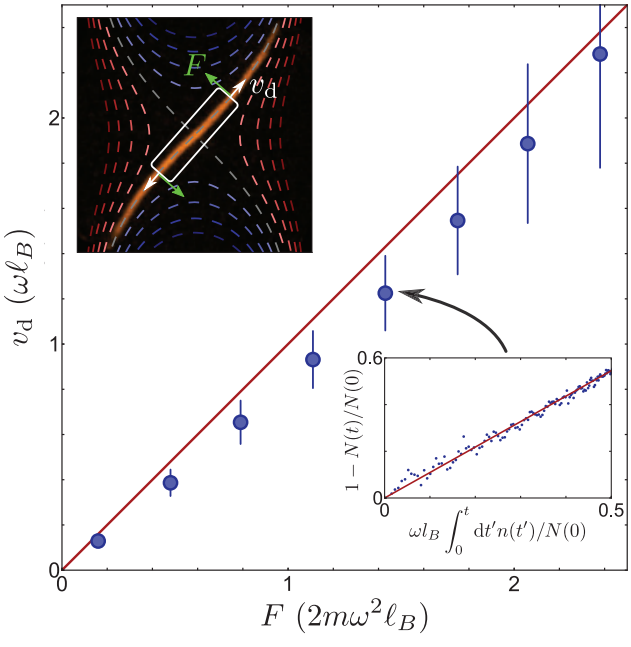


Fig. 2. Isopotential drift velocity. The main plot shows the particles' radial speed v_d in response to the azimuthal force *F*, measured at distances of (left to right) 4.1 μm, 12 μm, 20 μm, 28 μm, 36 μm, 44 μm, 52 μm, and 60 μm from the trap center. The speed is inferred from changes in the atom number *N* inside a bounding box (top inset) and the density *n* at its boundary. The bottom inset shows a typical plot constructed from *N*(*t*) and *n*(*t*), whose slope gives v_d (see text). The data show good agreement with the theoretical expectation (red line) without any free parameters. The force *F* is calculated assuming a harmonic trap, but the quartic corrections to the potential reduce the velocity along *x* = *y*; this results in a small downward shift of the data, which is captured by a GP simulation (*1*). Error bars show the variation in v_d measured across different time intervals.

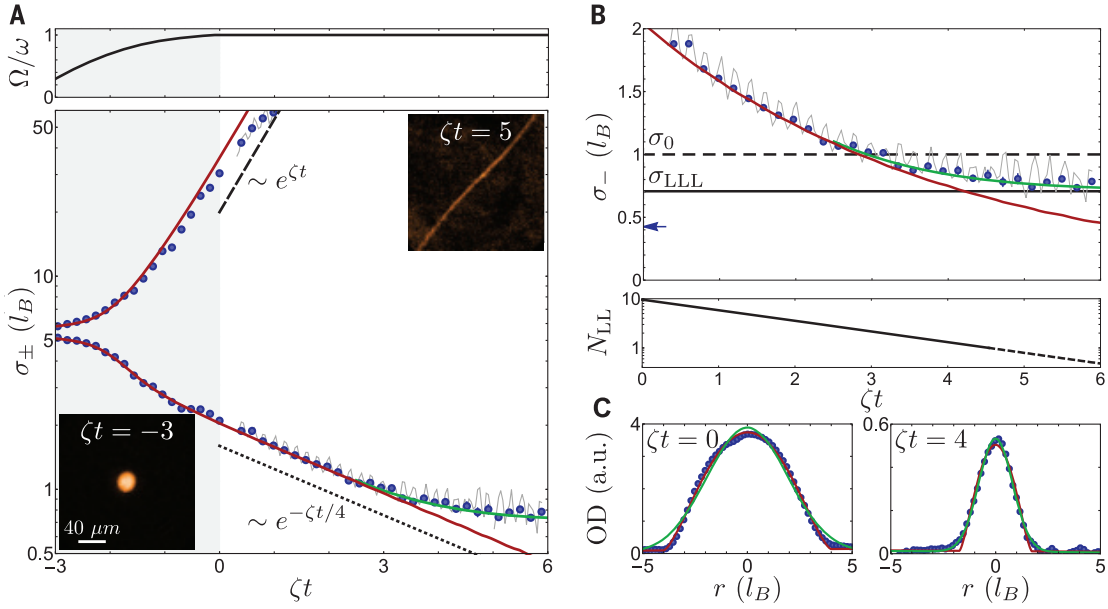


Fig. 3. Squeezing into the lowest Landau level. (**A**) Evolution of the major and minor cloud radii $\sigma_\pm$ in response to ramping the rotational angular frequency Ω of the trap from zero to the trap frequency ω (top). The gray shading shows the time period for which $\Omega < \omega$. The left and right insets show representative in situ images of the cloud at early and late times, respectively. Initially the condensate is approximately isotropic, whereas for long times the spatial aspect ratio exceeds 100. The red line shows the prediction of a hydrodynamic model for which the total atom number is the only free parameter, and whose behavior when $\Omega = \omega$ follows simple scalings shown by the dashed and dotted lines (see text). The green line shows the result of a Gross-Pitaevskii simulation of our experiment (*1*) and captures the deviation from classical hydrodynamic behavior as the LLL is approached. The gray data show a small cyclotron breathing oscillation driven by trap imperfections (see text); the blue points are averaged over one period. (**B**) Top: A zoom-in of the minor width evolution. Bottom: The inferred number of occupied Landau levels $N_{LL} \equiv \mu/(2\hbar\omega)$. As the condensate enters the LLL, its width saturates at $\sigma_{LLL} = \ell_B/\sqrt{2}$, shown by a solid line and corresponding to zero-point cyclotron motion. For comparison, the dashed line shows the width of the two-dimensional harmonic oscillator ground state, $\sigma_0 = \ell_B$. The blue arrow denotes the measured imaging resolution obtained using vortex cores (*1*). (**C**) The transverse optical density (OD) profile of the cloud along with fits of Thomas-Fermi (red) and Gaussian (green) functions. At early times, interactions dominate and the profile is Thomas-Fermi in character, whereas when $N_{LL} \lesssim 1$ we observe a Gaussian shape, which is characteristic of wave functions in the LLL.

vortices in situ with a contrast of ~60% (*1*). These have a characteristic size set by the healing length, which is ~300 nm in our system. This is much smaller than the quantum mechanical ground-state size of cyclotron orbits, set by the magnetic length $\ell_B = 1.6\ \mu m$.

In the frame rotating at Ω, the condensate evolves under both a vector potential and a scalar potential. The vector potential and the associated synthetic magnetic field are induced by the frame rotation (*1*). The scalar potential $V = m\left[(\omega^2 - \Omega^2)(X^2 + Y^2)/2\right] + m\varepsilon\omega^2\left[(X^2 - Y^2)/2\right]$ arises from both the TOP trap and the centrifugal force. For $\Omega/\omega < \sqrt{1-\varepsilon}$, the isopotentials of V are closed; the condensate remains confined but deforms into an ellipse. In earlier experiments, unstable density modulations mediated the nucleation of vortices for rotation frequencies $\Omega/\omega \gtrsim 0.8$ (*28*, *29*). By ramping sufficiently quickly, we precluded breakup of the condensate while allowing its ellipticity to adiabatically follow the equilibrium value (*30*).

When $\Omega = \omega$, the scalar potential forms a saddle $V_s = m\varepsilon\omega^2(X^2 - Y^2)/2$ (Fig. 1B). Without a vector potential, atoms would be lost along the anti-trapped y-direction. Instead, the guiding centers drift outward along the $x = y$ contours and inward along the $x = -y$ contours. This flow (illustrated by white arrows) mediates squeezing of the spatial distribution. In Fig. 1C we show the evolution of the condensate density viewed in the rotating frame. The final image is overlaid with the known isopotentials of V_s, whose coincidence with the atomic density provides a qualitative signature of isopotential drift. The small curvature of the diagonal contours arises from the known quartic corrections $\sim(X^2 + Y^2)^2$ to the trapping potential (*27*), and the spatial twisting of the condensate lies in close analogy to the twisting in optical phase space induced by the Kerr effect (*31*).

To measure the transverse Hall response, we obtained the radial drift speed as a function of the azimuthal force, which at a radius r is $F(r) = m\varepsilon\omega^2 r$. Our measurements are shown in Fig. 2 along with the theoretical relation $v_d = F/(2m\omega)$, valid for any quantum state, which shows good agreement without any free parameters. We used a continuity equation to infer the drift speed; the atom number N inside a bounding box (Fig. 2, white frame in top inset) centered on $r = 0$ and with length $2R$ varies as $\dot{N} = -2v_d n$, where v_d and n are the drift speed and one-dimensional number density at $r = R$. Integrating once gives $1 - [N(t)/N(0)] = 2v_d \int_0^t dt'[n(t')/N(0)]$, allowing straightforward evaluation of v_d as shown in the lower inset. This method offers a convenient protocol for measuring the Hall response of any fluid.

Whereas the drift velocity determines the local response to a force, the specific geometric transformation of the cloud depends upon the global shape of V_s. Qualitatively, isopotential flow on a saddle in the presence of a magnetic field results in elongation and contraction along orthogonal diagonals. More quantitatively, in terms of the oscillator ladder operators $\hat{a} = \sqrt{m\omega/\hbar}\left(\hat{\xi} + i\hat{\eta}\right)$ and $\hat{b} = \sqrt{m\omega/\hbar}\left(\hat{X} - i\hat{Y}\right)$, the single-particle Hamiltonian is

$$\hat{H}_s \approx 2\hbar\omega\left(\hat{a}^\dagger\hat{a} + \frac{1}{2}\right) + \frac{\hbar\zeta}{2}\left(\hat{b}\hat{b} + \hat{b}^\dagger\hat{b}^\dagger\right) \quad (4)$$

(*1*), where we define $\zeta = \varepsilon\omega/2$. Comparison with the one-mode squeezing operator $\hat{S}(\alpha) = \exp\left[\left(\alpha^*\hat{b}\hat{b} - \alpha\hat{b}^\dagger\hat{b}^\dagger\right)/2\right]$ reveals that time evolution under a saddle potential is equivalent to fully coherent squeezing of the guiding-center phase space distribution, analogous to phase squeezing in quantum optics (*32*–*34*). Consistent with the perspective based on isopotential flow, the imaginary squeezing parameter $\alpha = i\zeta t$ describes dilation of the cloud along the diagonals of phase space by factors $\exp(\pm\zeta t)$.

In the limit $\zeta t \gg 1$, the particles' guiding centers become widely distributed along one diagonal and sharply localized along the other. The residual transverse width of the cloud solely arises from the unsqueezed cyclotron orbits, which have a size $\sqrt{\langle\hat{\xi}^2\rangle} = \ell_B\sqrt{\nu+1/2}$ in the νth Landau level. The minimum orbit size

$\sigma_{LLL} = \ell_B/\sqrt{2}$ occurs in the LLL, where the cyclotron wave function is Gaussian and saturates the Heisenberg uncertainty relation $\Delta\xi\Delta\eta \geq \ell_B^2/2$. The density of any condensate in the LLL is therefore a convolution of the guiding-center distribution with a Gaussian of width σ_{LLL}. In the quantum optics analogy, this directly realizes the Husimi-Q representation of the guiding-center Wigner function (*1*). In our case, at long times the cloud is an extended strip of transverse width σ_{LLL}. Geometric squeezing therefore coherently transforms the condensate into a single Landau gauge wave function within the LLL (*1*).

In Fig. 3A we show images of the condensate before and after squeezing and plot the major and minor cloud widths, $\sigma_{\pm}$, which are defined as the $e^{-1/2}$ radii obtained from a Gaussian fit. Initially the chemical potential is $\mu_0 \approx h \times 3.4$ kHz, where h is the Planck constant, and the number of Landau levels admixed into the condensate wave function is $\sim\mu_0/(2\hbar\omega) \approx 20$, hence the evolution is well described by a hydrodynamic model that neglects quantum pressure (*29*). The prediction of this model is shown by the red line, for which the only free parameter is the total atom number (*35*).

For times $t > 0$, the cloud evolves under the squeezing Hamiltonian of Eq. 3 and the major width increases as $\sigma_+ \propto \exp(\zeta t)$, illustrated by the dashed line. However, the minor width decays more slowly. This difference arises because the condensate size contains contributions from both the guiding centers, which are squeezed at a rate ζ, and the cyclotron orbits, whose size depends on the number of occupied Landau levels $N_{LL} \equiv \mu/(2\hbar\omega)$. In our experiment, σ_- is generally dominated by cyclotron motion and its evolution is captured well by a simple scaling model. The chemical potential is proportional to the atomic number density $\sim N_{Tot}/(\sigma_+\sigma_-\sigma_z)$, where σ_z is the axial extent of the condensate. The major width always increases as $\sigma_+ \propto \exp(\zeta t)$, and $\sigma_{-,z} \propto \sqrt{\mu}$ when $N_{LL} \gg 1$. We therefore predict a time dependence $\sigma_- \propto \exp(-\zeta t/4)$ at early times, which is shown by the dotted line in Fig. 3A. The gray data show a small breathing of the cloud at the cyclotron frequency 2ω. This is driven by imperfections in the trap, which shows a ~0.3% root-mean-square variation in ω with ellipse orientation, giving a perturbation in the rotating frame with a frequency 2Ω. The blue points are averaged over one period.

The falling chemical potential $\mu \propto \exp(-\zeta t/2)$ guarantees that eventually $\mu < 2\hbar\omega$ and the condensate enters the LLL. As shown in Fig. 3B, we directly observe the saturation of σ_- at the zero-point cyclotron width σ_{LLL} imposed by Heisenberg uncertainty. Because the hydrodynamic model neglects quantum pressure, it predicts that $\sigma_- \to 0$. On the other hand, the saturation of the cloud width is captured very well by a Gross-Pitaevskii simulation with no free parameters (green solid line) (*1*). For comparison, the dashed line shows the width $\sigma_0 = \ell_B$ of the noninteracting harmonic oscillator ground state, which corresponds to minimal, but isotropic, Heisenberg uncertainty in both cyclotron and guiding-center coordinates. This lies above our data at long times, and from the last five data points we infer squeezing of the guiding centers by >7 dB below the standard quantum limit.

In the lower panel of Fig. 3B, we plot the number of occupied Landau levels, N_{LL}, inferred from the central density evaluated using the fitted hydrodynamic model. We indeed

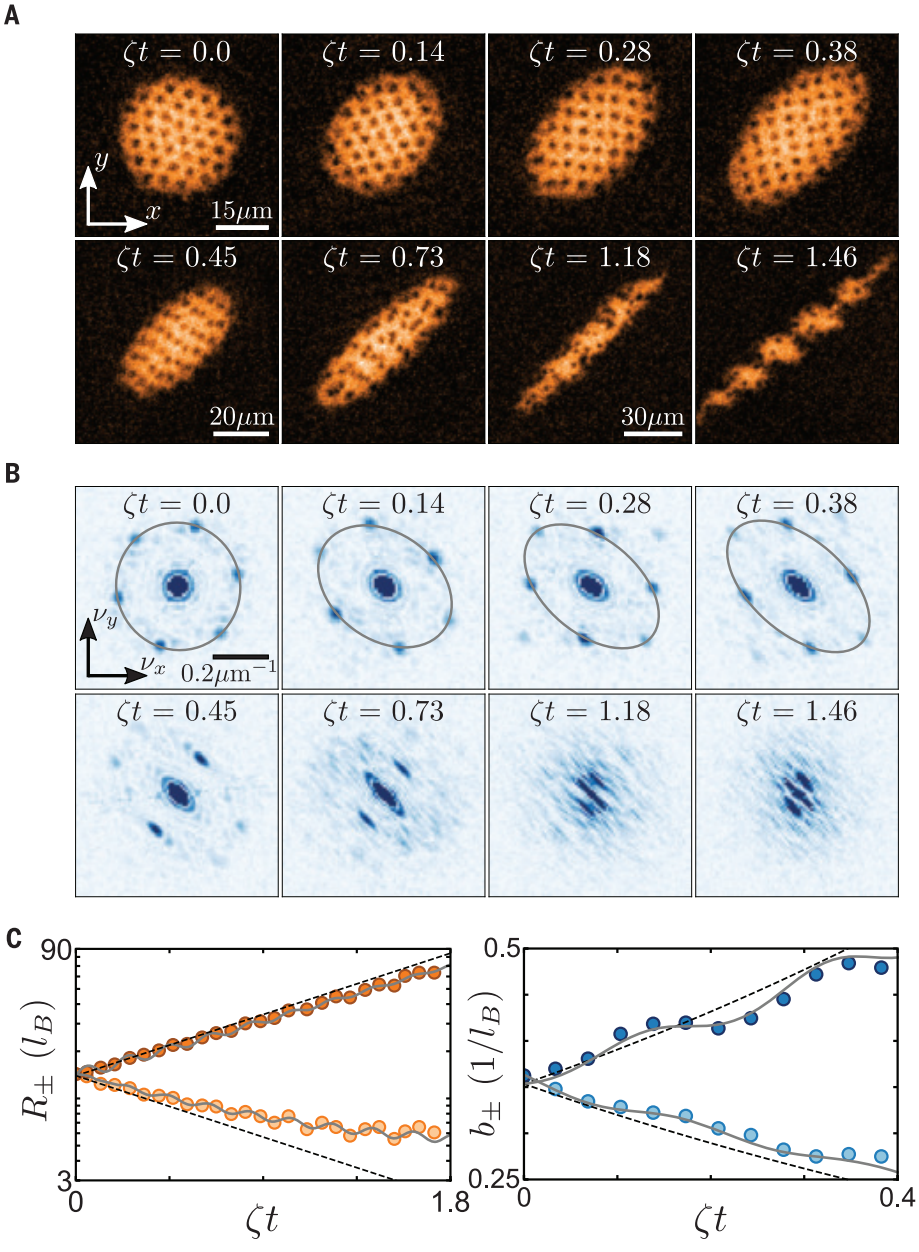


Fig. 4. Squeezing of a vortex lattice. (**A** and **B**) In situ evolution in real space (A) and reciprocal space (B) after suddenly applying the rotating saddle. Initially the cloud is round, and the reciprocal lattice vectors lie on a circle. Squeezing is evident in both the condensate spatial envelope and the vortex lattice spacing. At longer times, clustering of vortices causes the condensate to break up into droplets. (**C**) Time evolution of the major and minor Thomas-Fermi radii of the condensate, $R_\pm$, and the major and minor radii of the ellipse describing the lattice vectors, $b_\pm$. The black dashed lines show exponential functions $A\exp(\pm\zeta t)$, with A fixed by the data at $t = 0$; the solid lines include the small contributions of quadrupolar collective modes and the nonzero size of the cyclotron orbits (*1*). The longest squeezing time $\zeta t = 1.8$ corresponds to $t \approx 50$ ms. The ellipse widths in reciprocal space are shown for times for which the distribution of vortices remains periodic.

find that the crossover to LLL behavior occurs for $N_{LL} \sim 1$; the dashed region corresponds to $N_{LL} < 1$, where the hydrodynamic model is not applicable and this inference is no longer self-consistent. We also see a qualitative change in the shape of the cloud, which changes from a Thomas-Fermi to a Gaussian profile. This is shown in Fig. 3C, where we plot cuts along $x = -y$ at early and late times. If $N_{LL} \gg 1$, the healing length is much smaller than the magnetic length and the density profile is a Thomas-Fermi function (red line). On the other hand, if $N_{LL} < 1$, the profile is Gaussian (green line), reflecting the cyclotron ground state. We note that at our latest times, the interparticle distance has grown to ~500 nm, close to half the size of a zero-point cyclotron orbit ~σ_{LLL}. This signals the approach of the Bose gas toward the strongly correlated regime (*10*, *15*, *36*–*40*).

Microscopically, the squeezing operator mixes higher–angular momentum states into the condensate wave function, in analogy to the admixing of higher Fock states in squeezed light (*32*). In general, the angular momentum of a superfluid can be carried either by vortices or by deformations that break rotational symmetry (*41*). Because $\nabla \times \mathbf{v}_d = 0$, the induced flow is irrotational, but the large aspect ratio gives a moment of inertia $\Theta = mN_{Tot}\left[\left(\sigma_+^2 - \sigma_-^2\right)^2/\left(\sigma_+^2 + \sigma_-^2\right)\right] \approx mN_{Tot}\sigma_+^2$ which is close to the rigid-body value (*41*). For clouds with $\sigma_+ > 50\ell_B$, this gives a per-particle angular momentum $\langle l_z \rangle > 1000\hbar$ despite the absence of any vortices inside the condensate (*42*).

In the experiments above, geometric squeezing was seen in the evolution of the condensate widths. To directly observe the drift velocity field inside the cloud, we now introduce a dilute gas of vortices that correspond to nodes in the atomic wave function and can serve as "tracer particles" for the local flow. We prepare a ground-state condensate rotating at 0.8ω in an isotropic trap and instantaneously apply the saddle V_s rotating at $\Omega = \omega$. The initial chemical potential is $\mu \approx h \times 2.2$ kHz, giving a cyclotron orbit size $\sim\sqrt{\mu/(2\hbar\omega)}\ell_B = 5.5$ μm. This is much smaller than the cloud's Thomas-Fermi radius of 21 μm, hence the observed width is dominated by the guiding-center distribution. In Fig. 4, A and B, we show the in situ evolution in both real and reciprocal space. Initially, the condensate is circular and contains a triangular Abrikosov lattice with six-fold–symmetric reciprocal lattice vectors. Subsequently, squeezing is evident in both the cloud shape and the vortex lattice. Because the vortices are distributed throughout the whole cloud, this indicates that the coordinates of all particles evolve under the same squeezing transformation. For longer times, while the overall spatial envelope continues to squeeze, the density profile exhibits an intricate evolution. Squeezing of the initially triangular vortex lattice eventually leads to the formation of vortex rows (*43*, *44*). Subsequently, a hydrodynamic instability drives amalgamation of the vortices into clusters, as well as an intriguing fragmentation of the condensate into a persistent array of droplets.

In Fig. 4C, we show the evolution of the major and minor Thomas-Fermi radii of the cloud, $R_\pm$, and the major and minor radii of an ellipse fitted to the reciprocal lattice vectors, $b_\pm$. The dashed lines show exponential functions $A\exp(\pm\zeta t)$ (where the amplitude A is the only free parameter), which capture the initial evolution well. This confirms both the expected rate of squeezing and the incompressibility of the guiding-center distribution. The solid lines show a fit that includes the excitation of quadrupolar collective modes by the saddle turn-on and trap imperfections (see above) and additionally accounts for the nonzero cyclotron orbit size (*1*). This results in a slight reduction of the apparent squeezing rate and a slowdown of the decay in R_- as the guiding-center width approaches the cyclotron size.

The geometric squeezing protocol established here offers an alternative route to LLL physics in quantum gases. Crucially, simply turning off the saddle potential halts the outward flow of atoms. This controllably prepares an equilibrium condensate (*1*), which occupies a single Landau gauge wave function whose purely interaction-driven evolution in the flat single-particle dispersion of the LLL can then be cleanly observed. Natural immediate directions concern the collective excitation spectrum (*45*), quantum hydrodynamic stability, and the appearance of strongly correlated bosonic states (*36*–*40*, *42*, *46*). More generally, the ability to resolve cyclotron motion and vortices in situ allows the study of chiral edge states and quantum turbulence in rotating gases. From a metrology perspective, azimuthally squeezed condensates might offer benefits for rotation sensing and a route to spin-squeezing via a spatially dependent coupling between internal atomic states (*47*).

REFERENCES AND NOTES

1. See supplementary materials.
2. M. Wilkinson, *J. Phys. Math. Gen.* **20**, 1761–1771 (1987).
3. Semiclassically, this incompressibility is an instance of Liouville's theorem; the quantum analog of a phase-space distribution is provided by the Wigner function, which is incompressible in the case of a quadratic Hamiltonian (*48*).
4. J. G. Valatin, *Proc. R. Soc. London Ser. A* **238**, 132–141 (1956).
5. A. Bohr, *Rev. Mod. Phys.* **48**, 365–374 (1976).
6. B. R. Mottelson, *Rev. Mod. Phys.* **48**, 375–383 (1976).
7. R. P. Kerr, *Phys. Rev. Lett.* **11**, 237–238 (1963).
8. A. Maeder, G. Meynet, *Rev. Mod. Phys.* **84**, 25–63 (2012).
9. F. D. M. Haldane, *Phys. Rev. Lett.* **107**, 116801 (2011).
10. N. Cooper, *Adv. Phys.* **57**, 539–616 (2008).
11. V. Schweikhard, I. Coddington, P. Engels, V. P. Mogendorff, E. A. Cornell, *Phys. Rev. Lett.* **92**, 040404 (2004).
12. V. Bretin, S. Stock, Y. Seurin, J. Dalibard, *Phys. Rev. Lett.* **92**, 050403 (2004).
13. J. Dalibard, F. Gerbier, G. Juzeliūnas, P. Öhberg, *Rev. Mod. Phys.* **83**, 1523–1543 (2011).
14. N. Goldman, G. Juzeliūnas, P. Öhberg, I. B. Spielman, *Rep. Prog. Phys.* **77**, 126401 (2014).
15. V. Galitski, G. Juzeliūnas, I. B. Spielman, *Phys. Today* **72**, 38–44 (2019).
16. Y.-J. Lin, R. L. Compton, K. Jiménez-García, J. V. Porto, I. B. Spielman, *Nature* **462**, 628–632 (2009).
17. V. Galitski, I. B. Spielman, *Nature* **494**, 49–54 (2013).
18. J. Struck *et al*., *Phys. Rev. Lett.* **108**, 225304 (2012).
19. M. Aidelsburger *et al*., *Phys. Rev. Lett.* **111**, 185301 (2013).
20. H. Miyake, G. A. Siviloglou, C. J. Kennedy, W. C. Burton, W. Ketterle, *Phys. Rev. Lett.* **111**, 185302 (2013).
21. G. Jotzu *et al*., *Nature* **515**, 237–240 (2014).
22. A. Celi *et al*., *Phys. Rev. Lett.* **112**, 043001 (2014).
23. M. Aidelsburger *et al*., *Nat. Phys.* **11**, 162–166 (2014).
24. L. J. LeBlanc *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **109**, 10811–10814 (2012).
25. B. K. Stuhl, H.-I. Lu, L. M. Aycock, D. Genkina, I. B. Spielman, *Science* **349**, 1514–1518 (2015).
26. M. Mancini *et al*., *Science* **349**, 1510–1513 (2015).
27. W. Petrich, M. H. Anderson, J. R. Ensher, E. A. Cornell, *Phys. Rev. Lett.* **74**, 3352–3355 (1995).
28. K. W. Madison, F. Chevy, V. Bretin, J. Dalibard, *Phys. Rev. Lett.* **86**, 4443–4446 (2001).
29. S. Sinha, Y. Castin, *Phys. Rev. Lett.* **87**, 190402 (2001).
30. A. Recati, F. Zambelli, S. Stringari, *Phys. Rev. Lett.* **86**, 377–380 (2001).
31. G. Kirchmair *et al*., *Nature* **495**, 205–209 (2013).
32. R. Loudon, P. Knight, *J. Mod. Opt.* **34**, 709–759 (1987).
33. H. A. Fertig, B. I. Halperin, *Phys. Rev. B* **36**, 7969–7976 (1987).
34. S. Vishveshwara, N. R. Cooper, *Phys. Rev. B* **81**, 201306 (2010).
35. The hydrodynamic theory of (*29*) assumes a Thomas-Fermi cloud shape of radius R. A Thomas-Fermi profile has a fitted Gaussian $e^{-1/2}$ radius of $0.46R$, which is the quantity shown by the red curve in Fig. 3A for direct comparison to our data.
36. N. R. Cooper, N. K. Wilkin, J. M. F. Gunn, *Phys. Rev. Lett.* **87**, 120405 (2001).
37. S. Sinha, G. V. Shlyapnikov, *Phys. Rev. Lett.* **94**, 150401 (2005).
38. X. Chen, Z.-C. Gu, Z.-X. Liu, X.-G. Wen, *Science* **338**, 1604–1606 (2012).
39. T. Senthil, M. Levin, *Phys. Rev. Lett.* **110**, 046801 (2013).
40. A. Vishwanath, T. Senthil, *Phys. Rev. X* **3**, 011016 (2013).
41. F. Zambelli, S. Stringari, *Phys. Rev. A* **63**, 033602 (2001).
42. A. Aftalion, X. Blanc, N. Lerner, *Phys. Rev. A* **79**, 011603 (2009).
43. P. Engels, I. Coddington, P. C. Haljan, E. A. Cornell, *Phys. Rev. Lett.* **89**, 100403 (2002).
44. M. Cozzini, S. Stringari, *Phys. Rev. A* **67**, 041602 (2003).
45. S. M. Girvin, A. H. MacDonald, P. M. Platzman, *Phys. Rev. Lett.* **54**, 581–583 (1985).
46. M. O. Oktel, *Phys. Rev. A* **69**, 023618 (2004).
47. D. J. Wineland, J. J. Bollinger, W. M. Itano, D. J. Heinzen, *Phys. Rev. A* **50**, 67–88 (1994).
48. J. E. Moyal, *Math. Proc. Camb. Philos. Soc.* **45**, 99–124 (1949).
49. R. J. Fletcher *et al*., Replication data for "Geometric squeezing into the lowest Landau level". Harvard Dataverse (2021); DOI: 10.7910/DVN/I2PEGZ.

ACKNOWLEDGMENTS

We thank T. Yefsah and J. Struck for early contributions to the planning and construction of the lab space and apparatus and L. Kendrick and J. T. Chamberlin for experimental assistance. **Funding:** Supported by NSF (Center for Ultracold Atoms award PHY-1734011 and award PHY-1506019), Air Force Office of Scientific Research (FA9550-16-1-0324 and MURI Quantum Phases of Matter FA9550-14-1-0035), Office of Naval Research (N00014-17-1-2257), the DARPA A-PhI program through ARO grant W911NF-19-1-0511, the David and Lucile Packard Foundation, and the Vannevar Bush Faculty Fellowship. Also supported by a MIT Pappalardo Fellowship (R.J.F.) and by NSF GRFP (A.S.). **Author contributions:** R.J.F. and A.S. performed the experimental measurements and data analysis; R.J.F., V.C., B.M., and M.W.Z. developed the theoretical description and numerical calculations; R.J.F., A.S., C.C.W., P.B.P., Z.Y., and B.M. constructed the experimental apparatus; and R.J.F. and M.W.Z. supervised the study. All authors contributed to the interpretation of the data and the preparation of the manuscript. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** All data shown in this work can be found at Harvard Dataverse (*49*).

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/372/6548/1318/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S4
References (*50*–*59*)

28 December 2019; resubmitted 3 July 2020
Accepted 11 May 2021
10.1126/science.aba7202

MAGNETISM

Imaging orbital ferromagnetism in a moiré Chern insulator

C. L. Tschirhart[1]†, M. Serlin[1]†, H. Polshyn[1], A. Shragai[1]‡, Z. Xia[1]‡, J. Zhu[1]‡, Y. Zhang[1], K. Watanabe[2], T. Taniguchi[3], M. E. Huber[4], A. F. Young[1]*

Electrons in moiré flat band systems can spontaneously break time-reversal symmetry, giving rise to a quantized anomalous Hall effect. In this study, we use a superconducting quantum interference device to image stray magnetic fields in twisted bilayer graphene aligned to hexagonal boron nitride. We find a magnetization of several Bohr magnetons per charge carrier, demonstrating that the magnetism is primarily orbital in nature. Our measurements reveal a large change in the magnetization as the chemical potential is swept across the quantum anomalous Hall gap, consistent with the expected contribution of chiral edge states to the magnetization of an orbital Chern insulator. Mapping the spatial evolution of field-driven magnetic reversal, we find a series of reproducible micrometer-scale domains pinned to structural disorder.

In crystalline solids, orbital magnetization arises from the Berry curvature of the bands and intrinsic angular momentum of the Bloch electron wave packet (*1*). Although the orbital magnetization often contributes—at times substantially (*2*)—to the net magnetization of ferromagnets, all known ferromagnetism involves partial or full polarization of the electron spin. Theoretically, however, ferromagnetism can also arise through the spontaneous polarization of orbital magnetization without involvement of the electron spin. Recently, hysteretic transport consistent with ferromagnetic order has been observed in heterostructures composed of graphene and hexagonal boron nitride (hBN) (*3*–*7*), neither of which are intrinsically magnetic materials. Notably, spin-orbit coupling is thought to be vanishingly small in these systems (*8*), effectively precluding a spin-based mechanism. These results have consequently been interpreted as evidence for purely orbital ferromagnetism (*9*–*16*).

To host purely orbital ferromagnetic order, a system must have a time-reversal-symmetric electronic degree of freedom separate from the electron spin as well as strong electron-electron interactions. Both of these qualities are present in graphene heterostructures, where the valley degree of freedom provides degenerate electron species related by time-reversal symmetry and a moiré superlattice can be used to engineer strong interactions. In these materials, a long-wavelength moiré pattern, arising from interlayer coupling between mismatched lattices, modulates the underlying electronic structure and leads to the emergence of superlattice minibands within a reduced Brillouin zone. The small Brillouin zone means that low electron densities are sufficient to dope the two-dimensional (2D) system to full filling or depletion of the superlattice bands, which can be achieved using experimentally realizable electric fields (*17*). For appropriately chosen constituent materials and interlayer rotational alignment, the lowest energy bands can have bandwidths considerably smaller than the native scale of electron-electron interactions, $E_C \approx e^2/\lambda_M$, where λ_M is the moiré period and e is the electron's charge. The dominance of interactions typically manifests experimentally through the appearance of "correlated insulators" at integer electron or hole filling of the moiré unit cell (*18*, *19*), consistent with interaction-induced breaking of one or more of the spin, valley, or lattice symmetries. Orbital magnets are thought to constitute a subset of these states, in which exchange interactions favor a particular order that breaks time-reversal symmetry by causing the system to polarize into one or more valley-projected bands. Notably, the large Berry curvature endows the valley-projected bands with a finite Chern number (*20*, *21*), so that valley polarization naturally leads to a quantized anomalous Hall effect at integer band filling. To date, quantum anomalous Hall effects have been observed at band fillings $\nu = 1$ and $\nu = 3$ in various heterostructures (*4*–*6*), where $\nu = An$ corresponds to the number of electrons per unit cell area A with carrier density n.

Although orbital magnetism is generally expected theoretically in twisted bilayer graphene (tBLG) (*10*–*12*), no direct experimental probes of magnetism have been reported, because of the relative scarcity of magnetic samples, their small size, and the low expected magnetization density. The resulting magnetization density $m \lesssim 0.1\ \mu_B/\mathrm{nm}^2$ (*22*) (where $\mu_B \approx 0.06$ meV/T is the Bohr magneton) is consequently more than three orders of magnitude smaller than in typical magnetic systems with several spins per subnanometer-sized crystal unit cell. The absence of magnetic studies leaves open both quantitative questions, such as the magnitude of the orbital magnetization, and qualitative ones regarding the nature of the magnetic phase transitions as a function of magnetic field and carrier density.

In this study, we perform spatially resolved magnetometry to image the submicrometer magnetic structure of the same sample presented in (*4*) (see Fig. 1A), which consists of a twisted graphene bilayer aligned to one of the hexagonal boron nitride–encapsulating layers. Figure 1B shows a schematic representation of our experimental setup. We use a superconducting quantum interference device (SQUID) fabricated on the tip of a quartz tube from cryogenically deposited indium (*23*) with a magnetic field sensitivity of ~15 $\mathrm{nT/Hz^{1/2}}$ at select out-of-plane magnetic fields of <50 mT (see fig. S1). The SQUID is mounted to a quartz tuning fork (*24*, *25*) (see fig. S2) and rastered in a 2D plane parallel to, and at a fixed height above, the tBLG heterostructure. A finite electrical excitation applied to the tuning fork generates a lateral oscillation of the tip along vector $\hat{\mathbf{a}}$, and we measure the SQUID response at the tuning fork oscillation frequency, $B_{TF} \approx \hat{\mathbf{a}} \cdot \vec{\nabla}_r B_z$ (see fig. S3).

Figure 1, C and D, shows images of B_{TF} taken while the sample is doped to $n = 2.36 \times 10^{12}\ \mathrm{cm^{-2}}$, near the quantized Hall plateau corresponding to $\nu = 3$. Here, n is the nominal density inferred from a parallel plate capacitor model, with the capacitance determined from the low-field Hall density (*4*). Images are acquired in the same background magnetic field $B = 22$ mT but on opposite branches of the hysteresis loop shown in Fig. 1A. As discussed in (*24*) and fig. S7, the measured B_{TF} contains contributions from magnetic signals as well as other effects arising from electric fields or thermal gradients (*24*). To isolate the magnetic structure that gives rise to the observed hysteretic transport, we subtract the data in Fig. 1C and Fig. 1D from each other. The result is shown in Fig. 1E, which depicts the gradient magnetometry signal associated with the fully polarized orbital ferromagnet. To reconstruct the static out-of-plane magnetic field, B_z, we integrate B_{TF} along $\hat{\mathbf{a}}$ from the lower and left boundaries of the image (Fig. 1F). We infer the total magnetization density m from the B_z data using standard Fourier domain techniques (*24*), as shown in Fig. 1G. Figure 1, H and I, shows a comparison of B_{TF} and m plotted along the contours indicated in Fig. 1, E and G. The shaded regions in Fig. 1I denote absolute

[1]Department of Physics, University of California, Santa Barbara, CA 93106, USA. [2]Research Center for Functional Materials, National Institute for Materials Science, 1-1 Namiki, Tsukuba 305-0044, Japan. [3]International Center for Materials Nanoarchitectonics, National Institute for Materials Science, 1-1 Namiki, Tsukuba 305-0044, Japan. [4]Departments of Physics and Electrical Engineering, University of Colorado Denver, Denver, CO 80217, USA.
*Corresponding author. Email: andrea@physics.ucsb.edu
†These authors contributed equally to this work.
‡Present address: Department of Physics, Cornell University, Ithaca, NY 14853, USA.

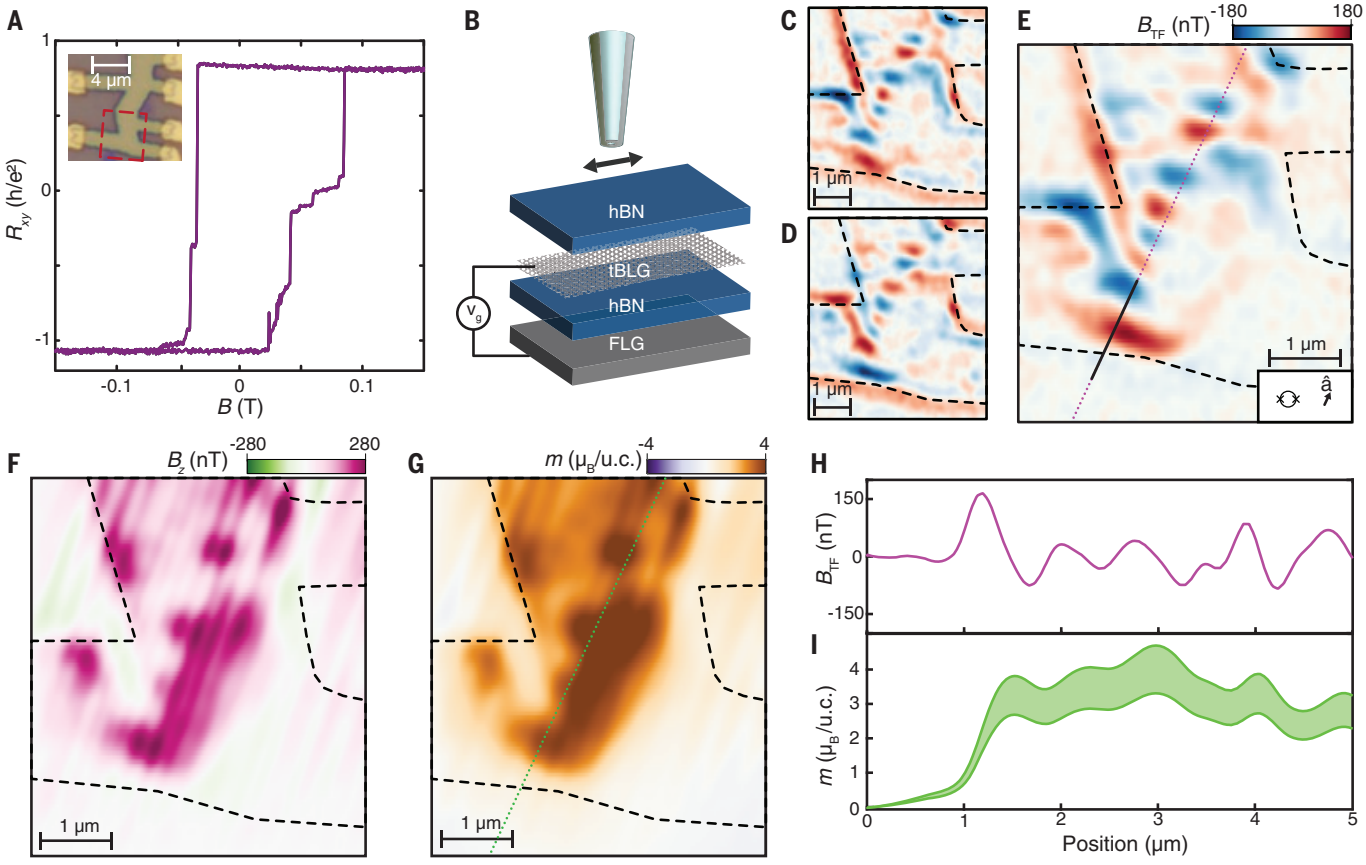


Fig. 1. Imaging orbital ferromagnetism. (**A**) Hall resistance of a twisted bilayer graphene device aligned to hexagonal boron nitride measured at T = 2.1 K and nominal electron density $n = 2.36 \times 10^{12}$ cm^{-2}. A thorough characterization of transport in this exact device is reported in (*4*). Inset: Optical micrograph of the device with the scan region marked in red. Scale bar, 4 μm. (**B**) Schematic illustration of the experimental setup, with hexagonal boron nitride (hBN) dielectric, tBLG (twisted bilayer graphene), and few-layer graphene (FLG) layers labeled. We raster a nanoscale indium SQUID with diameter of d = 215 nm at a height of $h \approx$ 153 nm above the plane of the twisted bilayer graphene. The SQUID is coupled to a quartz tuning fork whose excitation causes the tip to oscillate mechanically at $f_{TF} \approx$ 33 kHz in the plane of the sample. The SQUID response at this frequency $B_{TF} \approx \hat{\mathbf{a}} \cdot \vec{\nabla}_r B_z$, where B_z is the static magnetic field and $|\hat{\mathbf{a}}| \approx$ 190 nm is the tuning fork oscillation amplitude. (**C**) B_{TF} measured at B = 22 mT after field training to +200 mT and (**D**) −200 mT on the area boxed in red in the inset of (A). The black dashed lines in (C) to (G) indicate the edges of the sample. (**E**) Half the difference between data shown in (C) and (D). (C) to (E) share the same color scale. Inset: To-scale representations of the SQUID diameter and tuning fork amplitude $\hat{a}$. (**F**) B_z as determined by integrating data in (E). We assume B_z = 0 along the bottom and left edges of the scan range. (**G**) Magnetization density m. Data are presented in units of Bohr magnetons per moiré unit cell (u.c.) area $A \approx$ 130 nm^2 (*4*). (**H**) B_{TF} and (**I**) m plotted along the line segments indicated in corresponding colors in (E) and (G). The shaded regions in (H) show absolute bounds on m obtained by propagating the systematic uncertainty in $|\hat{\mathbf{a}}|$ (*24*).

error bounds from the dominant systematic uncertainty in $|\hat{\mathbf{a}}|$.

Our measurements are taken close to ν = 3, equivalent to a single hole per unit cell relative to the nonmagnetic state at ν = 4 that corresponds to full filling of the lowest energy bands. We find that the magnetization density is considerably larger than 1 μ_B per unit cell area $A \approx 130$ nm^2, where we have taken a g-factor of 2 as appropriate for graphene, in which spin orbit coupling is negligible. Without any assumptions about the nature of the broken symmetries, this state has a maximum spin magnetization of 1 μ_B per moiré unit cell. Our data reject this hypothesis, finding instead a maximum magnetization density of m in the range of 2 to 4 μ_B per moiré unit cell, corresponding to an orbital magnetization of 1.8×10^{-4} to 3.6×10^{-4} μ_B per carbon atom. We conclude that the magnetic moment associated with the quantum anomalous Hall phase in tBLG is dominated by its orbital component.

In an intrinsic orbital magnet in which all moments arise from conduction electrons, the magnetization depends strongly on the density. Additional density dependence arises from the fact that contributions to the orbital magnetization from both wave-packet angular momentum and Berry curvature need not be uniformly distributed within the Brillouin zone (*1*). Transport observations of a quantum anomalous Hall effect measure only the total Berry curvature of a completely filled band. At partial band filling, however, extrinsic contributions from scattering complicate the relationship between transport and band properties. In contrast, measuring m provides direct information about the density-dependent occupation of the Bloch states in momentum space. Figure 2A shows repeated measurements of B_{TF} for a series of gate voltages in the vicinity of ν = 3. B_{TF} is measured along a contour that runs over a region of the device showing magnetic inhomogeneity even at the saturation magnetization (Fig. 1E, solid black line; see also Fig. 1H). Assuming that the saturated magnetic state has density-independent spatial structure [i.e., $m(n, r) = \mu_B K(n) L(r)$, where K and L are functions of density n and position r, respectively], the amplitude of the position-dependent modulation of B_{TF} functions as a proxy for m.

To compare magnetization at different n, we fit the data for $n = 2.57 \times 10^{12}$ cm^{-2} to a seventh-order polynomial (see Fig. 2A); all other curves are then fit to the same polynomial with an overall scale factor, which we denote $\tilde{B}_{TF}$ and plot in Fig. 2B. Error bars

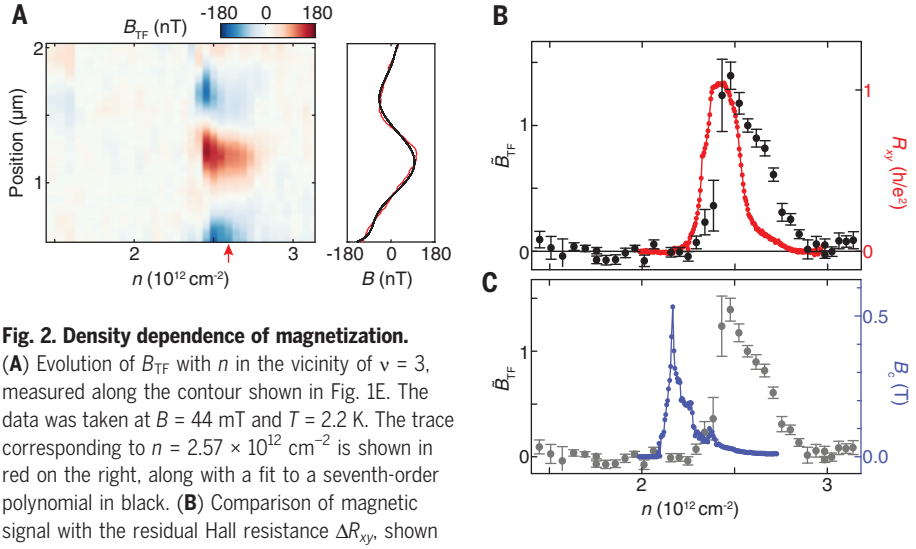


Fig. 2. Density dependence of magnetization. (**A**) Evolution of B_{TF} with n in the vicinity of $\nu = 3$, measured along the contour shown in Fig. 1E. The data was taken at $B = 44$ mT and $T = 2.2$ K. The trace corresponding to $n = 2.57 \times 10^{12}$ cm^{-2} is shown in red on the right, along with a fit to a seventh-order polynomial in black. (**B**) Comparison of magnetic signal with the residual Hall resistance ΔR_{xy}, shown in red. B_{TF} traces at different n are fit to the same polynomial as in (A), with a scale factor $\tilde{B}_{TF}$, which serves as a proxy for m. Error bars measure standard error of the mean of the residuals of these fits, normalized to the values shown in the inset of (A). This analysis is covered in more detail in fig. S11 (*24*). (**C**) Coercive field B_c determined from transport measurements (*24*) (fig. S8) plotted alongside $\tilde{B}_{TF}$. These data also appear in the supplementary materials of (*6*) and are reproduced with permission.

reflect standard error of the mean of the residuals of these fits. B_{TF} is below our noise floor for $n \lesssim 2.25 \times 10^{12}$ cm^{-2}. For 2.25×10^{12} cm$^{-2} < n < 2.52 \times 10^{12}$ cm^{-2}—a density range that overlaps with the quantization of the Hall conductivity (*4*)—B_{TF} increases rapidly. For $n > 2.52 \times 10^{12}$ cm^{-2}, B_{TF} then slowly decreases as a function of n, dropping below the noise floor at $n \approx 2.9 \times 10^{12}$ cm^{-2}, corresponding to a superlattice filling of $\nu \approx 3.7$ and approximately coinciding with the vanishing of ferromagnetic signatures in transport. Comparing the n-dependent $\tilde{B}_{TF}$ with the residual Hall resistance shows that $\tilde{B}_{TF}$ grows slowly as density is lowered toward $\nu = 3$ but then abruptly drops below the noise floor of our SQUID measurements within the range of n associated with the quantized R_{xy} plateau.

The drastic change of the inferred m within the quantum anomalous Hall plateau arises from the contribution of the chiral edge state to the total magnetization. Within an energy gap, changes in the chemical potential induce no additional charges in the sample bulk and, consequently, no change in the bulk magnetization. In a Chern insulator, however, charges may accumulate on the sample boundary owing to the presence of chiral edge states. This is predicted to give rise to a chemical potential–dependent contribution to the magnetization within the quantized transport plateau (*22*). We interpret the sharp change we observe in magnetization within the transport plateau as evidence for this contribution to the magnetization by the topological edge states. Assuming B_{TF} to be a good proxy for m, our measurements imply a $\Delta m \gtrsim 3\ \mu_B$ per unit cell across the quantum anomalous Hall gap. The magnitude of the theoretically expected jump in magnetization across a Chern insulator gap is $\Delta m = CE_{gap}/\Phi_0$, where C is the Chern number, E_{gap} is the topological band gap, and $\Phi_0 = h/e$ is the flux quantum. Thermally activated transport measurements of the quantum anomalous Hall state at $\nu = 3$ in this sample found $E_{gap} = 2.5$ meV (*4*), corresponding to a Δm of $1.4\ \mu_B$ per unit cell, somewhat smaller than our local measurement and potentially implying that the transport measurements underestimate E_{gap}.

The orbital magnetization contributed by Bloch states within a valley-projected subband is momentum dependent. As a result, the total magnetization is expected to depend on electron density, although the precise trend is sensitive to the details of the many-particle ground state wave function. For 2.52×10^{12} cm$^{-2} < n < 2.9 \times 10^{12}$ cm^{-2}, the observed gradual decrease in B_{TF} is concomitant with the hysteresis in transport. In contrast, no B_{TF} signal is observed for $n < 2.25 \times 10^{12}$ cm^{-2} despite hysteresis in transport persisting until $n = 2.05 \times 10^{12}$ cm^{-2}. Under the assumption that $B_{TF} \propto m$, absence of measured signal implies $|m| < 0.2\ \mu_B$ per moiré unit cell. Transport measurements in this density range show that the anomalous Hall effect changes sign at $n \approx 2.17 \times 10^{12}$ cm^{-2}, accompanied by a divergence in the coercive field (fig. S8 and Fig. 2C). These phenomena, observed in a regime of undetectably small m, point to a density-tuned transition in the valley occupation mediated by a sign change in the valley subband magnetization. In this picture, the sign change of the anomalous Hall effect arises directly from the opposite Berry curvatures of the contrasting valleys.

Although the coercive field behavior is difficult to model quantitatively, coercive fields in general result from competition between the energetic barrier to magnetic inversion and the coupling of magnetic order to the magnetic field: $B_c \sim E/m$. For densities where m approaches zero and E does not, B_c will to a leading order be sensitive to changes in $1/m$. The observation of a divergence in B_c concomitant with a sign change in the anomalous Hall effect is consistent with a vanishing m at this point. Our observation of an undetectably small magnetization in this regime strongly suggests these anomalies arise from a sign change in m at partial band filling rather than from density-dependent extrinsic contributions to the anomalous Hall effect or domain pinning dynamics.

Across the range of densities at which magnetic hysteresis is observed, B_c is not simply related to m, raising the question of the nature of magnetic pinning. Previous work on graphene-based Chern insulators has found Barkhausen noise jumps comparable to h/e^2 (*3*, *4*, *6*), suggesting a substructure of only a handful of ferromagnetic domains comparable in size to the distance between contacts. However, our magnetometry data show considerable submicrometer-scale inhomogeneity even at full magnetic saturation. This is similar to findings in transition metal–doped topological insulators, where the magnetic structure is dominated by inhomogeneous distribution and clustering of the Cr or V dopants. In those systems, magnetic imaging shows superparamagnetic dynamics characterized by the reversal of weakly correlated point-like microscopic magnetic dipoles (*26*–*28*). Transport, meanwhile, does not typically show substantial Barkhausen noise (*29*), with the exception of one study where jumps were reported in a narrow range of temperatures (*30*).

To investigate the domain dynamics directly, we compare magnetic structure across different states stabilized in the midst of magnetic field–driven reversal. Figure 3A shows a schematic depiction of our transport measurement, and Fig. 3B shows the resulting R_{xy} data for both a major hysteresis loop spanning the two fully polarized states at $R_{xy} = \pm h/e^2$ (in purple) and a minor loop that terminates in a mixed polarization state at $R_{xy} \approx 0$ (in red). All three states represented by these hysteresis loops can be stabilized at $B = 22$ mT for $T = 2.1$ K, where our nanoSQUID has excellent sensitivity, allowing a direct comparison of their respective magnetic structures (Fig. 3, C to E). Figure 3, F and G, shows images obtained by

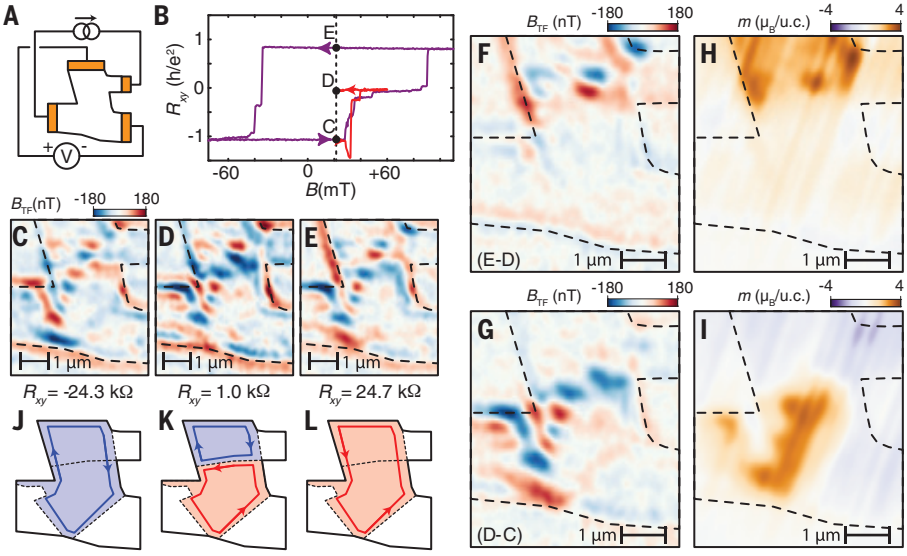


Fig. 3. Mesoscopic magnetic domains. (**A**) Transport measurement schematic and (**B**) Hall resistance data as a function of magnetic field for major (purple) and minor (red) hysteresis loops. (**C** to **E**) Gradient magnetometry images of the tBLG device at the three indicated values of the Hall resistance. All magnetic imaging presented here was performed at B = 22 mT, T = 2.1 K, and $n = 2.36 \times 10^{12}$ cm^{-2}. (**F** and **G**) Pairwise difference images based on the data presented in (C) to (E). The subtraction is indicated in the lower-left corner of each panel. The same domain structures were observed on multiple cooldowns, reminiscent of grains formed by crystalline domain walls in polycrystalline ferromagnetic metals. (**H** and **I**) Extracted magnetization m for the images in (F) and (G). (**J** to **L**) Schematic depiction of edge-state structure corresponding to magnetization states in (C) to (E). Assuming full edge-state equilibration, these states would result in $R_{xy} = -h/e^2$, 0, and h/e^2, close to the observed values.

subtracting one of the images at full positive or negative polarization from the mixed state, as indicated in the lower-left corners of the panels. Applying the same magnetic inversion algorithm used in Fig. 1 produces maps of m corresponding to these differences (Fig. 3, H and I), allowing us to visualize the domain structure generating the intermediate plateau $R_{xy} \approx 0$ seen in the major hysteresis loop. The domains presented in Fig. 3, H and I, are difference images; the domain structures actually realized in experiment (Fig. 3, C to E) are illustrated schematically in Fig. 3, J to L. Evidently, the Hall resistance of the device in this state is dominated by the interplay of two large magnetic domains, each comprising about half of the active area.

Armed with knowledge of the domain structure, it is straightforward to understand the behavior of the measured transport in the mixed state imaged in Fig. 3D. In particular, the state corresponds to the presence of a single domain wall that crosses the device, separating both the current and the Hall voltage contacts (see Fig. 3A). In the limit in which the topological edge states at the boundaries of each magnetic domain are in equilibrium, there will be no drop in chemical potential across the domain wall, leading to $R_{xy} = 0$. This is very close to the observed value of R_{xy} = 1.0 kilohms = 0.039 h/e^2. As shown in figs. S9 and S10, more-subtle features of the transport curve can also be associated with the reversal of domains that do not bridge contacts (*24*).

In the absence of substantial magnetic disorder, ferromagnetic domain walls minimize surface tension. In two dimensions, domain walls are pinned geometrically in devices of finite size with convex internal geometry. As discussed in fig. S10 (*24*), we observe pinning of domain walls at positions that do not correspond to minimal-length internal chords of our device geometry, suggesting that magnetic order couples to structural disorder directly. This is corroborated by the fact that the observed domain reversals associated with the Barkhausen jumps are consistent over repeated thermal cycles between cryogenic and room temperatures.

Although crystalline defects on the atomic scale are unlikely in tBLG thanks to the high quality of the constituent graphene and hBN layers, the thermodynamic instability of magic-angle twisted bilayer graphene makes it highly susceptible to inhomogeneity at scales larger than the moiré period, as shown in prior spatially resolved studies (*31*, *32*). For example, the twist angle between the layers as well as their position relative to the underlying hBN substrate may all vary spatially, providing potential pinning sites (*33*). Moiré disorder may thus be analogous to crystalline disorder in conventional ferromagnets, which gives rise to Barkhausen noise as it was originally described (*34*). A subtler issue raised by our data is the density dependence of magnetic pinning; as shown in Fig. 2, B_c does not simply track $1/m$ across the entire density range, in particular, failing to collapse with the rise in m in the topological gap. This suggests nontrivial dependence of the pinning potential on the realized many-body state. Understanding the pinning dynamics is critical for stabilizing magnetism in tBLG and the growing class of related orbital magnets, which includes both moiré systems (*3*–*6*) as well as more-traditional crystalline systems such as rhombohedral graphite (*35*).

REFERENCES AND NOTES

1. D. Xiao, M.-C. Chang, Q. Niu, *Rev. Mod. Phys.* **82**, 1959–2007 (2010).
2. H. Adachi, H. Ino, *Nature* **401**, 148–150 (1999).
3. A. L. Sharpe *et al.*, *Science* **365**, 605–608 (2019).
4. M. Serlin *et al.*, *Science* **367**, 900–903 (2020).
5. G. Chen *et al.*, *Nature* **579**, 56–61 (2020).
6. H. Polshyn *et al.*, *Nature* **588**, 66–70 (2020).
7. S. Chen *et al.*, *Nat. Phys.* **17**, 374–380 (2020).
8. J. Sichau *et al.*, *Phys. Rev. Lett.* **122**, 046403 (2019).
9. M. Xie, A. H. MacDonald, *Phys. Rev. Lett.* **124**, 097601 (2020).
10. N. Bultinck, S. Chatterjee, M. P. Zaletel, *Phys. Rev. Lett.* **124**, 166601 (2020).
11. Y.-H. Zhang, D. Mao, T. Senthil, *Phys. Rev. Res.* **1**, 033126 (2019).
12. J. Liu, X. Dai, *Phys. Rev. B* **103**, 035427 (2021).
13. F. Wu, S. Das Sarma, *Phys. Rev. Lett.* **124**, 046403 (2020).
14. S. Chatterjee, N. Bultinck, M. P. Zaletel, *Phys. Rev. B* **101**, 165141 (2020).
15. C. Repellin, Z. Dong, Y.-H. Zhang, T. Senthil, *Phys. Rev. Lett.* **124**, 187601 (2020).
16. Y. Alavirad, J. Sau, *Phys. Rev. B* **102**, 235123 (2020).
17. K. Kim *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **114**, 3364–3369 (2017).
18. Y. Cao *et al.*, *Nature* **556**, 80–84 (2018).
19. G. Chen *et al.*, *Nat. Phys.* **15**, 237–241 (2019).
20. J. C. W. Song, P. Samutpraphoot, L. S. Levitov, *Proc. Natl. Acad. Sci. U.S.A.* **112**, 10879–10883 (2015).
21. Y.-H. Zhang, D. Mao, Y. Cao, P. Jarillo-Herrero, T. Senthil, *Phys. Rev. B* **99**, 075127 (2019).
22. J. Zhu, J.-J. Su, A. H. MacDonald, *Phys. Rev. Lett.* **125**, 227702 (2020).
23. Y. Anahory *et al.*, *Nanoscale* **12**, 3174–3182 (2020).
24. See supplementary materials.
25. A. Uri *et al.*, *Nat. Phys.* **16**, 164–170 (2020).
26. E. O. Lachman *et al.*, *Sci. Adv.* **1**, e1500740 (2015).
27. E. O. Lachman *et al.*, *NPJ Quantum Mater.* **2**, 70 (2017).
28. K. Yasuda *et al.*, *Science* **358**, 1311–1314 (2017).
29. C.-Z. Chang *et al.*, *Science* **340**, 167–170 (2013).
30. M. Liu *et al.*, *Sci. Adv.* **2**, e1600167 (2016).
31. A. Uri *et al.*, *Nature* **581**, 47–52 (2020).
32. T. Benschop *et al.*, *Phys. Rev. Res.* **3**, 013153 (2021).
33. Y. H. Kwan, G. Wagner, N. Chakraborty, S. H. Simon, S. A. Parameswaran, arXiv:2007.07903 [cond-mat.str-el] (23 November 2020).
34. H. Barkhausen, *Phys. Ztschr.* **20**, 401 (1919).
35. Y. Shi *et al.*, *Nature* **584**, 210–214 (2020).
36. C. L. Tschirhart *et al.*, Imaging orbital ferromagnetism in a moire Chern insulator, Dryad, Dataset (2021); https://doi.org/10.25349/D9JC8B.

ACKNOWLEDGMENTS

We thank A. H. Macdonald, J. Zhu, M. Zaletel, and D. Xiao for discussions of the results and E. Lachman for comments on the manuscript. **Funding:** The work was primarily funded by the US Department of Energy under DE-SC0020043, with additional support for instrumentation development supported by the Army

Research Office under grant W911NF-16-1-0361. K.W. and T.T. acknowledge support from the Elemental Strategy Initiative conducted by MEXT, Japan, grant JPMXP0112101001; JSPS KAKENHI grant JP20H00354 and CREST grant JPMJCR15F3, JST. C.L.T. acknowledges support from the Hertz Foundation and from the National Science Foundation Graduate Research Fellowship Program under grant 1650114. This project is funded in part by the Gordon and Betty Moore Foundation's EPiQS Initiative, grant GBMF9471 to A.F.Y. **Author contributions:** H.P. and Y.Z. fabricated the graphene heterostructure. C.L.T., M.S., and Z.X. fabricated the scanning probe tips. C.L.T., M.S., A.S., Z.X., J.Z., and A.F.Y. constructed the scanning probe microscope. M.S., A.S., and J.Z. wrote the microscope control software. M.E.H. provided the cryogenic amplifier. C.L.T., M.S., and A.F.Y. performed the measurements. C.L.T., M.S., Z.X., and A.F.Y. analyzed the data. T.T. and K.W. provided the hBN crystals. C.L.T., M.S., H.P., Z.X., and A.F.Y. wrote the manuscript. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** All data shown in the main text and supplementary materials are available in the Dryad data repository (*36*).

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/372/6548/1323/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S11
References (*37*–*45*)

14 June 2020; accepted 13 May 2021
Published online 27 May 2021
10.1126/science.abd3190

SOLAR CELLS

Lead halide–templated crystallization of methylamine-free perovskite for efficient photovoltaic modules

Tongle Bu[1,2,3], Jing Li[1], Hengyi Li[1], Congcong Tian[1], Jie Su[4], Guoqing Tong[3], Luis K. Ono[3], Chao Wang[1], Zhipeng Lin[1], Nianyao Chai[1], Xiao-Li Zhang[5], Jingjing Chang[4], Jianfeng Lu[2,6], Jie Zhong[1,2], Wenchao Huang[1], Yabing Qi[3], Yi-Bing Cheng[2], Fuzhi Huang[1,2]*

Upscaling efficient and stable perovskite layers is one of the most challenging issues in the commercialization of perovskite solar cells. Here, a lead halide–templated crystallization strategy is developed for printing formamidinium (FA)–cesium (Cs) lead triiodide perovskite films. High-quality large-area films are achieved through controlled nucleation and growth of a lead halide•*N*-methyl-2-pyrrolidone adduct that can react in situ with embedded FAI/CsI to directly form α-phase perovskite, sidestepping the phase transformation from δ-phase. A nonencapsulated device with 23% efficiency and excellent long-term thermal stability (at 85°C) in ambient air (~80% efficiency retention after 500 hours) is achieved with further addition of potassium hexafluorophosphate. The slot die–printed minimodules achieve champion efficiencies of 20.42% (certified efficiency 19.3%) and 19.54% with an active area of 17.1 and 65.0 square centimeters, respectively.

Hybrid organic-inorganic metal halide perovskite solar cells (PSCs) have attracted intensive interest during the past decade, with power conversion efficiencies (PCEs) now greater than 25% (*1*). Such a development is attributed to the intrinsically superior photoelectric properties of the perovskite materials that possess tunable bandgaps, high absorption coefficients, and long carrier diffusion lengths (*2*–*4*). In particular, the PSCs can be fabricated through a myriad of low-cost solution processes, which offers great promise for future commercialization. However, scalability and stability issues have impeded their industrialization.

[1]State Key Laboratory of Advanced Technology for Materials Synthesis and Processing, Wuhan University of Technology, Wuhan 430070, PR China. [2]Foshan Xianhu Laboratory of the Advanced Energy Science and Technology Guangdong Laboratory, Foshan 528216, PR China. [3]Energy Materials and Surface Sciences Unit (EMSSU), Okinawa Institute of Science and Technology Graduate University (OIST), Okinawa 904-0495, Japan. [4]Xidian University, School of Microelectronics, State Key Discipline Lab of Wide Band Gap Semiconductor Technology, Shaanxi Joint Key Lab of Graphene, Advanced Interdisciplinary Research Center for Flexible Electronics, Xian 710071, PR China. [5]School of Materials Science and Engineering, Zhengzhou University, Zhengzhou 450001, PR China. [6]State Key Laboratory of Silicate Materials for Architectures, Wuhan University of Technology, Wuhan 430070, PR China.
***Corresponding author. Email: fuzhi.huang@whut.edu.cn**

The most important prerequisite for fabricating large-area PSCs is the deposition of high-quality perovskite thin films. The nucleation and crystal growth of the perovskite in solution are largely uncontrollable, often leading to a porous film that would greatly impair the device's performance (*5*, *6*). The larger the area is, the harder it will be to achieve a uniform crystalline film. Various efforts have been devoted to controlling the nucleation and crystal growth for scaling up perovskite films. Many strategies—including antisolvent bathing (*7*), softcover coating (*8*), gas flow (*9*), vacuum (*10*) or thermal assisting (*11*), and additive engineering (*12*)—have been successfully used to fabricate high-quality large-area perovskite films. For example, Hu and others used an air blade to quickly remove the solvent of the methylammonium lead triiodide ($MAPbI_3$) perovskite wet film to promote the concentration of the perovskite precursor and induce a higher nucleation rate, forming a dense perovskite film (*13*). Huang and others reported a thermally assisted blade-coated perovskite film with an efficiency of 14.6% on 57.2-cm^2 perovskite solar modules (PSMs) using a surfactant-added $MAPbI_{3-x}Cl_x$ perovskite ink (*11*).

However, $MAPbI_3$ perovskites have shown poor stability at high temperatures or under light illumination (*14*, *15*). Instead, the MA-free perovskites such as formamidinium lead triiodide ($FAPbI_3$) or formamidinium-caesium lead triiodide [$(FACs)PbI_3$] show promising thermal stability owing to their higher phase-transformation temperatures (*16*–*18*). In addition, the narrower optical bandgap of $FAPbI_3$ with respect to that of $MAPbI_3$ can contribute to higher efficiencies (*19*). Recently, the FA-based PSCs without MA have attracted intensive attention, especially for large-area devices (*20*). Unfortunately, the nucleation and crystal growth of the FA-based perovskites are even harder to control. Alternatively, it might be feasible to control the nucleation of perovskite intermediates such as their solvent-coordinated complexes (*9*, *21*). However, the nucleation rate of solvent-coordinated FA-based perovskite complexes is still not high enough. Indeed, besides the one-step method, the two-step method is also widely used in fabricating small-area perovskite films. It is rather easy to achieve a dense PbI_2 film in the first step, but it is hard for the FAI deposited in the second step to diffuse into the bottom of the PbI_2 film to induce a complete reaction (*22*). In the perovskite precursor solution, if the FAI and PbI_2 species do not form perovskites or solvent-coordinated perovskite complexes, the nucleation will be dominated by PbI_2 and therefore it will be easier to form a dense film. Because FAI is embedded during the formation of the PbI_2 film, it is easier to induce an in situ reaction between the PbI_2 and FAI by the subsequent thermal annealing, resulting in a dense perovskite film. Thus, the crystallization of the perovskite is templated by the PbI_2-derived crystals.

Here, we report a lead halide–templated crystallization strategy to prepare compact methylamine-free perovskite films for the fabrication of antisolvent–free and ambient air–printed high-performance PSMs. The key point to obtaining high-quality large-area FA-based perovskite films is to completely inhibit the formation of a solvent-coordinated perovskite intermediate complex via the formation of a stable PbI_2•*N*-methylpyrrolidone (NMP) adduct, which can react in situ with embedded FAI/CsI species. In addition, by using this process, we can lower the formation energy of α-phase perovskite, which is an unstable high-temperature phase, thus converting the α-phase FA-based perovskite film ($FA_{0.83}Cs_{0.17}PbI_3$) even

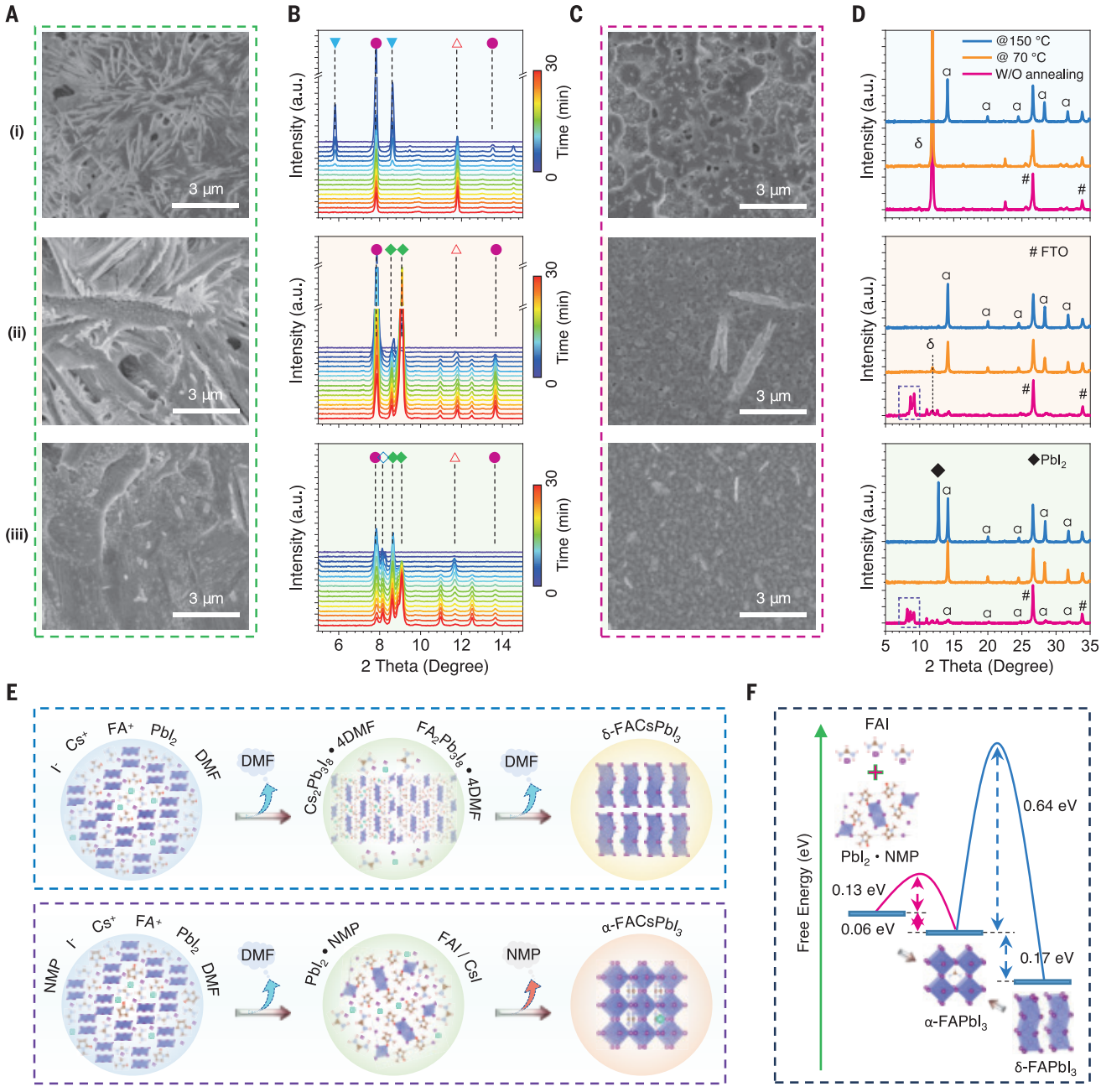


Fig. 1. Nucleation and crystallization of MA-free perovskites. (**A** and **B**) Shown are (A) SEM and (B) in situ XRD patterns of 10-μl perovskite precursor inks (1.1 M) with or without different additives drying on the 1.5-cm–by–1.5-cm FTO/glass substrates: (i) $FA_{0.83}Cs_{0.17}PbI_3$/DMF, (ii) $FA_{0.83}Cs_{0.17}PbI_3$-NMP/DMF, and (iii) $FA_{0.83}Cs_{0.17}PbI_3$-NMP-10%$PbCl_2$/DMF. $Cs_2Pb_3I_8$•4DMF, solid blue triangle; $FA_2Pb_3I_8$•4DMF, purple circle; open triangle, δ-(FACs)PbI_3; solid green diamond, PbI_2•NMP; open diamond, PbX_2•0.5NMP•0.5DMF; a.u., arbitrary units. (**C**) SEM images of the corresponding perovskite ink spun onto FTO/glass substrates at 3000 rpm. (**D**) The corresponding XRD patterns of spin-coated (at 3000 rpm) perovskite films annealed at different temperatures for 10 min. (**E**) Schematic diagram of crystal growth with or without NMP. (**F**) Free-energy calculation for the formation of $FAPbI_3$ perovskites with or without NMP.

at room temperature. The resulting perovskite films are further passivated by a KPF_6 salt, which contributes to high-performance hysteresis-free PSCs with an efficiency of 23.35% and dramatically enhanced thermal and light stability. Eventually, a slot die–printed high-quality large-area perovskite film is realized using this strategy. The corresponding solar minimodules demonstrate efficiencies of 20.42% on 17.1 cm^2 and 19.54% on 65.0 cm^2, respectively.

$FAPbI_3$ has the narrowest bandgap (~1.48 eV) among the Pb-based perovskites and much better thermal stability than MA-based perovskite. However, pure α-$FAPbI_3$ is unstable, so Cs is normally introduced to stabilize the phase (*23*). Here, $FA_{0.83}Cs_{0.17}PbI_3$ perovskite is used to study its nucleation and crystal growth kinetics. *N*,*N*-dimethylformamide (DMF) is a commonly used solvent for perovskite precursor ink because of its high solubility and volatility. We found that during the natural drying process, only a few nuclei form at the beginning (fig. S1i), then several flat, needle-shaped crystals

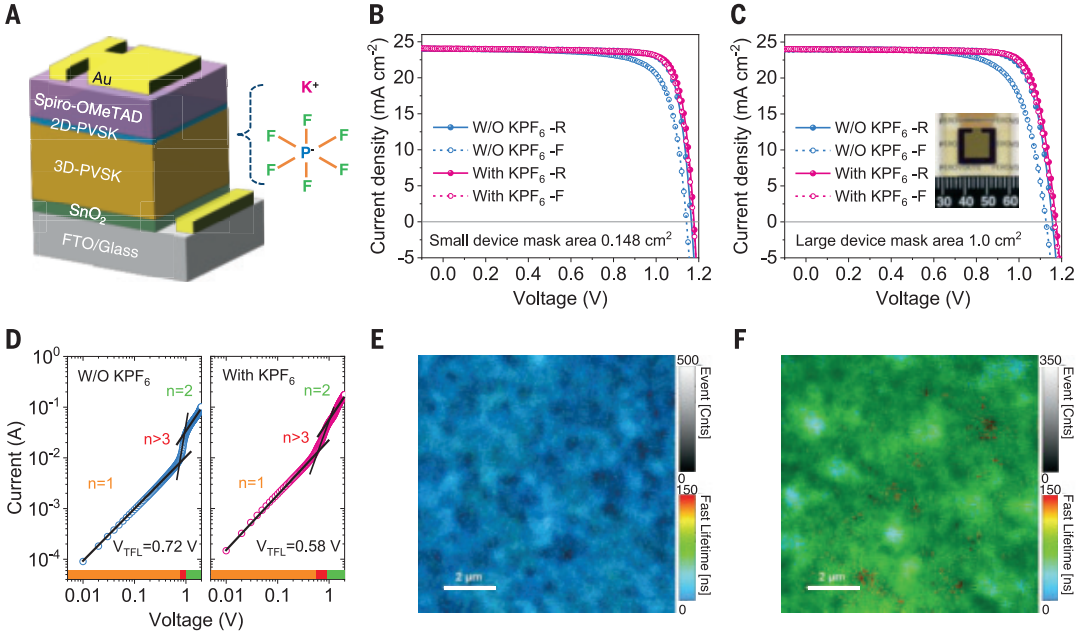


Fig. 2. Photovoltaic performance of antisolvent–free coated PSCs with KPF_6 passivation. (**A**) Schematic of the PSC with the structure FTO/SnO_2/3D-perovskite/2D-perovskite/spiro-OMeTAD/Au. PVSK, perovskite. (**B** and **C**) Champion *J-V* curves of 3D/2D perovskite–based devices with or without KPF_6 additive tested using a metal mask with an aperture area of (B) 0.148 cm^2 and (C) 1.0 cm^2. (**D**) *J-V* characteristics of the 3D/2D perovskite films with or without KPF_6 additive derived from the SCLC measurements with a structure of ITO/perovskite/Au (ITO, indium tin oxide). V_{TFL}, trap-filled limited voltage. (**E** and **F**) Time-resolved confocal PL lifetime maps of (E) 3D/2D and (F) 3D/2D-KPF_6 perovskite films, respectively.

grow, surrounding every nucleus. After drying completely, a scanning electron microscopy (SEM) image (Fig. 1Ai) shows a rough film with dendrites, large pores, and some densely packed large grains lying underneath the dendrites, clearly indicating that there are two types of structures. To trace the formation processes of such structures, in situ x-ray diffraction (XRD) was conducted to investigate phase change during the natural drying process. According to the density functional theory (DFT) fitting (fig. S2) and a comparison with the in situ XRD patterns from different components with additives (fig. S3), it can be inferred from Fig. 1Bi and fig. S4 that at the beginning, solvent-coordinated perovskite intermediate phases of $Cs_2Pb_3I_8$•4DMF and $FA_2Pb_3I_8$•4DMF form and then transform to δ-(FACs)PbI_3. Finally, $FA_2Pb_3I_8$•4DMF and δ-(FACs)PbI_3 are mainly present in the film. Although perovskites in the DMF form intermediate solvent-coordinated complexes, the nucleation rate is still too low during the natural drying process. To accelerate the nucleation rate, the precursor solution is spun at 3000 rpm to quickly remove the solvent. The morphology change is shown in Fig. 1Ci. There are some large-area dense zones that formed, with a few rods randomly distributed. So even with the assistance of solvent removal by spinning, the nucleation rate is not high enough, because there are still many large pores. This demonstrates that it is hard to change the nucleation kinetics of FA-based perovskites through the approach of forming intermediate solvent-coordinated complexes. The XRD patterns (Fig. 1Di) show the presence of major δ-(FACs)PbI_3 in the final film.

In our initial study, we found that NMP can form a strong PbI_2•NMP adduct (fig. S3). As proposed, if PbI_2•NMP can remain in the perovskite precursor solution, then the nucleation will be dominated by the PbI_2•NMP adduct. After the formation of the film, the FAI/CsI species are also homogeneously distributed in the film. It is very easy to induce the reaction of PbI_2 and FAI/CsI in the film by thermal annealing. Thus, the film morphology will be well controlled. Then NMP with a molar ratio of 1:1 to PbI_2 is added into the $FA_{0.83}Cs_{0.17}PbI_3$/DMF solution. Similar studies are carried out, and we find that there are two types of nuclei (fig. S1ii). The nuclei that induce the growth of needle-shaped crystals have no obvious changes, but the length of the needle becomes a little shorter, indicating that the growth of the solvent-coordinated perovskite complexes is suppressed. More importantly, a large number of the second type of nuclei form, growing into spherical particles. Through the in situ XRD (Fig. 1Bii), we find that the initial strong peaks of $FA_2Pb_3I_8$•4DMF become weaker and the later-appearing PbI_2•NMP peaks become stronger, leaving the final film consisting of $FA_2Pb_3I_8$•4DMF and PbI_2•NMP. The SEM image (Fig. 1Aii) also shows that there are two types of structures in the rough film: densely packed particles in the layered structure and some dendrite structures that originated from the $FA_2Pb_3I_8$•4DMF. When the film is prepared by spin-coating the precursor solution at 3000 rpm, the film becomes smoother and denser and has a light brown color (fig. S5), but it still has some small pores and several needles (Fig. 1Cii). XRD patterns (Fig. 1Dii) show the presence of an almost pure PbI_2•NMP phase. Interestingly, we also find that minor α-(FACs)PbI_3 emerges in the film. When we increase the spinning speed to 5000 rpm, the peak intensity of the α-phase becomes stronger (fig. S6). It is completely different from that of the film derived from the pure DMF solution. When the film is annealed at 70°C, an obvious α-(FACs)PbI_3 phase appears (Fig. 1Dii), and the film becomes a black color (fig. S5). With a further increase in annealing temperature to 150°C, the peak of the α-phase becomes even stronger. For the DMF-derived film, when annealed at 70°C, the peak of the δ-(FACs)PbI_3 phase becomes stronger (Fig. 1Di) and the film still remains a yellow color (fig. S5). However, when further annealed at 150°C, the δ-phase is completely transformed into α-phase (Fig. 1Di) and becomes a black color (fig. S5). According to the DFT calculation (Fig. 1F), because α-$FAPbI_3$ is a high-temperature phase, the conversion energy is high if it is from δ-$FAPbI_3$, which is rapidly transformed from the DMF-coordinated complexes. However, when PbI_2•NMP reacts with FAI to form perovskite, the α-$FAPbI_3$ formation energy is dramatically decreased. This is why the addition of NMP can induce an α-phase perovskite at room temperature with the incorporation of Cs. The study of pure $FAPbI_3$ films growing from 2-methoxyethanol (2-Me), an uncoordinated solvent, with or

without corresponding coordination solvent additives further confirms the above findings, as shown in fig. S7. In short, the presence of the intermediate phases (the perovskite-DMF complexes) will result in porous δ-phase perovskite films, whereas the existence of PbI_2-NMP will directly produce dense α-phase perovskite films (fig. S4E). Therefore, an improved quality of perovskite film can be obtained by inhibiting the formation of the perovskite-DMF complexes.

Thus, it is more preferable if α-(FACs)PbI_3 is directly formed from the solution without a second phase transformation, which would suppress the formation of defects and traps during the δ- to α-phase transition. From the above results, we can conclude that in the (FACs)PbI_3 perovskite DMF solution, solvent-coordinated perovskite intermediate phases of $Cs_2Pb_3I_8$•4DMF and $FA_2Pb_3I_8$•4DMF that later convert to δ-(FACs)PbI_3 are likely to form and thus result in poor morphology, as depicted in Fig. 1E. When NMP is added, the intermediate phases of $Cs_2Pb_3I_8$•4DMF and $FA_2Pb_3I_8$•4DMF are restrained by the competition of PbI_2•NMP, resulting in the formation of α-(FACs)PbI_3. To further improve the film quality, the intermediate phase of $FA_2Pb_3I_8$•4DMF should be completely inhibited. By the further introduction of excess $PbCl_2$ to the precursor solution, we find that the growth of the $FA_2Pb_3I_8$•4DMF nuclei is further suppressed by the formation of an additional PbX_2•0.5NMP•0.5DMF adduct, as shown in fig. S1iii, and the film becomes much denser (Fig. 1Aiii), more transparent, and brown in color (fig. S5), indicating much faster nucleation of lead halide–NMP adducts and effective suppression of δ-phase formation. Further evidence can be found from the in situ XRD patterns (Fig. 1Biii), which show the presence of a much lower peak of $FA_2Pb_3I_8$•4DMF, and the SEM image of the spin-coated film, which shows negligible pores (Fig. 1Ciii). The related XRD patterns (Fig. 1Diii) of the spin-coated film show the presence of PbI_2•NMP and PbX_2•NMP/DMF complexes without $FA_2Pb_3I_8$•4DMF. The XRD patterns (fig. S6) and SEM images (fig. S8) indicate that a relatively low volatilization rate of the precursor solution during the deposition (spin rate >3000 rpm) is sufficient to achieve dense perovskite films even without an antisolvent process. The annealing even at 70°C could induce a stronger peak of α-(FACs)PbI_3 phase. The additional peak belonging to PbI_2 at 150°C is due to the added excess PbX_2. The excess PbI_2 is generally beneficial to the device because of the passivation effect (*24*).

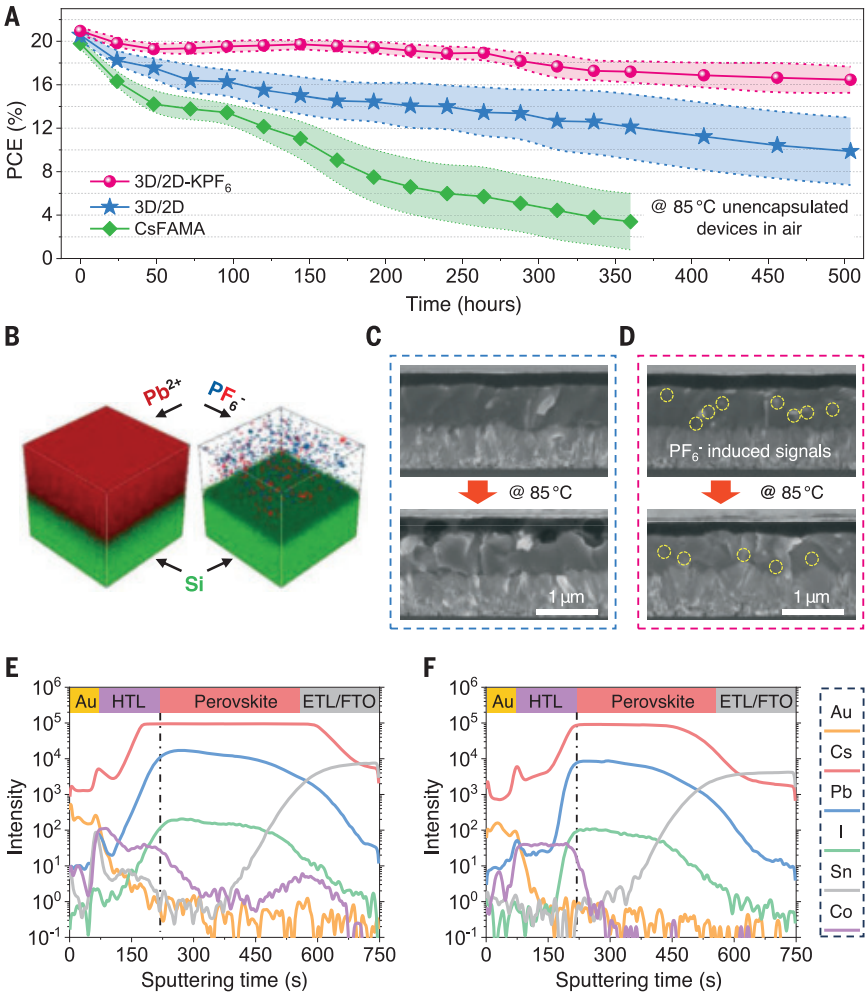


Fig. 3. Thermal stability characterization. (**A**) Average PCE evolution of the unencapsulated devices measured over a 500-hour stability test at 85°C in ambient air (relative humidity ~15 ± 5%). The shaded regions represent the variation range of the PCE obtained from eight cells. (**B**) TOF-SIMS of Pb, P, and F ions in the $FA_{0.83}Cs_{0.17}PbI_{3-x}Cl_x$ perovskite film with KPF_6 additive on a Si substrate (measurement area 80 μm by 80 μm). (**C** and **D**) Cross-sectional SEM images of different devices: (C) fresh 3D/2D device before and after 360 hours of 85°C aging and (D) fresh 3D/2D-KPF_6 device before and after 360 hours of 85°C aging. (**E** and **F**) TOF-SIMS spectra of (E) 3D/2D and (F) 3D/2D-KPF_6 devices after 360 hours of 85°C aging. HTL, hole transport layer; ETL/FTO, electron transport layer/fluorine-doped tinoxide.

To study the corresponding film's device performance, we used a normal structure of FTO/SnO_2/perovskite/spiro-OMeTAD/Au [FTO, fluorine-doped tin dioxide; spiro-OMeTAD, 2,2′,7,7′-tetrakis(*N*,*N*-di-*p*-methoxyphenyl-amine) 9,9′-spirobifluorene]. To facilitate industrial production, we adopted a two-step annealing process for perovskites, namely a 70°C annealing in the glovebox to dry the films followed by another 150°C annealing in air to promote crystal growth with the assistance of humidity. The champion current density–voltage (*J-V*) curves of PSCs are shown in fig. S9. The devices made from pure DMF-derived perovskite films exhibit a poor PCE of 7.64%, whereas the NMP-engineered perovskite devices show a much higher efficiency of greater than 20%. We further modulated the addition of $PbCl_2$ and obtained mirror-like black films (fig. S5). The efficiencies are substantially improved when the amount of $PbCl_2$ is increased to no more than 10% while the hysteresis continues to decrease. A champion efficiency of 21.92% is achieved by the addition of 10% $PbCl_2$ (fig. S9). After further characterization of the crystal properties by XRD, UV-visible (UV-Vis) absorption spectroscopy, and SEM for perovskite films with different amounts of added $PbCl_2$ (fig. S10), we found that the intrinsic reasons for the improved performance by introducing $PbCl_2$ are the suppression of δ-phase formation, the improvement of coverage with increased grain size, and the in situ formation of PbI_2 at the grain boundary as a passivator (fig. S10) due to the substitution of I^- or I^- vacancies by Cl^- to form a $FA_{0.83}Cs_{0.17}PbI_{3-x}Cl_x$ perovskite (*25*, *26*).

To further improve performance, we introduced a posttreatment of bromide-based large cation salt (isobutylamine bromide, iBABr) on

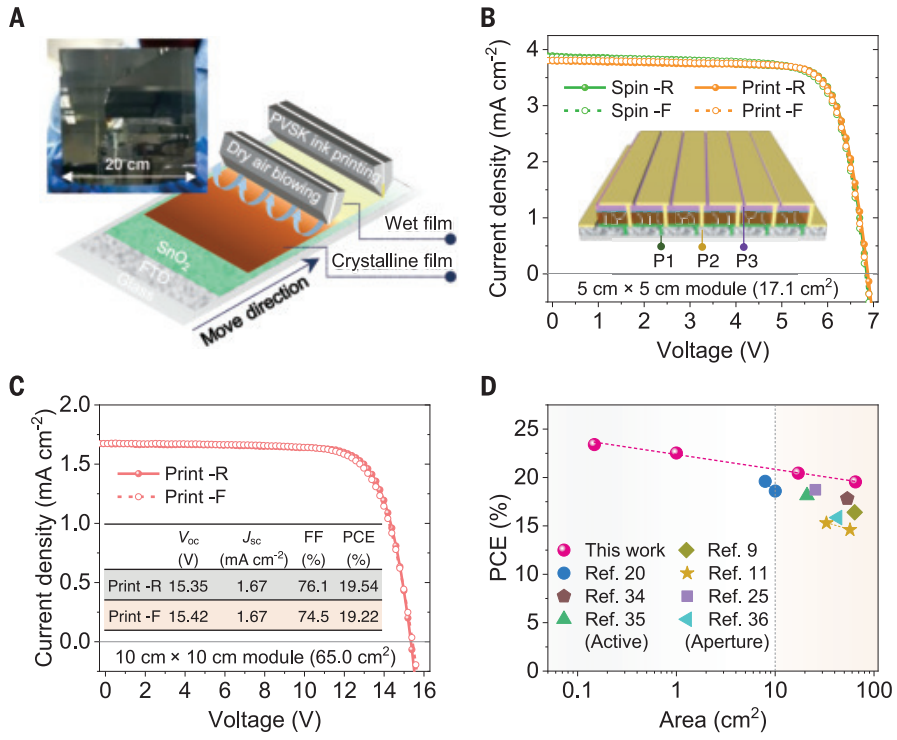


Fig. 4. Photovoltaic performance characterization of large-area modules. (**A**) Schematic illustration for the slot-die printing of perovskite films with low-pressure dry air blowing. The inset shows a photo of a printed 20-cm–by–20-cm perovskite film. (**B**) Champion *J-V* curves of the 5-cm–by–5-cm minimodules based on the antisolvent-free spin-coating method and slot-die printing method. The inset shows the schematic diagram of a six-subcell series-connected module. F, forward scan; R, reverse scan; P1, P2, and P3, three laser scribing steps. (**C**) Champion *J-V* curves of the 10-cm–by–10-cm minimodule using the large-scale perovskite film fabricated by the slot-die printing method. (**D**) Champion efficiencies of PSCs as a function of the area from this work and recent representative reports (*9*, *11*, *20*, *25*, *34*–*36*).

the as-fabricated $FA_{0.83}Cs_{0.17}PbI_{3-x}Cl_x$ [labeled as three-dimensional (3D) perovskite] surface to form a 3D/2D structured perovskite layer according to previous reports (*27*, *28*) (fig. S11), which would contribute to stability improvement and open-circuit voltage (V_{oc})–loss reduction for the solar cells. The corresponding device structure is shown in Fig. 2A. An improved efficiency of 23.02% under reverse scan (RS) for the 3D/2D perovskite-based solar cells is achieved, with a V_{oc} of 1.166 V, a short-circuit current density (J_{sc}) of 23.97 mA cm^{-2}, and a fill factor (FF) of 0.824. A lower efficiency of 20.71% is achieved under forward scan (FS), with a V_{oc} of 1.142 V, J_{sc} of 24.01 mA cm^{-2}, and a FF of 0.755, respectively (Fig. 2B and table S1). An obvious hysteresis with a hysteresis index (HI) of 0.10 is still present.

Hysteresis reflects the stability that is one of the major obstacles to the commercialization of PSCs (*29*). To eliminate the hysteresis and further improve the photovoltaic performance and stability, we used a potassium-based salt, KPF_6, as an additive to the perovskite precursor solution. XRD, UV-Vis absorption spectroscopy, and SEM characterizations (fig. S12) confirm the incorporation of KPF_6 to perovskites. The corresponding photovoltaic characteristics are shown in fig. S13. With the addition of KPF_6, the hysteresis continues to decrease and the HI reaches 0.00 at a 5% additive concentration. To accurately evaluate the real PCE, steady-state power output (SPO) measurements were conducted. With a lower HI, the PCE from SPO becomes closer to the PCE from *J-V* scans. When the concentration is above 0.5%, the hysteresis is actually almost negligible. A champion efficiency is obtained with the addition of 1.5 mol % KPF_6 salts (labeled as 3D/2D-KPF_6), showing a higher V_{oc} of 1.178 V, a J_{sc} of 24.03 mA cm^{-2}, a FF of 0.825, and a PCE of 23.35% under RS, and a V_{oc} of 1.175 V, a J_{sc} of 24.06 mA cm^{-2}, a FF of 0.818, and a PCE of 23.13% under FS. The average PCE of 23.24% is very close to the PCE from SPO of 23.2%, which is much higher than the control device (21.9%) (fig. S14A). It is not surprising to see that further addition of KPF_6 decreases the PCE, although there is no hysteresis, because too much KPF_6 will affect the nucleation behavior and lead to poorer film morphology (fig. S12C). The corresponding external quantum efficiency spectra show an integrated J_{sc} of 23.70 and 23.78 mA cm^{-2} for devices with and without KPF_6, respectively (fig. S14B), which match well with the measured J_{sc}. In addition, PSCs that are 1.0 cm^2 in size also exhibit substantial improvements in PCE, increasing from 21.92% (RS) and 19.31% (FS) to 22.53% (RS) and 22.27% (FS) with the addition of KPF_6 (Fig. 2C and table S2). The distribution of *J-V* parameters for both small-area and 1.0-cm^2-sized cells processed with or without KPF_6 additives are shown in fig. S15. Ultraviolet photoemission spectroscopy (UPS) and UV-Vis absorption spectroscopy (fig. S16) characterizations reveal the lifted conduction band minimum from −4.14 eV for 3D perovskite to −4.10 eV for 3D/2D perovskite, and further to −3.94 eV for 3D/2D-KPF_6 perovskite, which contributes to the enhanced V_{oc} of the modified devices.

We further used the space charge–limited current (SCLC) measurements to characterize the trap density of these different films (*30*–*32*). The calculated defect density decreases from 6.17×10^{15} to 4.97×10^{15} cm^{-3}, indicating the reduced defects of the perovskite film with the addition of KPF_6 (Fig. 2D). Time-resolved confocal photoluminescence (PL) microscopy measurements were also performed to characterize the charge carrier properties of these perovskite films in microsize (10 μm by 10 μm) (Fig. 2, E and F). The blue region with a short PL lifetime for the pristine 3D/2D perovskite transits to a green region with a much longer PL lifetime after the addition of KPF_6. It indicates a reduced trap-induced or nonradiative recombination by the defect passivation due to the addition of KPF_6, thus resulting in a small V_{oc} loss of ~0.37 eV (fig. S17). This passivation strategy would be beneficial to stability.

Figure 3A shows the thermal stability of widely researched CsFAMA triple-cation-based perovskite $Cs_{0.05}(FA_{0.85}MA_{0.15})_{0.95}Pb(I_{0.85}Br_{0.15})_3$ and our MA-free perovskite ($FA_{0.83}Cs_{0.17}PbI_3$)–based solar cells measured at 85°C in ambient air (15 ± 5% relative humidity). The efficiency of CsFAMA devices quickly decreases to 2.9 ± 3.1% within 360 hours. Obvious holes are observed from the cross-sectional SEM image because of the degradation of the CsFAMA device after aging at 85°C (fig. S18). Although the MA-free 3D/2D devices exhibit a considerable improvement and retain the black appearance, the devices still show a salient decrease after aging at 85°C for 500 hours, retaining less than 50% of the initial efficiency. However, this degradation can be substantially suppressed with the addition of KPF_6. The 3D/2D-KPF_6 devices exhibit ~80% of the initial efficiency after aging at 85°C for 500 hours. The *J-V* curves of different devices under the thermal stability measurements are plotted in fig. S19. There is a negligible change of FF in the device processed with KPF_6, indicating the undamaged interfaces of devices.

This substantial improvement in stability can be ascribed to the PF_6^--induced complex in perovskite films. The complex stays on the surface, passivating the grain boundaries, as seen from the SEM images (fig. S12C). In addition, to unravel the distribution of the PF_6^- complex across the perovskite layer, we conducted time-of-flight secondary-ion mass spectrometry (TOF-SIMS) for 1.5 mol % KPF_6–contained $FA_{0.83}Cs_{0.17}PbI_{3-x}Cl_x$ perovskite film (Fig. 3B). A block-by-block distribution of the agglomerated PF_6^- ions is shown in the 3D visualization images. The corresponding depth profiles of TOF-SIMS and x-ray photoelectron spectroscopy also show detected signals of K and F elements throughout the perovskite films (fig. S20, A to C). In addition, Fourier transform infrared spectra show that the N–H and P–F corresponding peaks shift to lower wave numbers with the increasing amount of KPF_6 additives, revealing the presence of the hydrogen bonding between perovskite and PF_6^- (fig. S20, D to F). Thus, the PF_6^- additives would greatly affect both the perovskite grain boundary and the interfacial properties of the devices, contributing to better performance by passivating the surface defects.

The corresponding cross-sectional SEM images of the thermally aged devices with or without KPF_6 additives are shown in Fig. 3, C and D. We found that there is no obvious change of the perovskite/spiro-OMeTAD interface with the KPF_6 additive, whereas the interface becomes intersected without the KPF_6 additive. We further used TOF-SIMS characterization to probe the thermally aged devices and fresh devices (Fig. 3, E and F, and fig. S21). In the device without KPF_6, Cs, Pb, and I ions are shifted toward the spiro-OMeTAD layer. The Co ions that are a dopant in spiro-OMeTAD diffuse into the perovskite layer after thermal aging (Fig. 3E), matching well with the above SEM finding. With KPF_6, the diffusion of ions is considerably suppressed (Fig. 3F).

We also measured the light illumination stability of the devices. The solar cells were continuously measured five times under 1 sun AM 1.5G solar illumination. A rapid degradation of V_{oc} and FF for the 1.0-cm^2 3D/2D PSC is observed, whereas the KPF_6-modified device exhibits negligible degradation (fig. S22). The long-term light illumination stability was also characterized, as shown in fig. S23A. The champion device with KPF_6 additives exhibits a better photostability than the pristine device, retaining 82% of its initial efficiency after 500 hours under continuous 1 sun AM 1.5G solar illumination without a UV filter, at an open-circuit condition that is harsher than the operating condition (*33*). A more stable statistic PCE evolution of the devices with KPF_6 additive compared with the pristine devices is shown in fig. S23B. Therefore, the KPF_6 additive plays an important role in improving the stability of perovskite devices as well.

Just as the dense perovskite films can be easily prepared by spin-coating with the developed ink at a relatively low speed, similarly a moderate dry-air gas can also promote nucleation (fig. S24A). Thus, a gas-assist slot-die printing technology toward the continuous deposition of large-area perovskite films is developed here, as demonstrated in Fig. 4A. With a low pressure of ~0.3-MPa dry air blowing, the printed wet perovskite film quickly changes to a brown color with a mirror-like surface, as shown in movie S1. The inset in Fig. 4A shows a photograph of the printed large-area 20-cm–by–20-cm perovskite film. A pinhole-free high-quality perovskite layer with clear grain boundaries is shown in the SEM image (fig. S24B).

The 5-cm–by–5-cm PSMs are then fabricated, and the *J-V* curves are shown in Fig. 4B. The inset shows the schematic diagram of a six-subcell series-connected 5-cm–by–5-cm module, and fig. S25 shows its photo. The slot die–printed PSM shows comparable performance to the spin-coated counterpart (Fig. 4B), with efficiencies greater than 20.4% (table S3). Certified efficiencies of 19.3 and 18.9% with a mask area of 17.1 cm^2 are achieved, respectively, for the slot die–printed and spin-coated PSMs (figs. S26 and S27). The hysteresis-suppression effect of KPF_6 is also proven in large solar modules, which exhibit a very stable efficiency (fig. S28, A and B). We also achieved a champion FF of 0.806 for a 4-cm–by–4-cm solar module (mask area of 10.0 cm^2) (fig. S28C), which is the highest FF recorded among the reported PSMs. Furthermore, the solar module also shows excellent performance under weak light illumination and continuous multiple testing (fig. S28D). Our results indicate that the very high uniformity of large-area perovskite films that is achieved from this lead halide–templated crystallization strategy contributes to high-performance solar modules.

We further scaled up the perovskite films to print a 10-cm–by–10-cm solar module with a series connection of 14 subcells and demonstrated a hysteresis-free solar module with high efficiency of 19.54% under RS and 19.22% under FS with a mask area of 65.0 cm^2 (Fig. 4C). The module efficiencies from different upscaling methods with different areas that were achieved in the past several years are summarized in Fig. 4D. The antisolvent-free modulated high-quality perovskite films in this work exhibit the highest efficiencies among all the reported works, indicating the high processability for achieving large-area high-quality perovskite films.

REFERENCES AND NOTES

1. National Renewable Energy Laboratory (NREL), Best research-cell efficiency chart (2021); www.nrel.gov/pv/cell-efficiency.html.
2. S. D. Stranks *et al*., *Science* **342**, 341–344 (2013).
3. D. Shi *et al*., *Science* **347**, 519–522 (2015).
4. Q. Dong *et al*., *Science* **347**, 967–970 (2015).
5. F. Huang *et al*., *Nano Energy* **10**, 10–18 (2014).
6. M. Xiao *et al*., *Angew. Chem. Int. Ed.* **53**, 9898–9903 (2014).
7. Y. Y. Kim *et al*., *Nat. Commun.* **11**, 5146 (2020).
8. H. Chen *et al*., *Nature* **550**, 92–95 (2017).
9. Y. Deng *et al*., *Sci. Adv.* **5**, eaax7537 (2019).
10. Z. Xu *et al*., *J. Mater. Chem. A Mater. Energy Sustain.* **7**, 26849–26857 (2019).
11. Y. Deng *et al*., *Nat. Energy* **3**, 560–566 (2018).
12. C. Li *et al*., *J. Am. Chem. Soc.* **141**, 6345–6351 (2019).
13. J. Ding *et al*., *Joule* **3**, 402–416 (2019).
14. B. Conings *et al*., *Adv. Energy Mater.* **5**, 1500477 (2015).
15. E. J. Juarez-Perez *et al*., *J. Mater. Chem. A Mater. Energy Sustain.* **6**, 9604–9612 (2018).
16. S.-H. Turren-Cruz, A. Hagfeldt, M. Saliba, *Science* **362**, 449–453 (2018).
17. Y.-H. Lin *et al*., *Science* **369**, 96–102 (2020).
18. X. X. Gao *et al*., *Adv. Mater.* **32**, e1905502 (2020).
19. M. Jeong *et al*., *Science* **369**, 1615–1620 (2020).
20. M. Du *et al*., *Adv. Mater.* **32**, e2004979 (2020).
21. N. Ahn *et al*., *J. Am. Chem. Soc.* **137**, 8696–8699 (2015).
22. F. Guo *et al*., *Adv. Funct. Mater.* **29**, 1900964 (2019).
23. M. Saliba *et al*., *Energy Environ. Sci.* **9**, 1989–1997 (2016).
24. Q. Jiang *et al*., *Nat. Energy* **2**, 16177 (2016).
25. A. Ren *et al*., *Joule* **4**, 1263–1277 (2020).
26. M. I. Saidaminov *et al*., *Nat. Energy* **3**, 648–654 (2018).
27. Y. Liu *et al*., *Angew. Chem. Int. Ed.* **59**, 15688–15694 (2020).
28. D. Luo *et al*., *Science* **360**, 1442–1446 (2018).
29. P. Liu, W. Wang, S. Liu, H. Yang, Z. Shao, *Adv. Energy Mater.* **9**, 1803017 (2019).
30. Z. Liu *et al*., *Nat. Commun.* **9**, 3880 (2018).
31. E. A. Duijnstee *et al*., *ACS Energy Lett.* **5**, 376–384 (2020).
32. J. Chen, S.-G. Kim, N.-G. Park, *Adv. Mater.* **30**, e1801948 (2018).
33. K. Domanski, E. A. Alharbi, A. Hagfeldt, M. Grätzel, W. Tress, *Nat. Energy* **3**, 61–67 (2018).
34. T. Bu *et al*., *Solar RRL* **4**, 1900263 (2019).
35. J. Li *et al*., *Joule* **4**, 1035–1053 (2020).
36. X. Dai *et al*., *Adv. Energy Mater.* **10**, 1903108 (2019).

ACKNOWLEDGMENTS

Funding: This work is financially supported by the National Key Research and Development Plan (2019YFE0107200 and 2017YFE0131900), the National Natural Science Foundation of China (91963209 and 21875178), the Technological Innovation Key Project of Hubei Province (2018AAA048), and the Foshan Xianhu Laboratory of the Advanced Energy Science and Technology Guangdong Laboratory (XHD2020-001 and XHT2020-005). The Analytical and Testing Centre of Wuhan University of Technology and Hubei Key Laboratory of Low Dimensional Optoelectronic Material and Devices, Hubei University of Arts and Science, are acknowledged for the XRD and SEM characterizations. G.T., L.K.O., and Y.Q. acknowledge funding support from the Energy Materials and Surface Sciences Unit of the Okinawa Institute of Science and Technology Graduate University. **Author contributions:** F.H. and T.B. conceived the ideas and designed the experiments. Y.-B.C. provided helpful advice on the work. T.B. conducted the corresponding device and module fabrication and basic characterization. J.L., H.L., and C.T. helped with the module fabrication and encapsulation. J.L. and C.W. helped with the efficiency certification of modules. T.B. and J.S. conducted the DFT calculations. G.T. and L.K.O. helped with the XPS, UPS characterization, and analyses. Z.L. and N.C. helped with the stability test. J.C., J.L., J.Z., X.-L.Z., W.H., Y.Q., and Y.-B.C. provided valuable suggestions for the manuscript. F.H. and T.B. participated in all of the data analyses. F.H. and T.B. wrote the paper, and all authors revised the paper. **Competing interests:** None declared. **Data and materials availability:** All data needed to evaluate the conclusions in the paper are present in the paper or the supplementary materials.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/372/6548/1327/suppl/DC1
Materials and Methods
Figs. S1 to S28
Tables S1 to S4
References (*37*–*43*)
Movie S1

16 February 2021; accepted 6 May 2021
10.1126/science.abh1035

OPTOMECHANICS

Approaching the motional ground state of a 10-kg object

Chris Whittle[1], Evan D. Hall[1], Sheila Dwyer[2], Nergis Mavalvala[1], Vivishek Sudhir[1,3]*, R. Abbott[4], A. Ananyeva[4], C. Austin[5], L. Barsotti[1], J. Betzwieser[6], C. D. Blair[6,7], A. F. Brooks[4], D. D. Brown[8], A. Buikema[1], C. Cahillane[4], J. C. Driggers[2], A. Effler[6], A. Fernandez-Galiana[1], P. Fritschel[1], V. V. Frolov[6], T. Hardwick[5], M. Kasprzack[4], K. Kawabe[2], N. Kijbunchoo9, J. S. Kissel2, G. L. Mansell1·2, F. Matichard1·4, L. McCuller1, T. McRae9, A. Mullavey[6], A. Pele[6], R. M. S. Schofield[10], D. Sigg[2], M. Tse[1], G. Vajente[4], D. C. Vander-Hyde[11], Hang Yu[1], Haocun Yu[1], C. Adams[6], R. X. Adhikari[4], S. Appert[4], K. Arai[4], J. S. Areeda[12], Y. Asali[13], S. M. Aston[6], A. M. Baer[14], M. Ball[10], S. W. Ballmer[11], S. Banagiri[15], D. Barker[2], J. Bartlett[2], B. K. Berger[16], D. Bhattacharjee[17], G. Billingsley[4], S. Biscans[1,4], R. M. Blair[2], N. Bode[18,19], P. Booker[18,19], R. Bork[4], A. Bramley[6], K. C. Cannon[20], X. Chen[7], A. A. Ciobanu[8], F. Clara[2], C. M. Compton[2], S. J. Cooper[21], K. R. Corley[13], S. T. Countryman[13], P. B. Covas[22], D. C. Coyne[4], L. E. H. Datrier[23], D. Davis[11], C. Di Fronzo[21], K. L. Dooley[24,25], P. Dupej[23], T. Etzel[4], M. Evans[1], T. M. Evans[6], J. Feicht[4], P. Fulda[26], M. Fyffe[6], J. A. Giaime[5,6], K. D. Giardina[6], P. Godwin[27], E. Goetz[5,17,28], S. Gras[1], C. Gray[2], R. Gray[23], A. C. Green[26], E. K. Gustafson[4], R. Gustafson[29], J. Hanks[2], J. Hanson[6], R. K. Hasskew[6], M. C. Heintze[6], A. F. Helmling-Cornell[10], N. A. Holland[9], J. D. Jones[2], S. Kandhasamy[30], S. Karki[10], P. J. King[2], Rahul Kumar[2], M. Landry[2], B. B. Lane[1], B. Lantz[16], M. Laxen[6], Y. K. Lecoeuche[28], J. Leviton[29], J. Liu[18,19], M. Lormand[6], A. P. Lundgren[31], R. Macas[24], M. MacInnis[1], D. M. Macleod[24], S. Márka[13], Z. Márka[13], D. V. Martynov[21], K. Mason[1], T. J. Massinger[1], R. McCarthy[2], D. E. McClelland[9], S. McCormick[6], J. McIver[4,28], G. Mendell[2], K. Merfeld[10], E. L. Merilh[2], F. Meylahn[18,19], T. Mistry[32], R. Mittleman[1], G. Moreno[2], C. M. Mow-Lowry[21], S. Mozzon[31], T. J. N. Nelson[6], P. Nguyen[10], L. K. Nuttall[31], J. Oberling[2], Richard J. Oram[6], C. Osthelder[4], D. J. Ottaway[8], H. Overmier[6], J. R. Palamos[10], W. Parker[6,33], E. Payne[34], R. Penhorwood[29], C. J. Perez[2], M. Pirello[2], H. Radkins[2], K. E. Ramirez[35], J. W. Richardson[4], K. Riles[29], N. A. Robertson[4,23], J. G. Rollins[4], C. L. Romel[2], J. H. Romie[6], M. P. Ross[36], K. Ryan[2], T. Sadecki[2], E. J. Sanchez[4], L. E. Sanchez[4], T. R. Saravanan[30], R. L. Savage[2], D. Schaetz[4], R. Schnabel[37], E. Schwartz[6], D. Sellers[6], T. Shaffer[2], B. J. J. Slagmolen[9], J. R. Smith[12], S. Soni[5], B. Sorazu[23], A. P. Spencer[23], K. A. Strain[23], L. Sun[4,9], M. J. Szczepa czyk[26], M. Thomas[6], P. Thomas[2], K. A. Thorne[6], K. Toland[23], C. I. Torrie[4], G. Traylor[6], A. L. Urban[5], G. Valdes[5], P. J. Veitch[8], K. Venkateswara[36], G. Venugopalan[4], A. D. Viets[38], T. Vo[11], C. Vorvick[2], M. Wade[39], R. L. Ward[9], J. Warner[2], B. Weaver[2], R. Weiss[1], B. Willke[18,19], C. C. Wipf[4], L. Xiao[4], H. Yamamoto[4], L. Zhang[4], M. E. Zucker[1,4], J. Zweizig[4]

The motion of a mechanical object, even a human-sized object, should be governed by the rules of quantum mechanics. Coaxing them into a quantum state is, however, difficult because the thermal environment masks any quantum signature of the object's motion. The thermal environment also masks the effects of proposed modifications of quantum mechanics at large mass scales. We prepared the center-of-mass motion of a 10-kilogram mechanical oscillator in a state with an average phonon occupation of 10.8. The reduction in temperature, from room temperature to 77 nanokelvin, is commensurate with an 11 orders-of-magnitude suppression of quantum back-action by feedback and a 13 orders-of-magnitude increase in the mass of an object prepared close to its motional ground state. Our approach will enable the possibility of probing gravity on massive quantum systems.

The apparent classical behavior of tangibly massive objects is, according to conventional quantum mechanics, the symptom of decoherence. Thermal decoherence, caused by the interaction of a system with a thermal environment, is by far the most pervasive. For a mechanical oscillator of mass m and natural frequency Ω_0, thermal decoherence induces motion characterized by the spectral density $S_x^{\mathrm{th}}[\Omega_0] = (2n_{\mathrm{th}}[\Omega_0] + 1)S_x^{\mathrm{zp}}[\Omega_0]$, where $n_{\mathrm{th}}[\Omega_0] \approx k_B T/\hbar\Omega_0$ is the average thermal phonon occupation caused by the environment (at temperature T) and $S_x^{\mathrm{zp}}[\Omega_0] = 8x_{\mathrm{zp}}^2/\Gamma_0[\Omega_0]$ is its motional zero-point fluctuation, $x_{\mathrm{zp}} = \sqrt{\hbar/(2m\Omega_0)}$, concentrated in a frequency band of width $\Gamma_0[\Omega_0]$. Thermal fluctuations obscure signatures of decoherence that allegedly arise from modifications of quantum mechanics at large masses (*1*–*4*) and limit the sensitivity of mechanical transducers in metrology applications (*5*). Techniques to probe both frontiers call for large mass mechanical objects prepared in pure quantum states.

Over the past decade, progressively larger objects all the way to nanomechanical oscillators have been prepared in their motional ground state (*6*–*10*). Most of these experiments relied on isolating the oscillator in an elastic or electromagnetic trap in the ≥100-kHz frequency range, embedded in a sideband-resolved electromagnetic cavity, typically in a cryogenic environment. These methods do not address a number of technical challenges specific to mechanical oscillators above the milligram-to-gram mass scale. For one, the large optical power required to trap massive oscillators introduces extraneous heating and other optomechanical nonlinearities. Furthermore, the low resonant frequency of large suspended oscillators doubly compounds the problem of thermal decoherence by increasing the intrinsic thermal motion ($n_{\mathrm{tot}} \propto 1/\Omega_0$) and precluding efficient cavity sideband cooling. Therefore, a different route is needed to prepare large-mass oscillators in pure quantum states.

The Advanced LIGO gravitational-wave detectors offer a unique perspective on this problem. Advanced LIGO is a pair of Michelson interferometers, each with 4-km-long Fabry–Pérot arm cavities formed by 40-kg mirrors that hang on fused silica fibers (Fig. 1A). The differential motion of each pair of arm cavity mirrors forms a mechanical oscillator with a reduced mass of 20 kg; the differential motion of each such oscillator in either arm, sensed by the Michelson interferometer, forms a mechanical oscillator of effective mass m = 10 kg that is the object of our attention. The oscillator follows the pendulum-like motion of the suspended mirror at a frequency $\Omega_0 \approx 2\pi \cdot 0.43$ Hz; gravitational stress dilution is expected to realize a quality factor of $Q_0 \approx 10^8$ (*11*). Its

[1]Laser Interferometer Gravitational Wave Observatory (LIGO), Massachusetts Institute of Technology, Cambridge, MA 02139, USA. [2]LIGO Hanford Observatory, Richland, WA 99352, USA. [3]Department of Mechanical Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139, USA. [4]LIGO, California Institute of Technology, Pasadena, CA 91125, USA. [5]Louisiana State University, Baton Rouge, LA 70803, USA. [6]LIGO Livingston Observatory, Livingston, LA 70754, USA. [7]OzGrav, University of Western Australia, Crawley, Western Australia 6009, Australia. [8]OzGrav, University of Adelaide, Adelaide, South Australia 5005, Australia. [9]OzGrav, Australian National University, Canberra, Australian Capital Territory 0200, Australia. [10]University of Oregon, Eugene, OR 97403, USA. [11]Syracuse University, Syracuse, NY 13244, USA. [12]California State University Fullerton, Fullerton, CA 92831, USA. [13]Columbia University, New York, NY 10027, USA. [14]Christopher Newport University, Newport News, VA 23606, USA. [15]University of Minnesota, Minneapolis, MN 55455, USA. [16]Stanford University, Stanford, CA 94305, USA. [17]Missouri University of Science and Technology, Rolla, MO 65409, USA. [18]Max Planck Institute for Gravitational Physics (Albert Einstein Institute), D-30167 Hannover, Germany. [19]Leibniz Universität Hannover, D-30167 Hannover, Germany. [20]RESCEU, University of Tokyo, Tokyo 113-0033, Japan. [21]University of Birmingham, Birmingham B15 2TT, UK. [22]Universitat de les Illes Balears, IAC3-IEEC, E-07122 Palma de Mallorca, Spain. [23]SUPA, University of Glasgow, Glasgow G12 8QQ, UK. [24]Cardiff University, Cardiff CF24 3AA, UK. [25]The University of Mississippi, University, MS 38677, USA. [26]University of Florida, Gainesville, FL 32611, USA. [27]The Pennsylvania State University, University Park, PA 16802, USA. [28]University of British Columbia, Vancouver, British Columbia V6T 1Z4, Canada. [29]University of Michigan, Ann Arbor, MI 48109, USA. [30]Inter-University Centre for Astronomy and Astrophysics, Pune 411007, India. [31]University of Portsmouth, Portsmouth PO1 3FX, UK. [32]The University of Sheffield, Sheffield S10 2TN, UK. [33]Southern University and A&M College, Baton Rouge, LA 70813, USA. [34]OzGrav, School of Physics & Astronomy, Monash University, Clayton 3800, Victoria, Australia. [35]The University of Texas Rio Grande Valley, Brownsville, TX 78520, USA. [36]University of Washington, Seattle, WA 98195, USA. [37]Universität Hamburg, D-22761 Hamburg, Germany. [38]Concordia University Wisconsin, Mequon, WI 53097, USA. [39]Kenyon College, Gambier, OH 43022, USA.
*Corresponding author. Email: vivishek@mit.edu

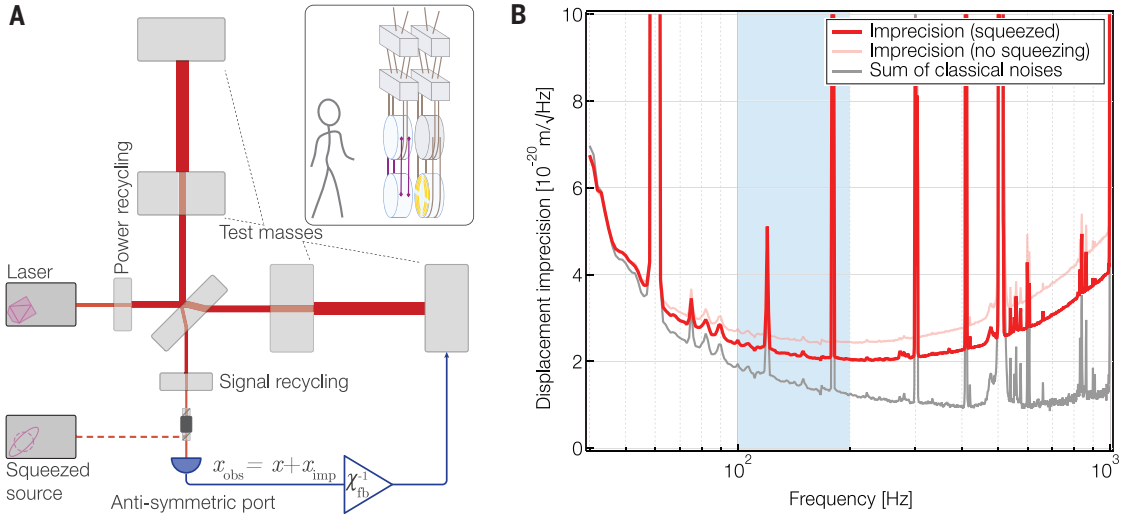


Fig. 1. Advanced LIGO interferometer. (**A**) Laser light is split and recombined at a beam-splitter, forming a Michelson interferometer. Its response is shaped by the Fabry–Pérot cavities in the arms and the signal-recycling mirror. The power-recycling mirror and injection of squeezed light enhances the sensitivity. Inset shows the suspension system of each of the four 40-kg mirrors. The final mass on the forward chain is the 40-kg mirror, suspended on fused silica wires (purple) featuring a quality factor $Q \approx 8 \cdot 10^7$; they can be displaced by electrostatic forces caused by voltages applied on electrodes (yellow) etched onto the reaction mass suspended behind it; average adult human sketched for scale. (**B**) The displacement sensitivity (red) of the interferometer is $2 \cdot 10^{-20}\text{m}/\sqrt{\text{Hz}}$ at 100 to 200 Hz, where it is largely shot noise (light red), suppressed by ~3 dB from injection of squeezed vacuum (red) and a combination of extraneous technical noises (gray). Blue band shows the frequency interval in which the pendulum mode is trapped and cooled.

displacement fluctuates because of the presence of $n_{\text{th}}[\Omega_0] \approx 10^{13}$ phonons. The interferometer resonantly transduces the differential arm motion into optical power fluctuations at the anti-symmetric port, which is sensed by homodyne detection. (During ordinary operation, these fluctuations encode the gravitational-wave signals). In this state, the homodyne photocurrent fluctuations bear the apparent displacement $\delta x_{\text{obs}} = \delta x + \delta x_{\text{imp}}$; here, δx is the physical motion of the differential arm, which contains the displacement of the oscillator, and δx_{imp} is the measurement imprecision. The imprecision noise, depicted in Fig. 1B, is $\sqrt{S_x^{\text{imp}}} \approx 2 \cdot 10^{-20}\text{m}/\sqrt{\text{Hz}}$ around 100 to 200 Hz and is largely quantum shot noise, suppressed by ~3 dB by injection of squeezed light (*12*) and shaped by the response of the signal recycling cavity, with a secondary contribution from mechanical dissipation in the mirror coatings (*13*). This sensitivity is equivalent to $n_{\text{imp}} \equiv S_x^{\text{imp}}/2S_x^{\text{zp}} \approx 3.5 \cdot 10^{-13}$ phonons for a 10-kg oscillator at ~150 Hz, a record low number [Rossi *et al.* (*9*) demonstrated $n_{\text{imp}} \approx 10^{-7}$] tantamount to resolving the zero-point motion of the oscillator with ~125 dB signal to squeezed-shot-noise ratio and comparable to the requirement to feedback cool the oscillator to its ground state [$n_{\text{imp}} \sim 1/2n_{\text{th}}$, for a viscously damped oscillator (*14*)].

To take advantage of this precision, we actively stiffened the pendulum mode by synthesizing a force proportional to the observed displacement (i.e., $\propto \Omega_{\text{fb}}^2 \delta x_{\text{obs}}$) and in-phase with the motion δx, trapping the pendulum mode as an oscillator around $\Omega_{\text{fb}} \approx 2\pi \cdot 148$ Hz. Two additional sources of decoherence plague this scheme. First, such measurement precision comes at the expense of additional quantum back-action on the pendulum mode: Radiation pressure shot noise from the 200 kW intracavity power and the anti-squeezed intracavity field produces motion (*15*) equivalent to $n_{\text{ba}}[\Omega_0] \approx 1.0 \cdot 10^{12}$ phonons. However, as long as the measurement record resolves the quantum back-action at a rate comparable to the thermal decoherence, active feedback can suppress it (*9*, *14*, *16*). Second, the feedback of amplified imprecision noise leads to an additional "feedback back-action," $n_{\text{fb}} \approx Q_0^2(\Omega_{\text{fb}}/\Omega_0)^4 n_{\text{imp}}$ [see section S1 of the supplementary materials (*17*)], which increases with the trap frequency. However, this is partially compensated by the $\Omega_{\text{fb}}/\Omega_0 \approx 300$-fold reduction in both the thermal occupation and decay rate of the trapped oscillator caused by structural damping (*5*).

To trap and damp the oscillator, we adjust the feedback control so that $\delta F_{\text{fb}} = \chi_{\text{fb}}^{-1}\delta x_{\text{obs}}$, with a feedback filter $\chi_{\text{fb}}^{-1} \propto \Omega_{\text{fb}}^2 + i\Omega\Gamma_{\text{fb}}$ between 100 and 200 Hz. This is implemented by careful shaping of the control loop that is otherwise used to stabilize the interferometer at its linear operating point. The feedback force is applied on the mirror electrostatically: Gold electrodes on the reaction mass (Fig. 1A) are held at a 400 V bias, and their fringing field polarizes the dielectric test mass. Control voltages added on interleaved electrodes produce a proportional force [extraneous force noise produces <1 phonon of excess occupation on average, see section S1.1 of the supplementary materials (*17*)]. The overall feedback gain is adjusted so that the system's effective susceptibility takes the form, $\chi_{\text{eff}}[\Omega] \propto \left(-\Omega^2 + \Omega_{\text{eff}}^2 + i\Omega\Gamma_{\text{eff}}[\Omega]\right)^{-1}/m$, of that of an oscillator with frequency $\Omega_{\text{eff}} = \sqrt{\Omega_0^2 + \Omega_{\text{fb}}^2} \approx \Omega_{\text{fb}} \approx 2\pi \cdot 148$ Hz. Delays in the feedback loop limit the trap frequency and cause the oscillator to be intrinsically "cold-damped." In particular, the phase response of the notch filters used to prevent excitation of the violin modes of the suspension (at 500 Hz and harmonics, featuring quality factors $\geq 10^9$) in conjunction with the feedback filter leaves the interferometer's length control system with a phase margin of 1° for a trap frequency of 148 Hz. Physical delay in the loop also cold-damps the trapped oscillator to a quality factor of ≈50 [Fig. 2B, red trace; see section S1.2 of the supplementary materials (*17*) for further details]. The oscillator is damped further by modifying the imaginary part of the feedback filter. Figure 2A shows the effective susceptibilities of the trapped and damped oscillator so realized. The largest damping rate, corresponding to a quality factor of ≈1, is limited by the gain margin ($\approx 10^{-3}$) of the control loop. Around the trap frequency, 100 to 200 Hz, additional force noise on the oscillator caused by feedback is dominated by subquantum fluctuations of the squeezed imprecision noise.

The calibrated in-loop signal, depicted in Fig. 2B, shows the apparent displacement fluctuations of the trapped and damped oscillator (δx_{obs}). This can be understood using a simple model [see section S1 of the supplementary materials (*17*)], $\delta x_{\text{obs}} = \chi_{\text{eff}}\left(\delta F_{\text{th}} + \delta F_{\text{ba}} - \chi_{\text{fb}}^{-1}\delta x_{\text{imp}}\right) + \delta x_{\text{imp}}$. It describes the oscillator, characterized by intrinsic susceptibility χ_0, with a displacement that responds through the feedback-modified effective susceptibility $\chi_{\text{eff}} = \left(\chi_0^{-1} + \chi_{\text{fb}}^{-1}\right)^{-1}$ to three forces: a frequency-dependent structural thermal force (δF_{th}), a white quantum measurement back-action force (δF_{ba}), and an additional force noise ($\propto \chi_{\text{fb}}^{-1}\delta x_{\text{imp}}$) caused by feedback of imprecision noise through the feedback filter. These physical displacements ride on the imprecision noise (δx_{imp}) to yield

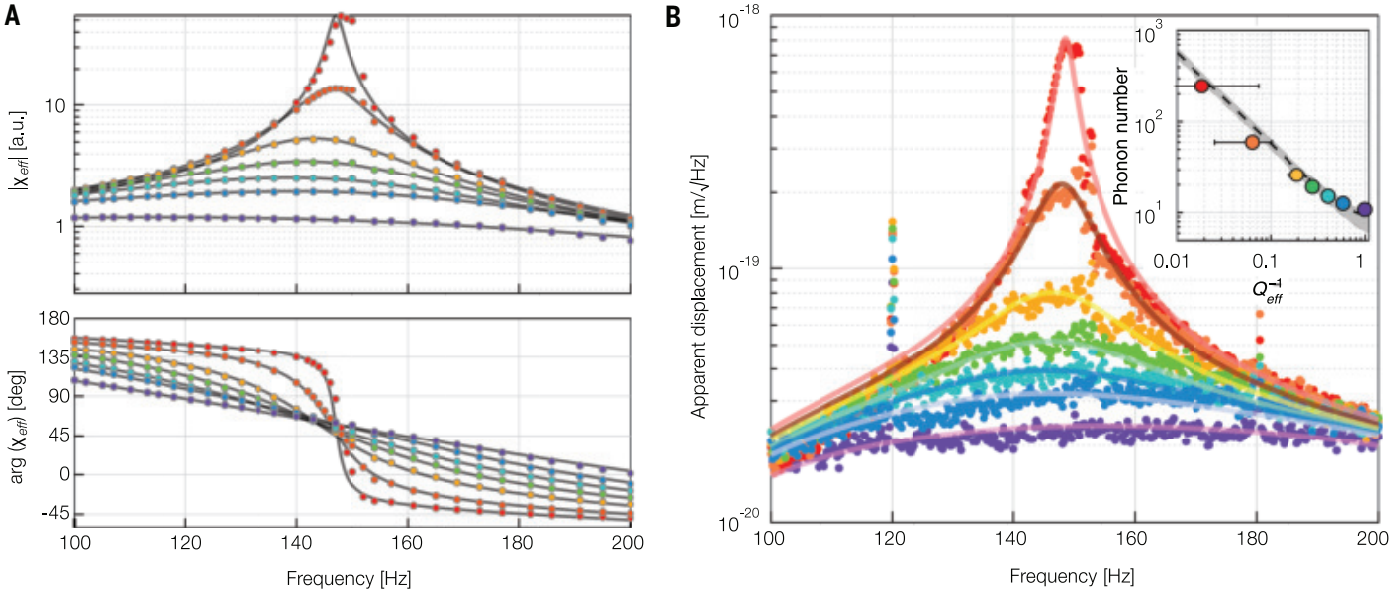


Fig. 2. Trapping and cooling of a 10-kg oscillator to 10 quanta. (**A**) Effective susceptibility of the oscillator for each setting of the damping filter, measured by exciting the feedback loop at each frequency and demodulating its response at the same frequency. The lines show fits to a model of the susceptibility of a damped harmonic oscillator with an additional delay, i.e., $\chi_{\mathrm{eff}}[\Omega]e^{i\Omega\tau}$. Fits to the phase response produce τ = 0.9 ms. (**B**) Displacement spectrum of the oscillator as the damping is increased. Solid lines show fits to a model of the observed spectrum S_x^{obs} (see text for details) where the effective susceptibility is determined by the response measurements in (A) and only the frequency-dependent imprecision noise and force noise are variable. Inset shows the inferred average phonon occupation for each of the curves in the main panel as a function of the damping quality factor. Also shown is a model (black dashed line) with model uncertainties (gray band). [The disagreement between the simple model and data, both in the transfer functions and spectra, around 150 to 155 Hz arises from a coupling between the motion of the pendulum and the upper intermediate mass of the suspension (*18*)].

the observed displacement δx_{obs}. The spectra of the observed displacement S_x^{obs} predicted by this model are shown as solid lines in Fig. 2B. In the model, the effective susceptibility is fully determined by the response measurements shown in Fig. 2A and is independent of the frequency-dependent force noise and imprecision noise. The latter, determined self-consistently among the displacement noise in Fig. 2B, shows a variation between the different feedback settings of <1%, consistent with expected drift in the Advanced LIGO interferometer over the ~2-hour time scale over which the experiment was performed. Several sources of uncertainty are accounted for in this process. Calibration of the displacement spectra contributes ≈2% uncertainty (*18*). Uncertainties in the effective susceptibility χ_{eff}, from fits to Fig. 2A, are at the 1% level, limited by the 1-s averaging used per point in measurements of the response [see section S2 of the supplementary materials (*17*)]. The dominant uncertainty is in the fits to the displacement spectra of Fig. 2B using the model for S_x^{obs}: The frequency dependence of the imprecision noise and structurally damped thermal force noise produce a ≈5% variation between the various spectra in Fig. 2B.

The effective phonon occupation (n_{eff}) of the cooled oscillator can be defined by $\hbar\Omega_{\mathrm{eff}}(n_{\mathrm{eff}}+\frac{1}{2}) = \langle p^2/(2m) + m\Omega_{\mathrm{eff}}^2 x^2/2\rangle$, where x (p) is the physical displacement (momentum) of the oscillator at the trap frequency. Assuming the displacement and momentum to be zero-mean, their second moments can be estimated as the integral of their spectral densities. In principle, two factors complicate this procedure: At lower frequencies, structural damping renders the displacement variance singular (*5*), whereas at higher frequencies,

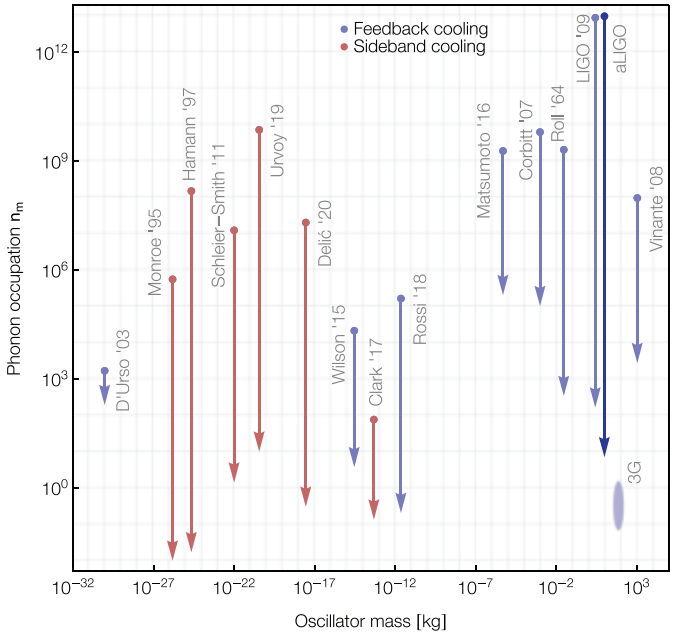


Fig. 3. A selection of oscillator cooling experiments. The initial occupations mentioned are those of the relevant oscillator mode as defined by the natural trap frequency at its ambient temperature (*9*, *10*, *14*, *20*–*30*). For atomic physics experiments, this is usually at room temperature in a harmonic electromagnetic trap. For most solid-state mechanical oscillators, it is the harmonic mode defined by the Hookean restoring force of its elastic suspension and typically at cryogenic temperatures [the exception is the recent work from Delić *et al.* (*10*), which demonstrated cavity cooling of an electromagnetically trapped nanoparticle to its ground state]. Our result ("aLIGO") sets a new record in the macroscopic mass range, reaching 10.8 ± 0.8 phonons. Experiments with future gravitational-wave interferometers ("3G") will achieve occupations <1.

feedback back-action precludes a finite momentum variance (*19*). In practice, the feedback filter $\chi_{\mathrm{fb}}^{-1} \propto \Omega_{\mathrm{fb}}^2 + i\Omega\Gamma_{\mathrm{fb}}$ is established around 100 to 200 Hz in an envelope that falls off at least as Ω^{-2} (at frequencies <10 Hz, the interferometer's length control loop picks up again), which regulates both of these problems. In this fashion, within 100 to 200 Hz, the trapped oscillator approximately satisfies the equipartition principle, so an effective phonon occupation can be assigned using the physical displacement spectrum as follows:

$$n_{\mathrm{eff}} \approx \int \frac{S_x[\Omega]}{2x_{\mathrm{zp}}^2} \frac{\mathrm{d}\Omega}{2\pi}$$

Note that the 100-Hz frequency band in which the oscillator is established is much larger than the expected decoherence rate of the trapped oscillator, $(n_{\mathrm{th}}[\Omega_{\mathrm{eff}}] + n_{\mathrm{ba}} + n_{\mathrm{fb}}[\Omega_{\mathrm{eff}}])\Gamma_0[\Omega_{\mathrm{eff}}] \approx 2\pi \cdot 10$ Hz. We evaluated the integral using the physical displacement spectrum reconstructed from fits to the observed displacement. The minimum phonon occupation of the 10-kg oscillator, corresponding to the purple trace in Fig. 2B, is thus inferred to be 10.8 ± 0.8; this is equivalent to an effective mode temperature of 77 nK. This demonstration sets a new record for the quantum state purity (≈10% ground state fidelity) for an object of such large mass (Fig. 3).

The preparation of massive objects progressively nearer their ground state opens the door for more sophisticated demonstrations and applications of macroscopic quantum phenomena and quantum metrology. The most intriguing possibility, however, harnesses the ready susceptibility of kilogram-scale masses to gravitational forces; with this work, it becomes possible to prepare them in near-quantum states. This hints at the tantalizing prospect of studying gravitational decoherence on massive quantum systems.

REFERENCES AND NOTES

1. F. Karolyhazy, *Nuovo Cim., A Gen. Phys.* **42**, 390–402 (1966).
2. L. Diósi, *Phys. Rev. A* **40**, 1165–1174 (1989).
3. R. Penrose, *Gen. Relativ. Gravit.* **28**, 581–600 (1996).
4. A. Bassi, A. Großardt, H. Ulbricht, *Class. Quantum Gravity* **34**, 193002 (2017).
5. P. R. Saulson, *Phys. Rev. D Part. Fields* **42**, 2437–2445 (1990).
6. J. Chan *et al.*, *Nature* **478**, 89–92 (2011).
7. J. D. Teufel *et al.*, *Nature* **475**, 359–363 (2011).
8. R. W. Peterson *et al.*, *Phys. Rev. Lett.* **116**, 063601 (2016).
9. M. Rossi, D. Mason, J. Chen, Y. Tsaturyan, A. Schliesser, *Nature* **563**, 53–58 (2018).
10. U. Delić *et al.*, *Science* **367**, 892–895 (2020).
11. A. V. Cumming *et al.*, *Class. Quantum Gravity* **29**, 035003 (2012).
12. M. Tse *et al.*, *Phys. Rev. Lett.* **123**, 231107 (2019).
13. A. Buikema *et al.*, *Phys. Rev. D* **102**, 062003 (2020).
14. D. J. Wilson *et al.*, *Nature* **524**, 325–329 (2015).
15. H. Yu *et al.*, *Nature* **583**, 43–47 (2020).
16. V. Sudhir *et al.*, *Phys. Rev. X* **7**, 011001 (2017).
17. See the supplementary materials.
18. L. Sun *et al.*, *Class. Quantum Gravity* **37**, 225008 (2020).
19. D. Vitali, S. Mancini, L. Ribichini, P. Tombesi, *J. Opt. Soc. Am. B* **20**, 1054 (2003).
20. B. D'Urso, B. Odom, G. Gabrielse, *Phys. Rev. Lett.* **90**, 043001 (2003).
21. C. Monroe *et al.*, *Phys. Rev. Lett.* **75**, 4011–4014 (1995).
22. S. E. Hamann *et al.*, *Phys. Rev. Lett.* **80**, 4149–4152 (1998).
23. M. H. Schleier-Smith, I. D. Leroux, H. Zhang, M. A. Van Camp, V. Vuletić, *Phys. Rev. Lett.* **107**, 143005 (2011).
24. A. Urvoy, Z. Vendeiro, J. Ramette, A. Adiyatullin, V. Vuletić, *Phys. Rev. Lett.* **122**, 203202 (2019).
25. J. B. Clark, F. Lecocq, R. W. Simmonds, J. Aumentado, J. D. Teufel, *Nature* **541**, 191–195 (2017).
26. N. Matsumoto, K. Komori, S. Ito, Y. Michimura, Y. Aso, *Phys. Rev. A (Coll. Park)* **94**, 033822 (2016).
27. T. Corbitt *et al.*, *Phys. Rev. Lett.* **99**, 160801 (2007).
28. P. G. Roll, R. Krotkov, R. H. Dicke, *Ann. Phys.* **26**, 442–517 (1964).
29. B. Abbott *et al.*, *New J. Phys.* **11**, 073032 (2009).
30. A. Vinante *et al.*, *Phys. Rev. Lett.* **101**, 033601 (2008).

ACKNOWLEDGMENTS

V.S. thanks V. Vuletić for useful conversations. **Funding:** This material is based upon work supported by the National Science Foundation's (NSF's) LIGO Laboratory, which is a major facility fully funded by the NSF. The authors gratefully acknowledge the support of the NSF for the construction and operation of the LIGO Laboratory and Advanced LIGO, as well as the Science and Technology Facilities Council (STFC) of the United Kingdom, and the Max-Planck-Society (MPS) for support of the construction of Advanced LIGO. Additional support for Advanced LIGO was provided by the Australian Research Council. The authors acknowledge the LIGO Scientific Collaboration Fellows program for additional support. LIGO was constructed by the California Institute of Technology and Massachusetts Institute of Technology with funding from the National Science Foundation and operates under cooperative agreement no. PHY-1764464. Advanced LIGO was built under award no. PHY-0823459. E.D.H. is supported by the MathWorks, Inc. This manuscript is LIGO document no. LIGO-P2000525. **Competing interests:** The authors declare no competing interests. **Author contributions:** V.S., E.D.H., and N.M. conceived this project. C.W., E.D.H., S.D., and V.S. designed and implemented the modifications to the Advanced LIGO detector that enabled the experiment. All authors contributed to the running, diagnostics, and calibration of the detector. C.W., E.D.H., and V.S. analyzed the data. V.S. wrote the manuscript with help from E.D.H., C.W., and N.M. V.S. developed the theoretical models and supervised the project. Other LIGO collaboration authors contributed to the design, construction, and operation of Advanced LIGO and the development and maintenance of data handling, data reduction, and data analysis. All authors meet the journal's authorship criteria and have reviewed, discussed, and commented on the results and the manuscript. **Data and materials availability:** All data needed to evaluate the conclusions of this study are provided in the main text or the supplementary materials.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/372/6548/1333/suppl/DC1
Supplementary Text
Fig. S1
References (*31*–*34*)

26 February 2021; accepted 5 May 2021
10.1126/science.abh2634

CORONAVIRUS

Clonal analysis of immunodominance and cross-reactivity of the CD4 T cell response to SARS-CoV-2

Jun Siong Low[1]†, Daniela Vaqueirinho[1]†, Federico Mele[1]†, Mathilde Foglierini[1], Josipa Jerak[1], Michela Perotti[1], David Jarrossay[1], Sandra Jovic[1], Laurent Perez[1], Rosalia Cacciatore[2], Tatiana Terrot[3], Alessandra Franzetti Pellanda[4], Maira Biggiogero[4], Christian Garzoni[4], Paolo Ferrari[5,6,7], Alessandro Ceschi[3,5,8,9], Antonio Lanzavecchia[10], Federica Sallusto[1,11]*‡, Antonino Cassotta[1]*‡

The identification of CD4$^+$ T cell epitopes is instrumental for the design of subunit vaccines for broad protection against coronaviruses. Here, we demonstrate in COVID-19–recovered individuals a robust CD4$^+$ T cell response to naturally processed severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) spike (S) protein and nucleoprotein (N), including effector, helper, and memory T cells. By characterizing 2943 S-reactive T cell clones from 34 individuals, we found that the receptor-binding domain (RBD) is highly immunogenic and that 33% of RBD-reactive clones and 94% of individuals recognized a conserved immunodominant S346–S365 region comprising nested human leukocyte antigen DR (HLA-DR)– and HLA-DP–restricted epitopes. Using pre– and post–COVID-19 samples and S proteins from endemic coronaviruses, we identified cross-reactive T cells targeting multiple S protein sites. The immunodominant and cross-reactive epitopes identified can inform vaccination strategies to counteract emerging SARS-CoV-2 variants.

The identification of T cell epitopes in disease-causing organisms is challenging in view of the polymorphism of human leukocyte antigen (HLA) molecules and the variability of rapidly mutating pathogens. In the context of the COVID-19 pandemic, bioinformatic analysis (*1*) has been used to predict T cell epitopes in severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) proteins (*2*, *3*) and to produce peptide pools to stimulate peripheral blood mononuclear cells (PBMCs) and enumerate antigen-specific T cells. These studies revealed a robust CD4$^+$ and CD8$^+$ T cell response against SARS-CoV-2 proteins in recovered patients (*2*–*6*) and a level of cross-reactivity with endemic coronaviruses in pre-pandemic samples (*7*–*9*).

A limitation of bioinformatics predictions is the difficulty in identifying immunodominant epitopes, because immunodominance is determined by multiple factors such as antigen processing, T cell repertoire, HLA alleles, and preexisting cross-reactive immunity (*10*–*12*). To identify naturally processed immunodominant CD4$^+$ T cell epitopes, we took the unbiased approach of stimulating T memory (T_M) cells with protein-pulsed antigen-presenting cells (APCs), followed by the isolation of T cell clones to precisely map the epitope recognized (*13*).

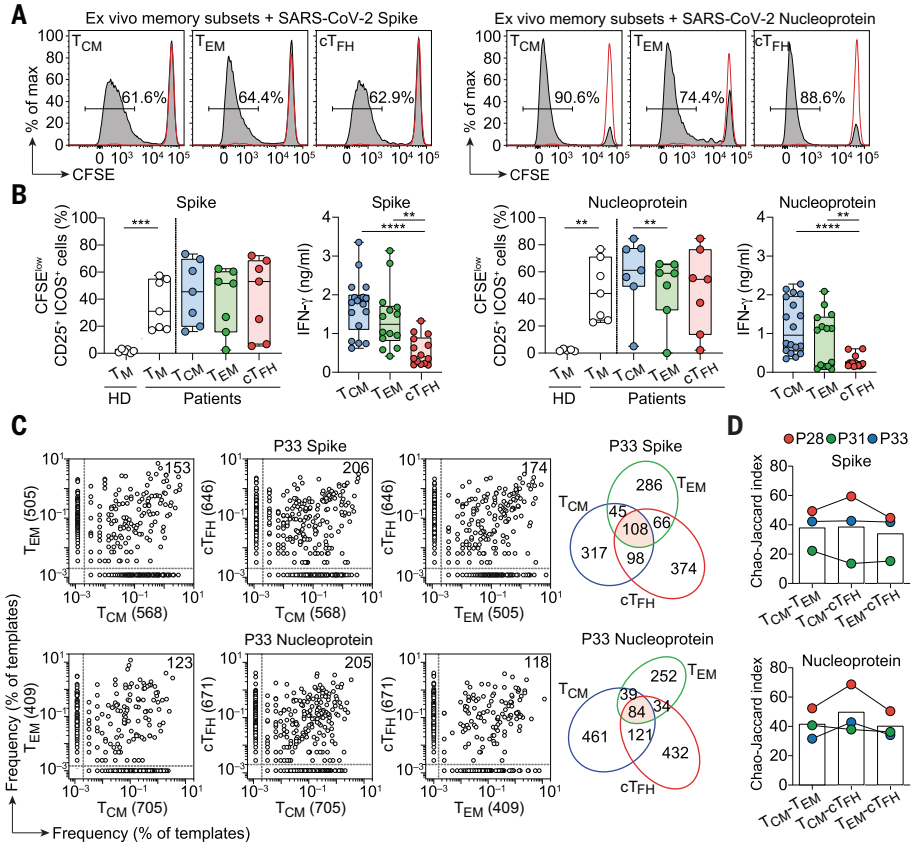


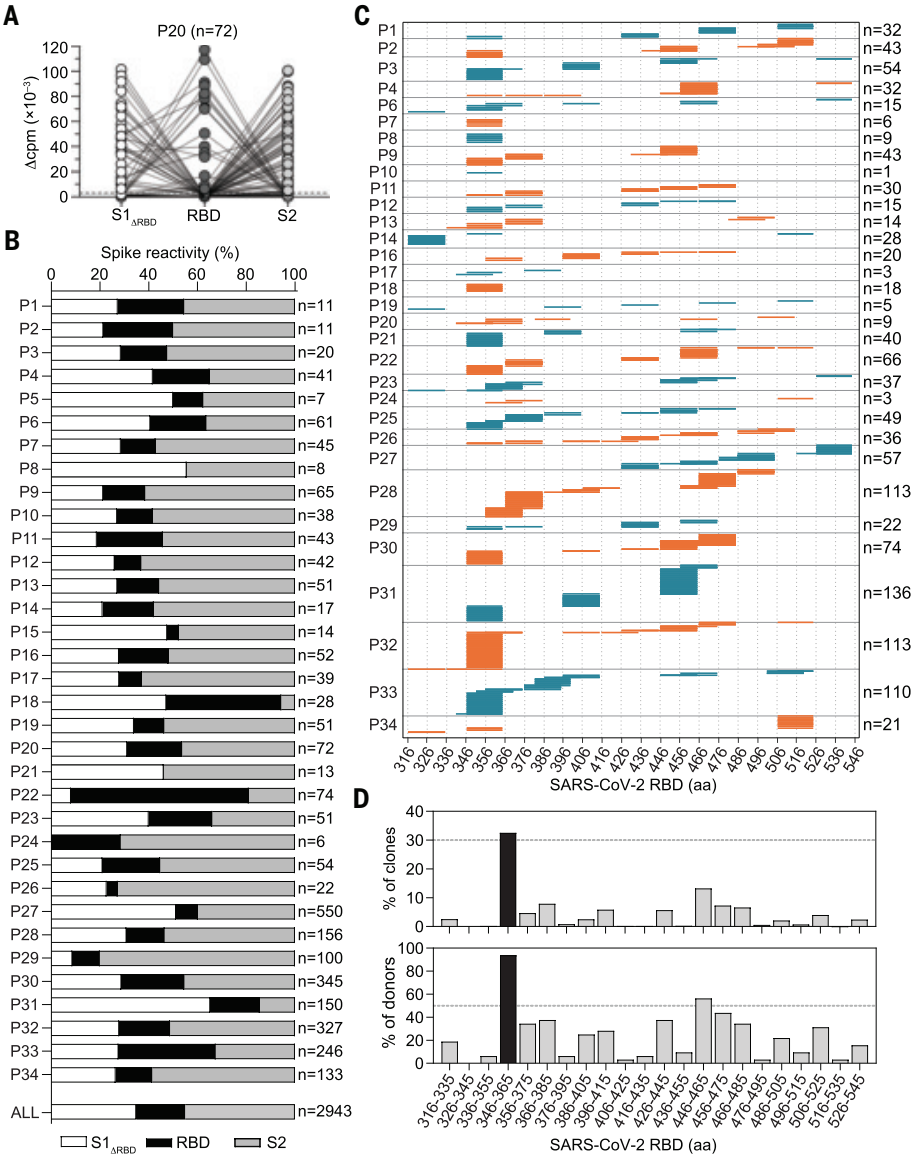
Fig. 1. Robust T cell response to SARS-CoV-2 S and N proteins in CD4$^+$ T_M cell subsets. Total CD4$^+$ memory T cells from seven COVID-19–recovered patients and six unexposed (pre–COVID-19) healthy donors (HD) or CD4$^+$ T_{CM}, T_{EM}, and cT_{FH} cells from seven COVID-19–recovered patients were labeled with CFSE and cultured with autologous monocytes in the presence or absence of recombinant SARS-CoV-2 S or N protein. (**A**) CFSE profiles on day 7 and percentage of CFSE^{lo} proliferating T_{CM}, T_{EM}, and cT_{FH} cells in a representative recovered patient. Negative controls of T cells cultured with monocytes in the absence of antigen are shown as red lines. (**B**) Individual values and median and quartile values of the percentage of CFSE^{lo}CD25$^+$ICOS$^+$ cells in total CD4$^+$ T_M cells and CD4$^+$ T_{CM}, T_{EM}, and cT_{FH} cell subsets in recovered patients and healthy donors. Also shown are IFN-γ concentrations in culture supernatants of T_M cell subsets from recovered patients at day 7 post stimulation with SARS-CoV-2 S or N protein. IFN-γ concentrations were below the detection limit in HD and in negative control cultures. ****$P < 0.0001$, ***$P < 0.001$, and **$P < 0.01$ as determined by two-tailed unpaired *t* test (total CD4$^+$ T_M and IFN-γ) or by two-tailed paired *t* test (CD4$^+$ T_{CM}, T_{EM}, and cT_{FH} cells). (**C**) Pairwise comparison of TCR Vβ clonotype frequency distribution in samples of T cells isolated from S protein–stimulated T_{CM}, T_{EM}, or cT_{FH} cell subsets (initial input, 5×10^5 cells per subset) from P33. Frequencies are shown as a percentage of productive templates. The total number of clonotypes is indicated in the *x*- and *y*-axes. Values in the upper right corner represent the number of clonotypes shared between the two samples. The Venn diagrams show the number of unique and shared clonotypes between the T_{CM}, T_{EM}, and cT_{FH} cell subsets. (**D**) Bar histograms showing the Chao–Jaccard similarity index between pairs of TCR Vβ repertoires in three donors.

PBMCs from a first cohort of 14 patients who had recovered from mild to severe COVID-19 (table S1) were used to isolate total CD4$^+$ T_M cells or T central memory (T_{CM}), T effector memory (T_{EM}), and circulating T follicular helper (cT_{FH}) cells (fig. S1A). The cells were labeled with carboxyfluorescein diacetate succinimidyl ester (CFSE) and stimulated with autologous monocytes in the presence of recombinant SARS-CoV-2 spike (S) protein or nucleoprotein (N). In all individuals, we observed a strong response to both antigens in terms of proliferation and interferon-γ (IFN-γ) production (Fig. 1, A and B, and fig. S1, B and C). Proliferating cells were detected at different levels in T_{CM}, T_{EM}, and cT_{FH} cells, consistent with a recent report (*14*), and over a 1-year period (fig. S1D). By contrast, the CD4$^+$ T_M cell response to SARS-CoV-2 proteins in unexposed individuals was low or undetectable (Fig. 1B and fig. S1C), consistent with the presence of a few cross-reactive T cells primed by endemic coronaviruses (*4*, *5*, *9*).

The clonal composition of SARS-CoV-2–reactive T cells and the relationship between different memory subsets was studied in three individuals (P28, P31, and P33) by T cell receptor (TCR) Vβ sequencing. The T_{CM}, T_{EM}, and cT_{FH} cell lines comprised, on average, 908, 480, and 697 S-reactive clonotypes and 1452, 623, and 908 N-reactive clonotypes, respectively (Fig. 1C and fig. S2). Unexpectedly, several of the most expanded clonotypes were shared between two subsets, and even among all three subsets (Fig. 1, C and D), indicating a polyfunctional response consistent with previous studies on intraclonal diversification of antigen-primed CD4$^+$ T cells (*15*, *16*).

In view of the interest in the design of a subunit vaccine, we analyzed in depth the CD4$^+$ T cell response to the S protein, in particular to the receptor-binding domain (RBD), which is the main target of neutralizing antibodies (*17*, *18*). CD4$^+$ T cells from a larger cohort of 34 COVID-19-recovered individuals (table S1) were stimulated with S protein–pulsed monocytes, and proliferating T cells were cloned by limiting dilution. We obtained 2943 T cell clones and mapped their specificity using three pools of peptides spanning $S1_{\Delta RBD}$, RBD, and S2 (Fig. 2, A and B). RBD-specific T cell clones were found in 32 out of 34 donors, accounting for, on average, 20% of the response to the S protein (Fig. 2B). Using a matrix-based approach, we mapped the epitope specificity of 1254 RBD-reactive CD4$^+$ T cell clones (Fig. 2C) and found that, in each individual, the clones recognized multiple sites that collectively spanned almost all of the RBD sequence. However, certain regions emerged as immunodominant, such as those spanning residues S346–S385 and S446–S485. A 20–amino acid region (S346–S365) was recognized by 94% of the individuals (30 out of

[1]Institute for Research in Biomedicine, Università della Svizzera italiana, 6500 Bellinzona, Switzerland. [2]Laboratory of Immunogenetics, Department of Transfusion Medicine and Immuno-Hematology, Fondazione I.R.C.C.S. Policlinico S. Matteo, 27100 Pavia, Italy. [3]Clinical Trial Unit, Ente Ospedaliero Cantonale, 6500 Bellinzona, Switzerland. [4]Clinic of Internal Medicine and Infectious Diseases, Clinica Luganese Moncucco, 6900 Lugano, Switzerland. [5]Faculty of Biomedical Sciences, Università della Svizzera italiana, 6900 Lugano, Switzerland. [6]Department of Internal Medicine, Ente Ospedaliero Cantonale, 6500 Bellinzona, Switzerland. [7]Prince of Wales Hospital Clinical School, University of New South Wales, Sydney, New South Wales 2052, Australia. [8]Division of Clinical Pharmacology and Toxicology, Institute of Pharmacological Sciences of Southern Switzerland, Ente Ospedaliero Cantonale, 6900 Lugano, Switzerland. [9]Department of Clinical Pharmacology and Toxicology, University Hospital Zurich, 8091 Zurich, Switzerland. [10]National Institute of Molecular Genetics, 20122 Milano, Italy. [11]Institute of Microbiology, ETH Zürich, 8093 Zurich, Switzerland.
***Corresponding author: federica.sallusto@irb.usi.ch (F.S.); antonino.cassotta@irb.usi.ch (A.Ca.)**
†These authors contributed equally to this work.
‡These authors contributed equally to this work.



E

Genus	Subgenus	Clade	Strain	346	347	348	349	350	351	352	353	354	355	356	357	358	359	360	361	362	363	364	365
β-CoV	Sarbecovirus	2	SARS-CoV-2 Wuhan-Hu-1/2019	R	F	A	S	V	Y	A	W	N	R	K	R	I	S	N	C	V	A	D	Y
β-CoV	Sarbecovirus	2	SARS-CoV-2 B.1.1.7	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
β-CoV	Sarbecovirus	2	SARS-CoV-2 B.1.351	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
β-CoV	Sarbecovirus	2	SARS-CoV-2 P.1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
β-CoV	Sarbecovirus	2	Bat-CoV-RaTG13	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
β-CoV	Sarbecovirus	2	GD Pangolin	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
β-CoV	Sarbecovirus	2	GX Pangolin	K	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
β-CoV	Sarbecovirus	2	Bat-SL-CoVs ZC45; ZXC21	.	.	P	.	.	.	.	.	E	.	T	K	.	.	D	.	I	.	.	.
β-CoV	Sarbecovirus	3	SARS-CoV	K	.	P	.	.	.	.	.	E	.	.	K	.	.	.	.	.	.	.	.
β-CoV	Sarbecovirus	3	Civet SARS-CoV SZ3	K	.	P	.	.	.	.	.	E	.	.	.	.	.	.	.	.	.	.	.
β-CoV	Sarbecovirus	3	Raccoon dog SARS-CoV SZ13	K	.	P	.	.	.	.	.	E	.	.	.	.	.	.	.	.	.	.	.
β-CoV	Sarbecovirus	3	Bat CoV WIV1; WIV16	T	.	P	.	.	.	.	.	E	.	.	.	.	.	.	.	.	.	.	.
β-CoV	Sarbecovirus	1	Bat CoV BtKY72	N	.	P	.	.	.	.	.	E	.	L	.	.	.	D	.	.	.	.	.
β-CoV	Sarbecovirus	1	Bat CoV BGR2008	S	.	P	.	.	.	.	.	E	.	M	.	.	T	.	.	.	.	.	.
β-CoV	Embecovirus		OC43	S	V	P	.	P	L	N	.	E	.	.	T	F	.	.	.	N	F	N	M
β-CoV	Embecovirus		HKU1 (isolate N5)	S	V	P	.	P	L	N	.	E	.	R	I	F	.	.	.	N	F	N	L
α-CoV	Setracovirus		NL63	D	I	N	F	T	A	T	-	-	-	A	S	-	-	-	-	-	-	-	-
α-CoV	Duvinacovirus		229E	F	I	V	L	H	V	K	F	E	H	Q	.	-	-	-	-	-	-	-	-

Fig. 2. CD4⁺ T cell clones target multiple sites on the S protein. (**A** and **B**) CD4⁺ T cell clones (n = 2943) were isolated from S-reactive T cell cultures from 34 COVID-19-recovered individuals, and their specificity was mapped by stimulation with autologous B cells and three pools of 15-mer peptides overlapping of 10 amino acids spanning the S1–S325 and S536–S685 sequences ($S1_{\Delta RBD}$, 91 peptides), the S316–S545 sequence (RBD, 44 peptides), and the S676–S1273 sequence (S2, 118 peptides), using as readout ^{3}H-thymidine incorporation. (A) Characterization of representative T cell clones (n = 72) from P20. Proliferation was assessed on day 3 after a 16-hour pulse with ^{3}H-thymidine and is expressed as counts per minute after subtraction of the unstimulated control value (Δcpm). (B) Percentage of T cell clones specific for $S1_{\Delta RBD}$ (white), RBD (black), and S2 (gray) in the 34 individuals tested. The number of clones tested is indicated on the right. The distribution of all S protein–reactive T cell clones isolated from all 34 individuals (ALL, n = 2943) is also indicated. (**C**) RBD-specific T cell clones (n = 1254) isolated from 32 individuals were further characterized for their epitope specificity using 15-mer peptides overlapping of 10 amino acids spanning the S316–S545 RBD sequence. The 20-mer specificity of each clone is represented by a horizontal line, and the total number of clones mapped for each individual is indicated on the right. (**D**) Percentage of clones specific and percentage of individuals carrying T cells specific for different 20-mer segments of the RBD. Data for the immunodominant region S346–S365 is shown in black. (**E**) Sequence alignments of the SARS-CoV-2–immunodominant region S346–S365 with homologous sequences in different sarbecoviruses, human and animal SARS-related coronaviruses, and alpha and beta coronaviruses. Dots indicate amino acid residues identical to SARS-CoV-2 reference strain; dashes indicate deletions.

32) and by 33% of the clones (408 out of 1254) (Fig. 2D). This region is highly conserved among human sarbecoviruses, including the recently emerged variants of concern and zoonotic sarbecoviruses (Fig. 2E) (*19*). RBD- and S346–S365-specific T cell clones were found in different memory subsets of COVID-19-recovered individuals and were also isolated from individuals after SARS-CoV-2 mRNA vaccination (fig. S3). Thus, RBD is highly immunogenic in vivo and contains a large number of naturally processed T cell epitopes, including a conserved immunodominant region.

To study the CD4⁺ T cell response to the immunodominant S346–S365 region, we

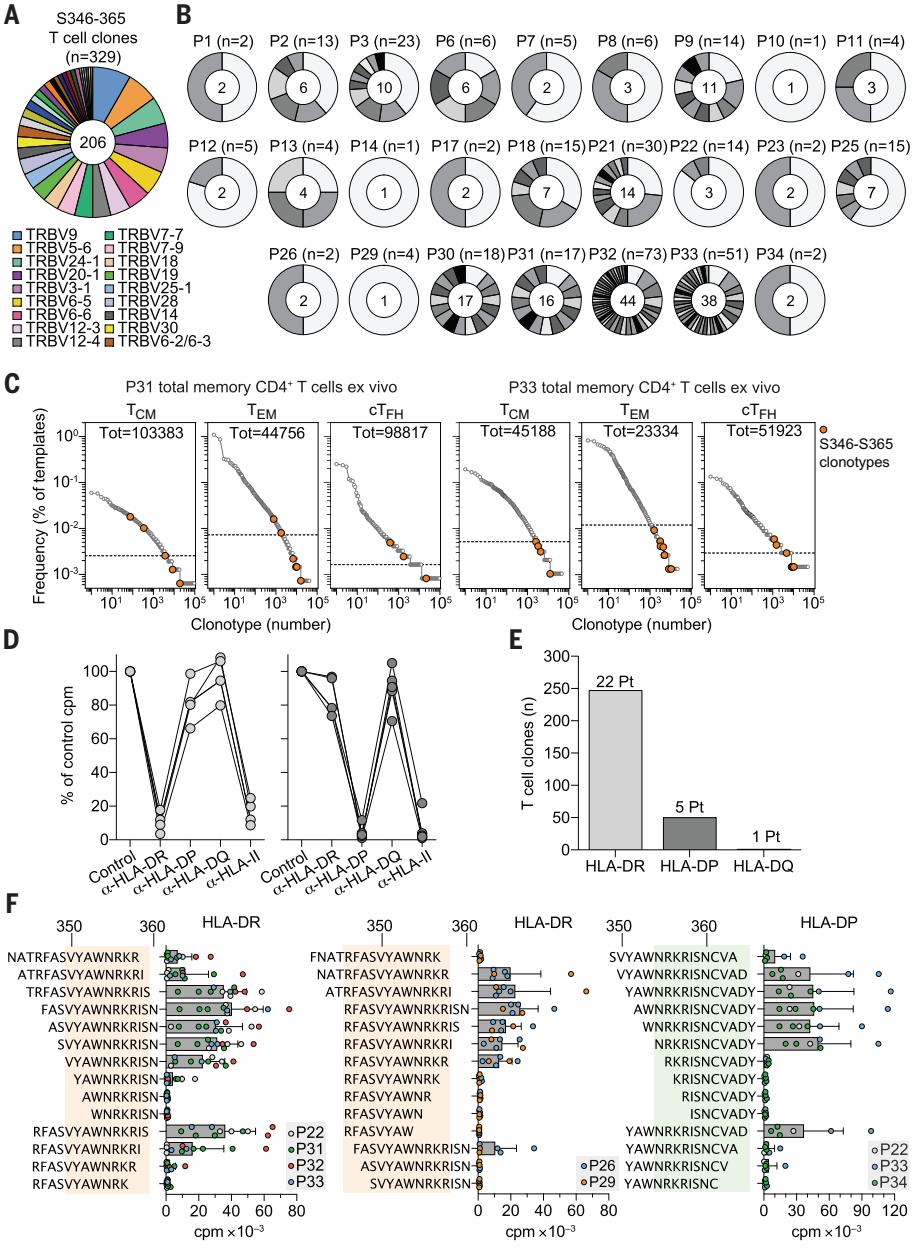


Fig. 3. The immunodominant S346–S365 RBD region contains nested epitopes targeted by a diverse repertoire of T cells restricted by HLA-DR and HLA-DP. (**A** and **B**) Rearranged TCR Vβ sequences of S346–S365-reactive CD4$^+$ T cell clones (n = 329) isolated from 25 COVID-19–recovered individuals as determined by reverse transcription polymerase chain reaction and Sanger sequencing. (A) TCR Vβ gene usage of the 206 unique clonotypes. Slices in the chart represent different Vβ genes, and their size is proportional to the number of clonotypes using that particular gene. Color-coded legend is reported for the top 18 Vβ genes (used by at least five different TCR Vβ clonotypes). (B) Number of S346–S365-reactive T cell clones and clonotypes identified in the 25 individuals. Slices in the charts represent different TCR Vβ clonotypes, and their size is proportional to the number of sister clones bearing the same sequence. The number of clones is reported on the top and the number of clonotypes at the center of the pie chart. (**C**) Frequency distribution of TCR Vβ clonotypes from CD4$^+$ T_{CM}, T_{EM}, and cT_{FH} cell subsets sequenced directly after ex vivo isolation from P31 and P33. Colored circles mark the TCR Vβ clonotypes found among the S346–S365-specific T cell clones isolated from the same individual. Dotted lines in the graphs indicate the frequency threshold of the top 5% expanded clonotypes. (**D**) HLA class II isotype restriction of S346–S365-specific T cell clones (n = 10) isolated from P33 as determined by stimulation with peptide-pulsed autologous APCs in the absence (control) or presence of blocking antibodies to HLA-DR, HLA-DP, HLA-DQ, or pan–HLA class II. Proliferation was assessed on day 3 after a 16-hour pulse with ^{3}H-thymidine. Data are expressed as a percentage of control counts per minute (cpm). (**E**) HLA class II isotype usage by S346–S365-reactive CD4$^+$ T cell clones (n = 298) from 24 individuals as determined by >80% inhibition of proliferation. (**F**) Identification of the minimal peptide recognized by HLA-DR– or HLA-DP–restricted S346–S365-reactive CD4$^+$ T cell clones (n = 23) isolated from seven individuals, as determined by stimulation with autologous APCs pulsed with a panel of truncated peptides. Proliferation was assessed on day 3 and is expressed as counts per minute (cpm). Bars indicate mean ± SD; circles indicate individual clones. The minimal amino acid sequences recognized by T cell clones are highlighted with colored shading.

sequenced TCR Vβ chains of 329 specific T cell clones. The 206 clonotypes identified used a broad spectrum of TCR Vβ genes and, even in the same individual, carried different CDR3 sequences (Fig. 3, A and B, and table S2). In P31 and P33, certain S346–S365 clonotypes were detected among the top 5% expanded T_M cells ex vivo (Fig. 3C). Using blocking antibodies, we determined that most of the T cell clones analyzed (n = 247 from 22 individuals) were HLA-DR restricted, whereas the remaining clones (n = 50 from five individuals) were HLA-DP restricted and one was HLA-DQ restricted (Fig. 3, D and E). Using truncated peptides and T cell clones from individuals with different HLA types (table S3), we defined two HLA-DR–restricted epitopes (VYAWNRKRIS and RFASVYAWNRKR) and one HLA-DP–restricted epitope (NRKRISNCVAD) (Fig. 3F). Thus, the S346–S365 region comprises at least three nested epitopes recognized in association with different allelic forms of HLA-DR or HLA-DP by T cell clones that use a large set of TCR Vβ genes and CDR3 of different sequence and length.

To address the extent of T cell cross-reactivity between different S proteins, SARS-CoV-2 S protein–specific T cell lines from P28 and P33 were relabeled with CFSE and stimulated with S proteins from endemic human coronaviruses. In these secondary cultures, a robust proliferation was observed in response to SARS-CoV and HKU1 (Fig. 4A). Unexpectedly, a sizeable fraction of clonotypes in SARS-CoV-2 primary cultures (ranging from 7 to 25%) were found in SARS-CoV and/or HKU1 secondary cultures, consistent with a substantial degree of T cell cross-reactivity (fig. S4). To corroborate this finding, we isolated from secondary cultures several T cell clones that proliferated in response to two or even three different naturally processed S proteins (Fig. 4B and table S4).

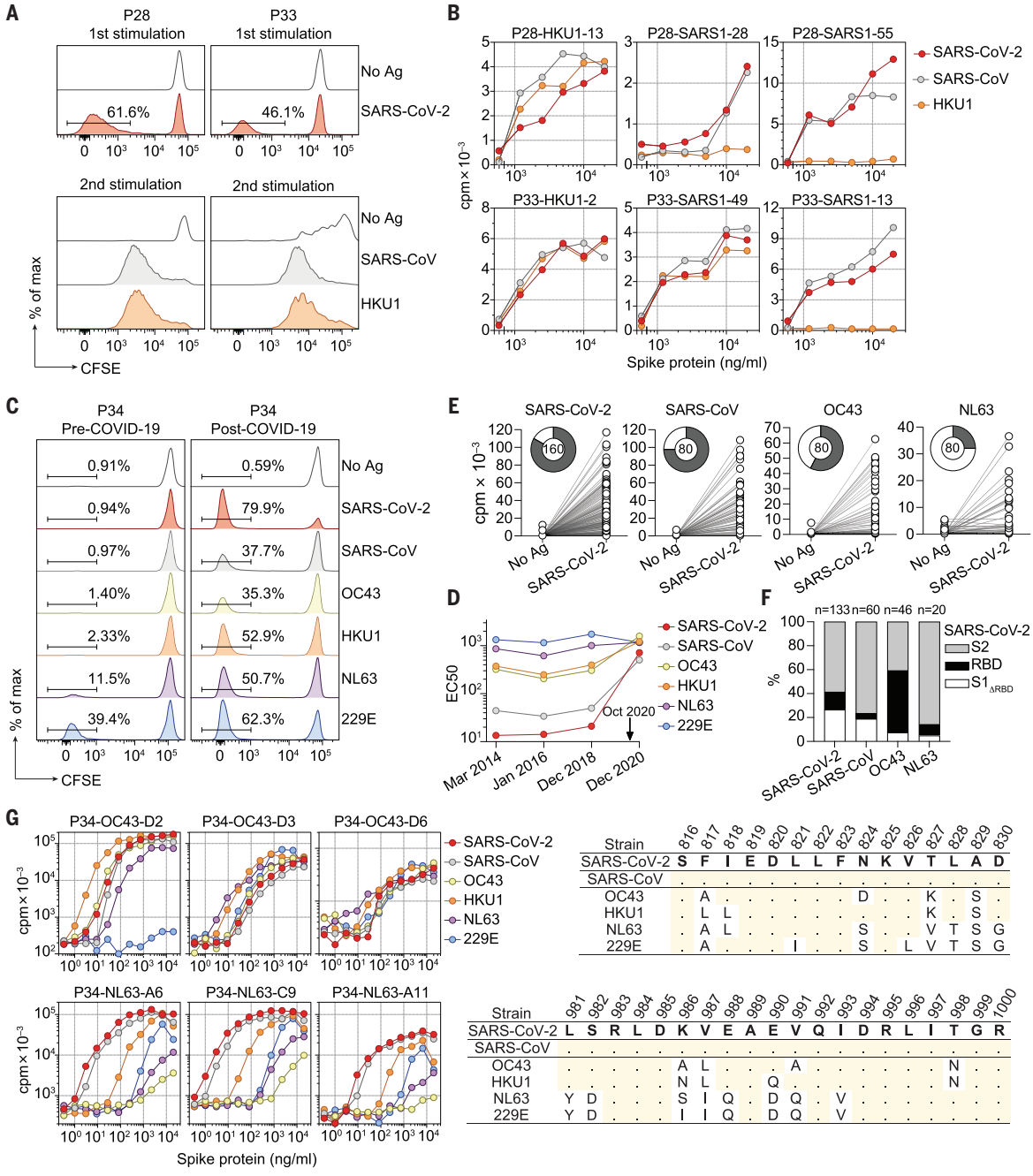


Fig. 4. Identification of coronavirus S protein cross-reactive T cell clones. (**A** and **B**) CFSE-labeled CD4$^+$ T_M cells from P28 and P33 were stimulated with autologous monocytes pulsed with recombinant SARS-CoV-2 S protein. CFSE^{lo} cells were expanded with interleukin-2 for 10 days, relabeled with CFSE, and restimulated with S proteins from human beta (SARS-CoV, HKU1, and OC43) or alpha (NL63 and 229E) coronaviruses. T cell clones from proliferating secondary cultures were isolated and tested for cross-reactivity against different S proteins. (A) CFSE profiles from primary and secondary stimulation in the absence or presence of the indicated antigens. (B) Proliferative response of representative T cell clones, isolated from secondary cultures, to autologous APCs pulsed with titrated doses of recombinant S proteins from SARS-CoV-2, SARS-CoV, or HKU1. Proliferation was assessed on day 3 after a 16-hour pulse with ^{3}H-thymidine and is expressed as counts per minute (cpm). (**C** to **G**) Multiple blood samples were obtained from donor P34 several years before and 1.5 months after COVID-19 (disease onset Oct 2020) and characterized regarding T_M cells and serum antibody levels. (C) T cell proliferation measured by CFSE dilution in pre–COVID-19 (Dec 2018) and post–COVID-19 (Dec 2020) samples in response to autologous monocytes pulsed with different S proteins. (D) Time course of serum immunoglobulin G antibodies against different coronavirus S proteins as determined by enzyme-linked immunosorbent assay [half-maximal serum dilution (EC_{50}) values]. These data demonstrate that, together with a strong induction of serum antibodies to SARS-CoV-2, antibody titers against HKU1 and OC43 also increased in the post–COVID-19 sample. (E) Proliferative response (day 3 cpm) to a pool of peptides spanning SARS-CoV-2 S protein of T cell clones obtained from post–COVID-19 CFSE^{lo} cultures stimulated by SARS-CoV-2, SARS-CoV, OC43, or NL63. Pie charts show the total number of clones tested and the fraction of responsive clones is shown in gray. (F) Specificity of SARS-CoV-2 S peptide pool-reactive T cell clones isolated from each culture (E) was further mapped by stimulation

with pools of peptides spanning the $S1_{\Delta RBD}$, RBD, and S2 regions of the SARS-CoV-2 S protein. Histograms show the percentage of clones specific for each region. The total number of clones tested is indicated at the top. (G) Characterization of representative cross-reactive T cell clones isolated from P34 post–COVID-19 sample. Left panels report the proliferative response (day 3 cpm) of T cell clones stimulated with titrated doses of recombinant S proteins in the presence of autologous monocytes. The peptides recognized are indicated on the right panels. Shown are sequence alignments of the recognized SARS-CoV-2 epitopes (S816–S830 and S981–S1000) with homologous sequences of endemic alpha and beta coronaviruses. Dots indicate amino acid residues identical to the SARS-CoV-2 reference strain.

Cross-reactive T cells may derive from preexisting memory T cells or from the priming of naïve T cells. We therefore analyzed a COVID-19–recovered individual from whom we had previously cryopreserved PBMCs. A robust $CD4^+$ T_M cell proliferation in the pre–COVID-19 sample was detected against NL63 and 229E S proteins, whereas the response to HKU1 and OC43 was limited and the response to SARS-CoV and SARS-CoV-2 undetectable (Fig. 4C). Conversely, in the post–COVID-19 sample, strong T cell proliferation was observed not only in response to SARS-CoV-2, but also in response to all other alpha and beta coronavirus S proteins (Fig. 4, C and D), and shared clonotypes were detected between SARS-CoV-2 and endemic coronavirus S protein–stimulated cultures (fig. S5A). Furthermore, T cell clones isolated from cultures stimulated by SARS-CoV, OC43, or NL63 proliferated in response to the SARS-CoV-2 S peptide pool, and their specificity was mapped primarily to the S2 region (Fig. 4, E and F), consistent with its high degree of sequence conservation (*20*–*22*). T cell clones that fully cross-reacted with all S proteins were mapped to the highly conserved fusion peptide (Fig. 4G).

To determine whether S-reactive T cells in the post–COVID-19 sample could be detected in pre-pandemic samples, we performed clonotypic analysis of total T_M cells on the post–COVID-19 sample and on samples collected in 2014 and 2017. Most of the SARS-CoV-2–specific clonotypes identified above were found only in the post–COVID-19 sample, consistent with priming of naïve T cells (fig. S5B). By contrast, clonotypes specific to endemic coronaviruses were found at a comparable number at all time points. Some T cell clonotypes against the highly conserved fusion peptide could be tracked back to the 2014 sample and were found to be expanded in the post–COVID-19 sample (fig. S5C). These findings demonstrate that preexisting cross-reactive T_M cells are recalled and expanded upon SARS-CoV-2 infection.

The robust $CD4^+$ T cell response to the RBD and the identification of the S346–S365 immunodominant region conserved in the emerging SARS-CoV-2 variants of concern provide the rationale for the development of a subunit vaccine based on the RBD because it is the target of most neutralizing antibodies (*17*, *18*). These findings were not anticipated in previous studies based on bioinformatics predictions (*2*, *3*) and short-term peptide stimulation of PBMCs, highlighting the value of combining T cell stimulation with protein antigens with cloning and TCR sequencing for the analysis of antigen-specific T cell repertoires.

The immunodominance of RBD S346–S365 at the individual level and at the population level may be due to the presence of three nested T cell epitopes presented by HLA-DR and HLA-DP and to the relative abundance of naturally processed peptides, as recently reported in an immunopeptidomics study (*23*). The S346–S365 region is also a contact site for the broadly reactive neutralizing antibody S309 (*24*), providing a good example of convergence of B and T cells around a conserved epitope.

Our study also provides evidence for the recall of preexisting cross-reactive T_M cells upon SARS-CoV-2 infection. However, this phenomenon, reminiscent of the "original antigenic sin" (*25*), does not prevent a robust and persistent primary response to new epitopes of SARS-CoV-2 that is characterized by extensive intraclonal diversification into T_{EM}, cT_{FH}, and T_{CM} cells, which represent inflammatory, helper, and long-lived T_M cells, respectively (*26*, *27*). The availability of a large number of cross-reactive T cell clones is not only instrumental for defining target sites in relevant pathogens but also for understanding whether cross-reactivity is due to epitope structural similarities or to TCR-binding degeneracy (*11*, *28*).

The possibility of leveraging a robust, cross-reactive T helper cell function against conserved sites will be instrumental in driving neutralizing antibody responses to adaptive vaccines that incorporate escape mutations found in emerging SARS-CoV-2 variants.

REFERENCES AND NOTES

1. R. Vita *et al.*, *Nucleic Acids Res.* **47** (D1), D339–D343 (2019).
2. A. Grifoni *et al.*, *Cell Host Microbe* **27**, 671–680.e2 (2020).
3. A. Nelde *et al.*, *Nat. Immunol.* **22**, 74–85 (2021).
4. A. Grifoni *et al.*, *Cell* **181**, 1489–1501.e15 (2020).
5. J. Braun *et al.*, *Nature* **587**, 270–274 (2020).
6. A. Tarke *et al.*, *Cell Rep. Med.* **2**, 100204 (2021).
7. J. Mateus *et al.*, *Science* **370**, 89–94 (2020).
8. N. Le Bert *et al.*, *Nature* **584**, 457–462 (2020).
9. P. Bacher *et al.*, *Immunity* **53**, 1258–1271.e5 (2020).
10. A. Kim, S. Sadegh-Nasseri, *Curr. Opin. Immunol.* **34**, 9–15 (2015).
11. A. K. Sewell, *Nat. Rev. Immunol.* **12**, 669–677 (2012).
12. J. W. Yewdell, *Immunity* **25**, 533–543 (2006).
13. A. Cassotta *et al.*, *J. Exp. Med.* **217**, e20200206 (2020).
14. L. B. Rodda *et al.*, *Cell* **184**, 169–183.e17 (2021).
15. N. J. Tubo *et al.*, *Cell* **153**, 785–796 (2013).
16. S. Becattini *et al.*, *Science* **347**, 400–406 (2015).
17. L. Piccoli *et al.*, *Cell* **183**, 1024–1042.e21 (2020).
18. A. J. Greaney *et al.*, *Cell Host Microbe* **29**, 44–57.e9 (2021).
19. R. Lu *et al.*, *Lancet* **395**, 565–574 (2020).
20. J. A. Jaimes, N. M. André, J. S. Chappie, J. K. Millet, G. R. Whittaker, *J. Mol. Biol.* **432**, 3309–3325 (2020).
21. M. A. Tortorici, D. Veesler, *Adv. Virus Res.* **105**, 93–116 (2019).
22. A. C. Walls *et al.*, *Nature* **531**, 114–117 (2016).
23. M. D. Knierman *et al.*, *Cell Rep.* **33**, 108454 (2020).
24. D. Pinto *et al.*, *Nature* **583**, 290–295 (2020).
25. J. W. Yewdell, J. J. S. Santos, *Cold Spring Harb. Perspect. Med.* **11**, a038786 (2021).
26. S. Crotty, *Immunity* **41**, 529–542 (2014).
27. R. Ahmed, M. J. Bevan, S. L. Reiner, D. T. Fearon, *Nat. Rev. Immunol.* **9**, 662–668 (2009).
28. J. B. Huppa, M. M. Davis, *Adv. Immunol.* **119**, 1–50 (2013).
29. TCR Vβ sequence data for: J. S. Low *et al.*, Clonal analysis of immunodominance and cross-reactivity of the CD4 T cell response to SARS-CoV-2, immuneACCESS Database (2021); https://doi.org/10.21417/JSL2021S.

ACKNOWLEDGMENTS

We thank all participants in the study and the personnel at the hospitals for blood collection; G. Durini for cell isolation; W. Pertoldi, S. Tettamanti, and E. Picciocchi from the Istituti Sociali Chiasso for blood samples from vaccinated donors; and D. Lilleri and members of the Tipizzazione laboratory of the IRCCS San Matteo Hospital Foundation San Matteo Pavia, Italy, for HLA typing. **Funding:** This work was supported in part by the Henry Krenter Foundation, the Cariplo Foundation (grant 2020-1374 CoVIM), the Swiss National Science Foundation (grant 189331), and by EOC research funds. F.S. and the Institute for Research in Biomedicine are supported by the Helmut Horten Foundation. **Author contributions:** A.Ca. and F.S. contributed to study concept and design. J.S.L., D.V., F.M., and J.J. contributed to experimental work. M.P. and L.P. produced recombinant proteins. D.J. performed cell sorting. S.J. provided technical support. M.F. performed bioinformatics analysis. R.C. performed HLA typing. T.T., A.F.P., M.B., C.G., P.F., and A.Ce. contributed clinical samples. A.Ca., A.L., and F.S. analyzed the data and wrote the manuscript. All authors contributed to interpretation of data and critical revision of the manuscript. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** TCR Vβ sequences have been deposited in the ImmuneACCESS database (29). All other data are available in the main text or the supplementary materials.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/372/6548/1336/suppl/DC1
Materials and Methods
Tables S1 to S5
Figs. S1 to S5
References (*30*, *31*)
MDAR Reproducibility Checklist

3 February 2021; accepted 13 May 2021
Published online 18 May 2021
10.1126/science.abg8985

ICE SHEETS

Transition to marine ice cliff instability controlled by ice thickness gradients and velocity

J. N. Bassis[1]*, B. Berg[1,2], A. J. Crawford[3], D. I. Benn[3]

Portions of ice sheets grounded deep beneath sea level can disintegrate if tall ice cliffs at the ice-ocean boundary start to collapse under their own weight. This process, called marine ice cliff instability, could lead to catastrophic retreat of sections of West Antarctica on decadal-to-century time scales. Here we use a model that resolves flow and failure of ice to show that dynamic thinning can slow or stabilize cliff retreat, but when ice thickness increases rapidly upstream from the ice cliff, there is a transition to catastrophic collapse. However, even if vulnerable locations like Thwaites Glacier start to collapse, small resistive forces from sea-ice and calved debris can slow down or arrest retreat, reducing the potential for sustained ice sheet collapse.

The Antarctic and Greenland ice sheets are drained by glaciers and ice shelves that terminate in near vertical ice cliffs submerged in the ocean. The portions of these ice sheets grounded beneath sea level have the potential to catastrophically collapse through a spectrum of instabilities, including the marine ice sheet instability and the marine ice cliff instability (MICI) (*1*–*4*). MICI was only recently proposed and occurs because the height of ice cliffs is limited by the strength of ice (*4*). When glaciers retreat into an overdeepening basin—or ice shelf collapse exposes a tall ice cliff—cliffs become structurally unstable at a threshold cliff height, leading to runaway cliff failure and ice sheet disintegration (*3*, *4*).

Because MICI proceeds through brittle failure, it could lead to rapid ice sheet mass loss, with serious implications for sea level rise in the 21st century and beyond (*3*). Although there is evidence supporting the MICI in the paleo-record (*5*), MICI remains controversial because it has yet to be observed in modern-day glaciers. Moreover, current models of MICI rely on quasi-empirical parameterizations extrapolated from limited observations to simulate retreat (*3*, *6*, *7*). Without understanding the processes that limit rates of collapse, projections of sea level rise remain uncertain.

Here we show that, contrary to the MICI hypothesis, ice cliffs perched just above the maximum cliff height will not always catastrophically collapse, even when grounded on retrograde bed slopes with ice thickness increasing upstream. Instead, we find that catastrophic collapse is triggered when the ice thickness gradient exceeds a critical threshold. To probe ice cliff stability, we use the m-ice model (*8*), which treats ice like a power-law viscous material only until a yield strength is reached (*9*, *10*). Once the yield strength is reached, the ice deforms rapidly and accumulation of plastic strain in failed portions of the ice reduces the strength of ice, resulting in failure localization. Our m-ice simulations neglect transient elastic stresses. The starting

[1]Department of Climate and Space Science and Engineering, University of Michigan, Ann Arbor, MI 48109, USA. [2]Department of Physics, University of Michigan, Ann Arbor, MI, 48109, USA. [3]School of Geography and Sustainable Development, University of St Andrews, St Andrews KY16 9AL, UK.
*Corresponding author. Email: jbassis@umich.edu

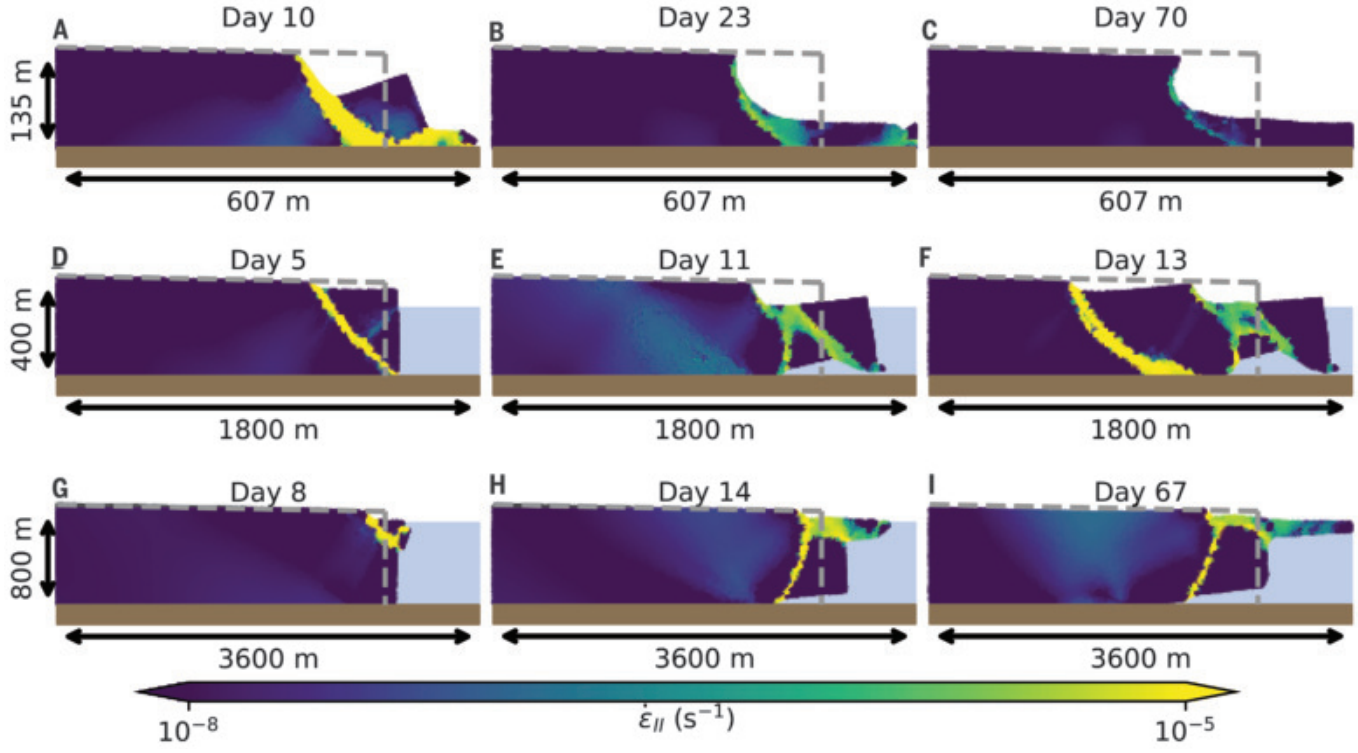


Fig. 1. Snapshots showing the evolution of the ice cliff for different water depths and ice thicknesses. The color bar shows the effective strain rate $\dot{\epsilon}_{II}$. Dashed gray lines show the initial glacier geometry. (**A** to **C**) Initial slump for a dry calving cliff 135 m high. After the slump, the ice cliff stabilizes and advances. (**D** to **F**) A full-thickness berg detaching from an ice cliff 400 m high, grounded in 290 m of water. The calving event exposes thicker ice upstream, triggering runaway failure and complete glacier disintegration. (**G** to **I**) Snapshots from an ice cliff 800 m high, grounded in 690-m water depth. An initial slump triggers a buoyant calving event. Continued erosion of mass from the top of the cliff eventually triggers another buoyant calving event in a cyclical process. See movies S1 to S3.

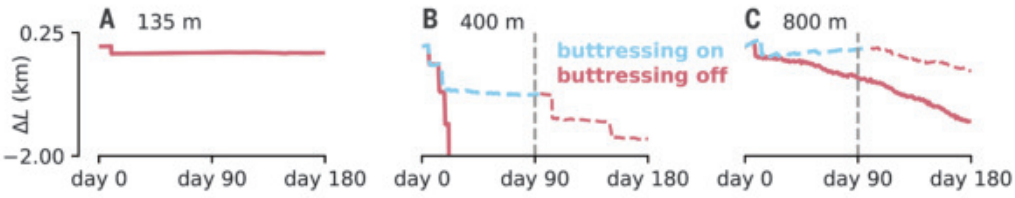


Fig. 2. Change in ice cliff position (Δ*L*) for the three ice thicknesses and water depth combinations shown in Fig. 1. (**A**) Initial retreat followed by advance of the 135-m-thick glacier. (**B** and **C**) The solid red lines show retreat of the cliff for the 400- and 800-m-thick glacier without buttressing. The dashed blue lines show the stabilizing effect of 25 kPa of buttressing. Buttressing was removed after day 90 (shown with a vertical gray line), triggering subsequent retreat (dashed red line).

point of our simulations thus roughly corresponds to the end point of the viscoelastic simulations in (*11*). Our simulations, however, use a representation of the strength of ice that is more appropriate for modern-day Greenland calving glaciers and Antarctic ice shelves (*8*).

In our experiments, we first examined idealized slabs of ice on a flat bed with varying ice cliff thickness (*H*) and water depth (*D*), with ice thickness increasing upstream. For each simulation, the ice thickness or water depth was chosen empirically so that the stress at the ice cliff was initially perched slightly above the failure threshold. This resulted in simulations with an initially 135-m-thick dry ice cliff, a 400-m-thick glacier grounded in 290 m of water, and an 800-m-thick glacier grounded in 690 m of water.

In the dry case (D = 0, H ~135 m; Fig. 1, A to C), failure occurs as a slump. Calved debris accumulates ahead of the ice cliff, but the cliff is stable. After the initial slump, the glacier slowly thins and advances, with cascades of smaller calving events avalanching from the cliff top (movie S1 and Fig. 2A). The slope of the cliff is ~55°, similar to observations of Eqip Sermia, Greenland, which has a ~100-m-tall cliff that terminates in tens of meters of water (*12*). Moreover, additional simulations using a discrete element model to simulate brittle failure reveal similar patterns of failure, with a stable slumped cliff (movies S4 and S5), supporting the simplified continuum representation of failure used here.

We next simulated cliff collapse from a 400-m-thick glacier terminating in ~290 m of water (Fig. 1, D to F, and movie S2), comparable to typical Greenland outlet glaciers. An initial full-thickness fracture results in an iceberg that detaches. The berg is buoyant, drifts away and, unlike the dry case, does not provide a stabilizing compressive stress. The new calving face has a slight slope, resulting in buoyancy forces that trigger a second calving event. This exposes thicker ice upstream and causes another full-thickness fracture, leading to a cycle of catastrophic cliff collapse (Fig. 2B and movie S2). However, the addition of a small ~25-kPa back-stress at the ice cliff, similar to or smaller than the back-stress inferred in iceberg-choked Greenlandic fjords (*13*), slows retreat and prevents complete collapse (Fig. 2B). Retreat continues once the back-stress or "buttressing" is removed. This is consistent with observations of Greenland glaciers that show that the seasonal presence of sea-ice and icebergs clogging fjords stabilizes glacier retreat (*13*–*16*).

Finally, we examined an 800-m-tall cliff terminating in 690 m of water (~25 m height-above-buoyancy (Fig. 1, G to I), comparable to Greenland's largest outlet glaciers and to the current grounding line thickness of Thwaites and Pine Island glaciers. Here, calving initiates with a subaerial slump, which triggers a buoyant calving event (Fig. 1G and movie S3). This is similar to the "footloose" theory of buoyant calving and observations of large Greenland glaciers (*17*–*19*). Thicker ice upstream is exposed as icebergs are quickly evacuated. However, episodic "serac" failure from the ice cliff results in a sequence of repeated events where wasting from the top of the ice cliff exposes a buoyant foot that episodically detaches (Fig. 1I). A small 25-kPa back-stress again stabilizes the ice cliff, this time resulting in cliff advance (Fig. 2C).

Our model predicts a distinct pattern of uplift near the ice cliff associated with progressive "serac" failure that precedes calving. This pattern of uplift is markedly similar to observed patterns of uplift observed near the cliffs of thick Greenland glaciers (Fig. 3), although additional processes, like submarine melt and formation of a super-buoyant tongue, may also play a role in Greenland calving cliff evolution. Nonetheless, the agreement between observations and simulations hint that cliff failure may already be underway in sections of Greenland.

To determine if stable cliff positions are possible in the absence of buttressing, we varied upstream velocity and bed slope to assess stability of the 800-m cliff. We ran simulations for at least 1 year or until the glacier completely collapsed to determine mean rates

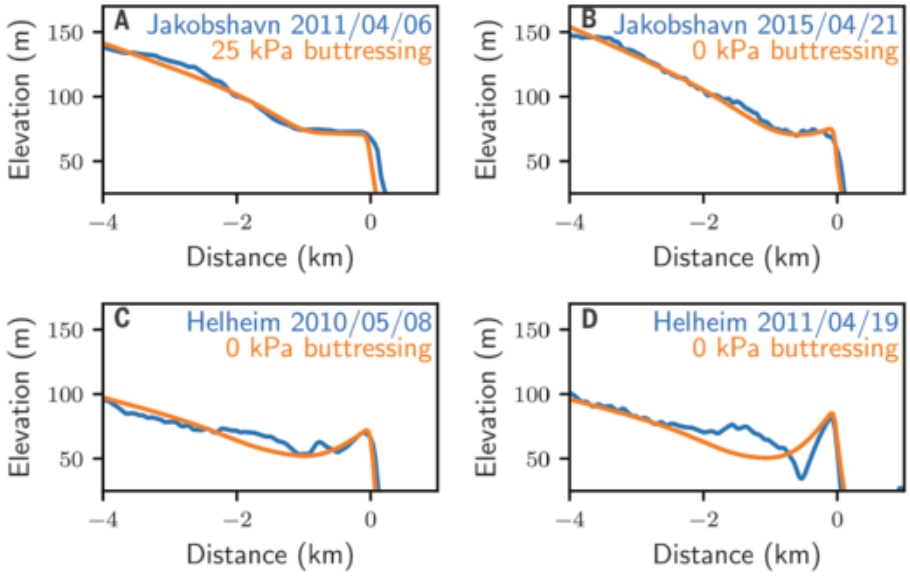


Fig. 3. Comparison between model snapshots (orange) and Operation IceBridge altimetry profiles (blue) over Jakobshavn Isbræ (Sermeq Kujalleq) and Helheim Glacier, Greenland. (**A** and **B**) Comparison of snapshots from simulations computed using a retrograde bed slope of 0.01 and upstream velocity of 6 km/annum. Panel (A) shows a snapshot from a simulation with 25 kPa of buttressing at the ice cliff, resulting in a flatter profile near the ice cliff. (**C** and **D**) Snapshots computed using a flat bed and 4 km/annum upstream velocity. Models and observations in (B) to (D) show substantial uplift at the terminus that is especially pronounced for Helheim Glacier.

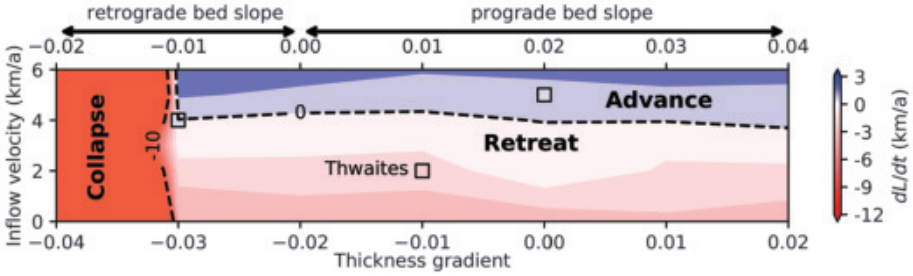


Fig. 4. Average rate of terminus advance over 1 year for an initial 800-m-thick glacier terminating in 690 m of water for a range of thickness gradients and upstream velocities. The bed slope is shown on the top axis. The filled squares are color coded to show the rate of terminus advance computed for half the initial surface slope. The approximate current thickness gradient and inflow velocity of the Thwaites Glacier grounding line are also marked, indicating retreat at a few kilometers per year, but not catastrophic collapse if the grounding line was exposed.

of cliff advance (Fig. 4). Simulations were initialized with identical surface slopes. To examine the role of bed slope and ice thickness gradients in controlling cliff stability, we also performed a few simulations with half the initial surface slope but equivalent ice thickness gradient (Fig. 4, filled squares).

For modest ice thickness gradients, we see patterns of retreat and advance largely controlled by the upstream velocity rather than the bed slope. Experiments performed with identical initial thickness gradients, but different bed slopes, result in comparable rates of terminus advance and retreat. For modest ice thickness gradients, larger upstream velocities result in advancing cliffs, whereas smaller upstream velocities result in retreating cliffs. Glacier advance and retreat is separated by a transition region, where rates of terminus advance are small (<100 m/annum) and quasi-stable over our 1- to 2-year simulation period (fig. S1).

As the ice thickness gradient becomes increasingly negative—ice thickness increases upstream faster—there is an abrupt transition at a critical thickness gradient to catastrophic collapse for all inflow velocities (Fig. 4). This "marine ice cliff collapse" regime is a consequence of retreat (by calving) exposing thicker ice upstream much faster than dynamic thinning can reduce the thickness of the upstream ice exposed. Glacier advance on a steeply sloping bed also increases cliff height, leading to runaway cliff failure and retreat rates exceeding tens of kilometers per year. This mechanism suggests that cliff stability is a strong function of dynamic thinning and hence, ice temperature. We confirmed this hypothesis with an additional set of simulations showing that warmer ice stabilizes retreat and results in a larger-magnitude (and flux dependent) critical thickness gradient (figs. S2 and S3).

Crucially, our results highlight the key role that dynamic thinning of the ice plays in controlling cliff failure. Resistance to collapse is controlled by a balance between upstream flux, dynamic thinning, and advection of thicker ice from upstream (supplementary text). This results in two regimes of cliff collapse. In the first regime, dynamic thinning keeps pace with calving, preventing runaway collapse by restricting growth of the cliff height during retreat. This regime is characterized by uplift near the cliff that precedes calving and strongly resembles observed patterns of uplift in thick Greenland outlet glaciers (Fig. 3). The second regime, marine ice cliff collapse, occurs when ice thickness increases rapidly upstream. However, even if a glacier enters into a regime where marine ice cliff collapse is imminent, a relatively small back-stress on the ice cliff of a few tens of kilopascals can slow or even stabilize retreat, making sustained ice sheet collapse less likely. This back-stress can be provided by the mixture of icebergs, sea-ice, and land-fast ice that abuts pinning points or ice margins. We also examined the possibility that gradual removal or weakening of an ice shelf could stabilize and prevent runaway retreat (*11*). Initializing simulations with 25 to 50 kPa of buttressing and then ramping the buttressing down over 1 to 50 days (fig. S4) shows that, consistent with our previous experiments, retreat and collapse can be postponed by a modest back-stress. The ice-cliff, however, remains precarious and retreat eventually accelerates, leading to collapse. These results support our previous interpretation but further emphasize that glacier geometry plays a dominant role in controlling rates of retreat associated with the marine ice cliff instability.

Thwaites Glacier, located in the Amundsen Sea Embayment of West Antarctica, is hypothesized to be one of the glaciers most vulnerable to cliff collapse (*3*). Our results suggest that disintegration or weakening of the floating ice shelf that currently buttresses Thwaites Glacier will expose a grounding line thickness large enough to initiate cliff retreat. At present grounding line conditions, Thwaites is unlikely to initially collapse. However, exposing the grounding line could trigger glacier retreat of a few kilometers per year (Fig. 3), comparable to the current retreat rate of large Greenland outlet glaciers like Jakobshavn Isbræ (Greenlandic: Sermeq Kujalleq) (*20*). Thwaites, however, is more than an order-of-magnitude wider than Jakobshavn and, even if Thwaites does not transition to catastrophic cliff collapse, initiating retreat would result in a substantial increase in the contribution of Thwaites Glacier to sea level rise.

REFERENCES AND NOTES

1. J. Weertman, *J. Glaciol.* **13**, 3–11 (1974).
2. J. H. Mercer, *Nature* **271**, 321–325 (1978).
3. R. M. DeConto, D. Pollard, *Nature* **531**, 591–597 (2016).
4. J. N. Bassis, C. C. Walker, *Eng. Sci.* **468**, 913–931 (2012).
5. M. G. Wise, J. A. Dowdeswell, M. Jakobsson, R. D. Larter, *Nature* **550**, 506–510 (2017).
6. T. L. Edwards *et al*., *Nature* **566**, 58–64 (2019).
7. T. Schlemm, A. Levermann, *Cryosphere* **13**, 2475–2488 (2019).
8. Materials and methods are available as supplementary materials.
9. J. N. Bassis, L. Ultee, *J. Geophys. Res. Earth Surf.* **124**, 2036–2055 (2019).
10. J. A. Olive, M. D. Behn, E. Mittelstaedt, G. Ito, B. Z. Klein, *Geophys. J. Int.* **205**, 728–743 (2016).
11. F. Clerc, B. M. Minchew, M. D. Behn, *Geophys. Res. Lett.* **46**, 12108–12116 (2019).
12. A. Walter, M. P. Luthi, A. Vieli, *Cryosphere* **14**, 1051–1066 (2020).
13. R. Cassotto, M. Fahnestock, J. M. Amundson, M. Truffer, I. Joughin, *J. Glaciol.* **61**, 76–88 (2015).
14. J. I. Walter *et al*., *Ann. Glaciol.* **53**, 181–192 (2012).
15. J. C. Burton, J. M. Amundson, R. Cassotto, C. C. Kuo, M. Dennin, *Proc. Natl. Acad. Sci. U.S.A.* **115**, 5105–5110 (2018).
16. A. A. Robel, *Nat. Commun.* **8**, 14596 (2017).
17. B. R. Parizek *et al*., *Geology* **47**, 449–452 (2019).
18. M. Trevers, A. J. Payne, S. L. Cornford, T. Moon, *Cryosphere* **13**, 1877–1887 (2019).
19. T. J. W. Wagner, T. D. James, T. Murray, D. Vella, *Geophys. Res. Lett.* **43**, 232–240 (2016).
20. E. M. McFadden, I. M. Howat, I. Joughin, B. Smith, Y. Ahn, *J. Geophys. Res.* **116** (F2), F02022 (2011).

ACKNOWLEDGMENTS

We thank the four anonymous reviewers whose insight improved this manuscript. **Funding:** This work is from the DOMINOS project, a component of the International Thwaites Glacier Collaboration (ITGC). Support came from NSF grant 1738896 and Natural Environment Research Council (NERC) grant NE/S006605/1. Logistics were provided by NSF–U.S. Antarctic Program and NERC–British Antarctic Survey. This study is ITGC contribution no. ITGC-044. **Author contributions:** J.N.B. and B.B. wrote the m-ice model. J.N.B. performed all simulations and data analysis. All authors contributed to planning experiments, interpreting results, and writing and editing the manuscript. **Competing interests:** The authors declare no competing interests; **Data and materials availability:** The development version of the m-ice model is publicly available at https://github.com/jbassis/m-ice. The version of m-ice used in this study along with all scripts used to generate figures is publicly available at https://github.com/jbassis/MICI-2021.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/372/6548/1342/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S10
Tables S1 and S2
References (*21*–*27*)
Movies S1 to S5

8 November 2020; accepted 3 May 2021
10.1126/science.abf6271

HISTORY OF INNOVATION

Who do we invent for? Patents by women focus more on women's health, but few women get to invent

Rembrand Koning[1]*, Sampsa Samila[2], John-Paul Ferguson[3]

Women engage in less commercial patenting and invention than do men, which may affect what is invented. Using text analysis of all U.S. biomedical patents filed from 1976 through 2010, we found that patents with all-female inventor teams are 35% more likely than all-male teams to focus on women's health. This effect holds over decades and across research areas. We also found that female researchers are more likely to discover female-focused ideas. These findings suggest that the inventor gender gap is partially responsible for thousands of missing female-focused inventions since 1976. More generally, our findings suggest that who benefits from innovation depends on who gets to invent.

The inventor gender gap is well established. Although progress has been made, women are still less likely to enter STEM (science, technology, engineering, and mathematics) occupations, less likely to continue in scientific careers, and less likely to become inventors (*1*–*6*). Even today, only 35% of STEM scientists and 13% of U.S. patent inventors are women, suggesting that there are many "lost Curies," talented girls who never grow up to discover and invent (*7*–*9*).

Although the gap itself is cause for concern, its consequences may extend beyond the labor market (*10*). It is possible that women are more likely to invent for women, and if so, the dearth of female inventors may cause society to supply too few inventions that benefit women. Recent work suggests a shortage of projects, procedures, and products aimed at female patients' needs (*11*–*13*). Anecdotal evidence abounds that a disproportionate share of female-focused inventions—such as ovarian cancer tests, personalized breast cancer trials, and cataract-removal procedures [cataracts are more common in women than in men (*14*)]—are invented and brought to market by women (*15*–*17*).

Yet, there is little systematic evidence on whether female-invented products are more likely to focus on the needs of women, especially in knowledge-intensive domains such as biomedicine. On the one hand, the lived experiences of female inventors may lead them to see opportunities to invent for people like themselves, specialize in clinical and disease areas that men may overlook, and invent with more knowledge of women's health (*18*, *19*). On the other hand, medical inventors generally have extensive training in MD and PhD programs, which could eliminate knowledge differences between male and female researchers. Furthermore, inventions follow market demand (*20*). It could be that inventors, male or female, simply invent where the market pulls them.

We measured the gender of inventors and the sex focus of their inventions for all U.S. biomedical patents filed between 1976 and 2010. We found that in this field, inventions by women are more likely to focus on the medical needs of women. This pattern is strongest for all-female invention teams, holds over decades, and is present even within narrow areas of invention. This last finding suggests that the female inventor-invention link is both the result of women working in more female-focused research areas and female inventors identifying opportunities to invent for women regardless of the area in which they work. We also analyzed biomedical research articles between 2002 and 2020 and found that female-discovered ideas are also more likely to be female-focused. That upstream research ideas also exhibit a female inventor-invention link further suggests that the gender gap in who commercializes their ideas has contributed to the sex gap in what types of ideas become inventions.

The starting point for our analysis was a new measure for a patent's focus on the medical needs of men and/or women. We extracted the title, abstract, and start of the summary text from the 441,504 "Drugs and Medical" patents in the PatentsView-NBER (National Bureau of Economic Research) dataset. We then fed this text through the National Library of Medicine's Medical Text Indexer (MTI) (*21*).

The MTI algorithm uses machine learning to map text to the Medical Subject Heading (MeSH) terms that most likely reflect the text's content. More details are provided in supplementary text S1.

The MeSH ontology includes two tags for the sex focus of the patent. A patent's text gets the "Female" tag when it covers "female organs, diseases, physiologic processes, genetics, etc." There is a parallel "Male" tag. We refer to the 12.76% of patents that the MTI tagged "Female" as female-focused patents. Conversely, 13.27% of patents are male-focused. In supplementary text S2, we validate the MTI algorithm, showing that it works well with patent text and that its accuracy is unbiased with respect to inventor gender. In supplementary text S3, we show that although the female and male MeSH tags are a coarse measure of sex focus, they capture meaningful differences in the likelihood that a patent benefits women or men. For example, with word-embedding methods and clinical-trial data, we show that female- and male-tagged patents are roughly 400% more likely to describe inventions that would be evaluated by a female- or male-only trial. Furthermore, with data on disease incidence, we show that patents tagged as female address diseases that affect roughly twice as many women; we found a similar pattern for male-tagged patents and men. We measured inventor gender using a standard dictionary-matching process (supplementary text S4). Our last sample includes 430,060 patents with inventor-gender and sex-focus measures.

Our first set of analyses explored raw trends in the gender composition of inventors. We did so by splitting our data into two groups: patents with majority-female inventing teams (≥50% women) and patents with strictly more male inventors (>50% men). In Fig. 1, we show that women-invented biomedical patents are on the rise, growing from 6.3% in 1976 to 16.2% in 2010. Yet the 3014 inventions by female-majority teams in 2010 are still outnumbered by the 3347 patents invented by men nearly four decades earlier, in 1976. In total, male-majority teams generated 373,774 patents, whereas female-majority teams generated 56,286.

Trends in the sex focus of invention for male-majority teams are shown in Fig. 2A and those of female-majority teams in Fig. 2B, with the black solid line and gray dashed line showing the percentage of patents that are female-focused and male-focused, respectively. Male-majority teams invented more for men than for women in all but 1 year of our data (Fig. 2A), although the difference has shrunk. In 1976, patents by male-majority teams were 15% more male-focused (11.2%) than female-focused (9.7%); in 2010, patents by male-majority teams were 6.2% more male-focused (13.7%) than female-focused (12.9%). This narrowing follows efforts by the National Institutes of Health and National Science Foundation starting in the late 1980s to promote research on women's health (*22*). Overall, of the 373,774 patents invented by male-majority teams, 49,886 (13.3%) were male-focused, 46,453 (12.4%) were female-focused, and the rest were tagged as not sex-focused.

Female-majority teams are substantially more likely to invent for women and only

[1]Harvard Business School, Harvard University, Boston, MA, USA. [2]IESE Business School, Universidad de Navarra, Barcelona, Spain. [3]Desautels Faculty of Management, McGill University, Montreal, Quebec, Canada.
*Corresponding author. Email: rem@hbs.edu

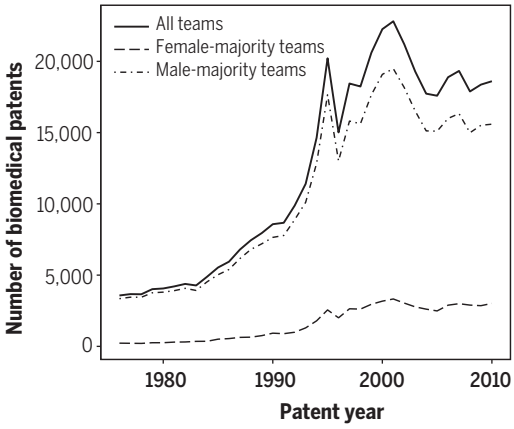


Fig. 1. Total number of U.S. biomedical patents over time along with the number of patents with majority female (≥ 50% women) and majority male (> 50% men) inventor teams. The figure includes 430,060 patents from 1976 through 2010.

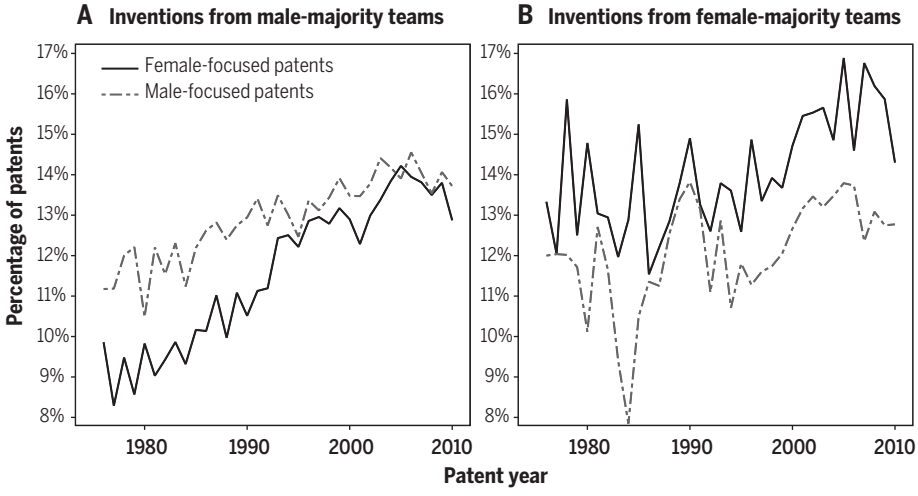


Fig. 2. Percentage of U.S. biomedical patents that are male-focused and female-focused broken out by the gender composition of the inventor team. (**A**) The percentages for patents with majority-male teams (>50% men). (**B**) The percentages for patents with majority-female teams (≥50% women).

slightly less likely to invent for men than are male-majority teams (Fig. 2B). In 1976, patents by female-majority teams were 7% more male-focused (12 versus 11.2%) and 37% were more female-focused (13.3 versus 9.7%) than patents by men. As with male-majority teams, by 2010, female-majority teams invented more often for both sexes. On average, of the 56,286 patents invented by female-majority teams, 8246 (14.7%) were female-focused, 7053 (12.5%) were male-focused, and the rest were not tagged as sex-focused. Compared with male-majority teams, female-majority teams are only 6% (12.5 versus 13.3%) less likely to invent for men but are 18.5% (14.7 versus12.4%) more likely to invent for women.

The inventor gender gap amplifies these invention sex gaps (Fig. 3). The number of female-focused minus male-focused patents invented by male-majority teams over time is shown in Fig. 3A. Such teams regularly invent more than 100 additional male-focused inventions per year. The same difference for female-majority teams is shown in Fig. 3B. Male and female inventors seem to have offsetting tendencies, but because there have been so many more male inventors, in aggregate invention has been skewed toward men's health (Fig. 3C). The rising share of female inventors, and their increased focus on the needs of women, has begun to close and in some years flip the invention sex gap.

To further explore the magnitude and consequences of this female inventor-invention link, we turned to patent-level models that regress a patent's female focus on whether the invention team falls into one of four buckets: all female, majority (50% or more) but not all female, minority female, and all male (our excluded baseline). Our model includes fixed effects for the year interacted with the patent's subcategory (drugs, surgery and instruments, biotechnology, or other) and the patent team's size (*23*). Additional details and tables are provided in supplementary text S5. The estimates from this model, shown in black in Fig. 4A, show a clear dose-response relationship, increasing from 1.3 ± 0.16 (±SE) ($P < 0.001$) percentage points for minority-female teams to 2.4 ± 0.19 ($P < 0.001$) for majority-female teams, and to 4.6 ± 0.30 ($P < 0.001$) for all-female teams. These reflect relative increases of 10, 18, and 35% over the baseline female-focused invention rate of all-male teams.

These estimates are the result of two possible pathways. First, women might work in more female-focused research areas. Second, women might draw on different experiences and knowledge, regardless of their research area. If the first pathway is the main one, then the benefit to women of there being more female inventors will be concentrated in areas that are already female-focused (for example, additional pre-eclampsia treatments). If the second pathway matters more, then women would also see benefits in research areas without a strong prior sex focus (for example, female-focused atrial fibrillation treatments).

To isolate the within-research-area mechanism, we matched female-invented patents to male-invented patents using the patent's publication year, subcategory, team size, and the narrow MeSH level-four disease area most associated with the invention [for example, 1983 drug patents for atrial flutter (C14.280.067.248) with two inventors, or 1990 surgery patents for atrial fibrillation (C14.280.067.198) with three inventors]. The estimates in Fig. 4A in gray come from this matched sample, which is roughly 1/10th the size of our full sample and includes year × subcategory × team size × disease fixed effects. After tightly matching within research areas, just over half of the total effect remains, with effect sizes for minority, majority, and all-female teams of 0.67 ± 0.37 ($P = 0.36$), 1.3 ± 0.37 ($P = 0.002$), and 2.8 ± 0.55 ($P < 0.001$) percentage points, respectively. The female inventor-invention link stems equally from women working in female-focused research areas and from women spotting opportunities in other research areas.

In the supplementary materials, we report additional analyses. In supplementary text S6, we show that inventions by women are not merely more sex-focused in general, which could be so if women work in applied areas in which anatomical sex differences are more central or if women are more likely to run sex and gender analyses (*24*). Instead, and consistent with Fig. 2, we found that teams with female inventors, as compared with all-male inventor teams, are neither more nor less likely to invent for men. Robustness checks are provided in supplementary text S7, including analyses showing that patents with all-female teams target diseases that affect 16% more women and are roughly 40% more likely to describe an invention that would be evaluated by a clinical trial that only enrolls women. Our

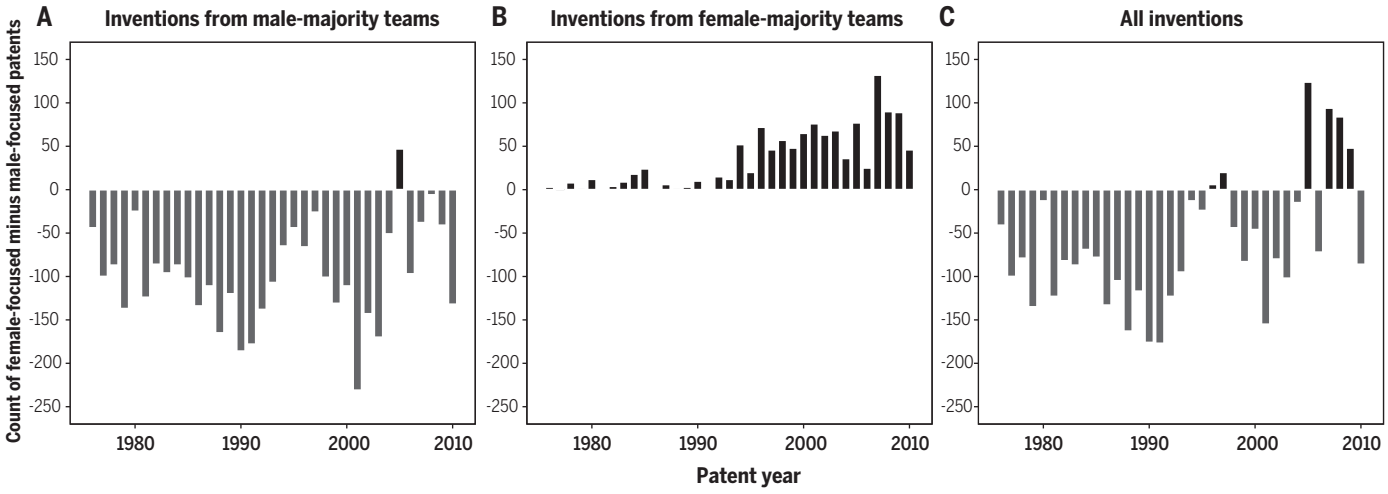


Fig. 3. The number of female-focused patents minus male-focused patents over time. (**A**) The gap for patents with male-majority inventor teams (>50% men). (**B**) The gap for patents with female-majority teams (≥50% women). (**C**) The gap when including both types of teams.

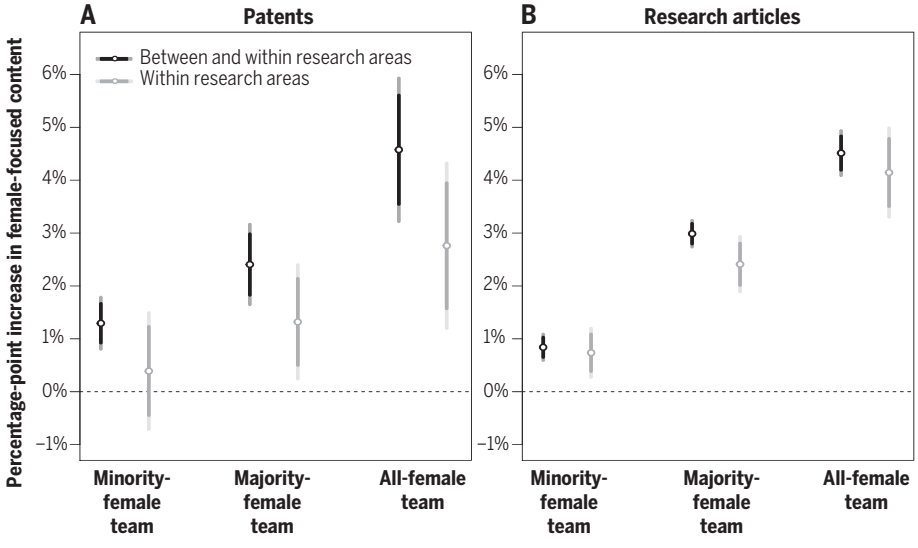


Fig. 4. The increased chance, in percentage points, that a female-invented patent or research article is female-focused, compared with a patent or article from an all-male invention team. The bars show robust 95 and 99% confidence intervals. (**A**) Estimates using the patent data. Black estimates come from a model fit by using OLS with year × subcategory and year × team size fixed effects. The gray estimates are from models that match teams with female inventors to all-male teams by using the patent's year, subcategory, team size, and disease focus. The full sample includes 430,060 patents; the largest matched sample includes 45,751 patents. (**B**) Estimates using the research article data. Black estimates are from models fit by using OLS with year × journal and year × team size fixed effects. The gray estimates are similar but also match teams with female scientists to all-male teams by using the article's year, team size, and disease focus. The full sample includes 2,062,695 articles published in the 1000 journals in PubMed with the highest commercialization impact factors from 2002 through 2020; the largest matched sample includes 119,650 papers.

within-disease-area findings hold if we instead rely on different matching criteria, multidimensional disease fixed effects, or lasso-style machine-learning methods (supplementary text S8), all of which further account for differences in which areas women and men invent.

In supplementary text S9, we show our exploration into heterogeneity in the female inventor-invention link. We found little evidence that the female inventor-invention link has consistently increased or decreased with time. We found suggestive evidence that all-female teams are 2.4 percentage points more likely to invent for women in noncorporate than in corporate settings, which is consistent with the idea that limited women's representation in corporate management leads firms to overlook female-focused research and development opportunities (*25*). We also show that all-female teams are 8.7 percentage points more likely to invent for women in the surgery subcategory, versus 2.3 percentage points in drugs. This last difference is consistent with the idea that the historical hostility toward women in surgery (*26*) has both involved men dismissing promising female-focused surgical ideas and women being especially likely to develop those ideas.

Overall, we found that women's biomedical inventions are more likely than men's inventions to focus on women's needs. That said, patents are a downstream outcome in the process of discovery and invention. Perhaps nascent female inventors discover ideas for both sexes equally, but gender stereotypes lead women to not patent their male-focused ideas (or men to not patent their female-focused ones). Differential selection into patenting, not differences in the knowledge and experiences that women bring to the discovery process, might be responsible.

To test for differential selection, we used data on the upstream research ideas that biomedical patents build on. Specifically, we analyzed 2,062,695 original biomedical research articles from the National Library of Medicine's PubMed database that are at risk of being commercialized and were published between 2002 (the first year we have data on the authors' gender) and 2020 (supplementary

text S10). We again used the MeSH ontology's "Male" and "Female" check tags to asses articles' sex focus, but we did not use the MTI algorithm because articles in PubMed are assigned MeSH tags by human medical indexers.

Research teams with more women are more likely to discover for women (Fig. 4, A and B). There is again a clear dose-response relationship, with the effect increasing from 0.84 ± 0.08 ($P < 0.001$) to 3 ± 0.08 ($P < 0.001$) to 4.5 ± 0.12 ($P < 0.001$) percentage points for minority, majority, and all-female teams, respectively. As in Fig. 4A, the gray estimates in Fig. 4B involve matching and fixed effects to account for differences in men's and women's publishing rates in different narrow disease areas (supplementary text S11). Although the coefficients shrink, the differences are small and nonsignificant. Compared with the baseline female-focused publication rate, papers with all-female teams are about 12% more likely to focus on the health needs of women. Additional robustness tests are reported in supplementary text S12, including analyses that show that discoveries by women are not more likely to be male-focused, that our findings hold under alternative matching schemes, and that discoveries by all-female teams target diseases that affect 47% more women and are >40% more likely to describe an idea that would be evaluated by a clinical trial that enrolls only women.

These findings imply that many promising female-focused discoveries have yet to be commercialized because women are less likely to obtain patents (*1*, *27*). Rough calculations (supplementary text S13) suggest that if all the patents invented between 1976 and 2010 had been produced by men and women equally, then there would have been around 6500 more female-focused inventions. Even when we focused within research areas, thus holding constant the number of patents produced across areas, we still found that inventor gender equality would have resulted in just over 3500 more female-focused inventions. If research articles were produced equally by men and women, then from 2002 to 2020 there would have potentially been 40,000 more female-focused discoveries. This suggests that there may still be many untapped market opportunities to invent for women—opportunities that could in turn improve women's health.

These estimates of the number of lost female inventions are suggestive. The primary goal of this study is to establish a link between the background of inventors and who might benefit from their inventions. We did not observe the counterfactual world in which there are more (or fewer) female inventors or scientists. Hence, the evidence presented here cannot tell us whether women's female-focused inventions could have instead been invented by all-male teams, or the extent to which female inventors might displace men from working on women's health, or whether women might instead pursue other, non–female-focused ideas if there were a greater share of female inventors and inventions. Our analyses also cannot tell us whether the gender mix of inventions is or ever has been socially optimal. Last, although recent work has found that demographic similarity between doctors and patients improves female and African American health outcomes (*19*, *28*, *29*), our analysis does not shed light on whether female-invented and -focused patents are more original or impactful. We see each of these limitations of the present study as a promising avenue for more research.

That research should not be limited to the differences between men and women nor to biomedicine. Our findings sketch a road map for future research on demographics and innovation. Reducing the barriers that disadvantaged groups face when working and inventing has accounted for a nontrivial share of economic growth over the past century (*30*). In virtually all cases, though, whether the population under study has been 19th-century Black inventors (*31*), early-20th-century Eastern European scientists (*32*), or some other group, the focus has been on how discrimination has reduced the overall level of innovation. By contrast, we found a robust and sizable connection between inventor gender and the sex focus of inventions. This is one step in showing how labor-market bias might spill over into product-market bias. Whereas earlier work on gender and innovation has focused on estimating biases against entrepreneurs and inventors while holding the underlying idea constant [for example, (*33*)], we hope that future work will focus on understanding bias against ideas that especially benefit historically disadvantaged groups (*34*). Our findings here suggest that inequities in who invents may lead to inequities in who benefits from invention.

REFERENCES AND NOTES

1. W. W. Ding, F. Murray, T. E. Stuart, *Science* **313**, 665–667 (2006).
2. J. S. Long, M. F. Fox, *Annu. Rev. Sociol.* **21**, 45–71 (1995).
3. P. Azoulay, R. Michigan, B. N. Sampat, *N. Engl. J. Med.* **357**, 2049–2056 (2007).
4. F. Murray, L. Graham, *Ind. Corp. Change* **16**, 657–689 (2007).
5. M. J. Lerchenmueller, O. Sorenson, *Res. Policy* **47**, 1007–1017 (2018).
6. S. Kahn, D. Ginther, "Women and STEM," technical report 23525, National Bureau of Economic Research, June 2017.
7. J. Huang, A. J. Gates, R. Sinatra, A.-L. Barabási, *Proc. Natl. Acad. Sci. U.S.A.* **117**, 4609–4616 (2020).
8. A. Bell, R. Chetty, X. Jaravel, N. Petkova, J. Van Reenen, *Q. J. Econ.* **134**, 647–713 (2019).
9. J. Milli, E. Williams-Baron, M. Berlan, J. Xia, B. Gault, "Equity in innovation: Women inventors and patents," an Institute for Women's Policy Research Report (Institute for Women's Policy Research, 2016).
10. L. S. Rotenstein, A. B. Jena, *N. Engl. J. Med.* **378**, 2255–2257 (2018).
11. M. Dusenbery, *Doing Harm: The Truth about How Bad Medicine and Lazy Science Leave Women Dismissed, Misdiagnosed, and Sick* (HarperOne, 2018).
12. C. C. Perez, *Invisible Women: Data Bias in a World Designed for Men* (Abrams Press, 2019).
13. Office of Research on Women's Health, "Report of the advisory committee on research on women's health," Technical report, NIH publication no. 17 OD 7995 (National Institutes of Health, 2016).
14. L. Lou *et al.*, *JAMA Ophthalmol.* **136**, 116–121 (2018).
15. M. Herper, *Forbes*, 16 April 2018.
16. K. Hafner, A breast cancer surgeon who keeps challenging the status quo. *New York Times*, 30 September 2015.
17. National Institutes of Health, "Changing the face of medicine: Dr. Patricia E. Bath," technical report, National Institutes of Health, 2019; https://cfmedicine.nlm.nih.gov/physicians/biography_26.html.
18. L. Schiebinger, editor, *Gendered Innovations in Science and Engineering* (Stanford Univ. Press, 2008).
19. B. N. Greenwood, S. Carnahan, L. Huang, *Proc. Natl. Acad. Sci. U.S.A.* **115**, 8569–8574 (2018).
20. D. Acemoglu, J. Linn, *Q. J. Econ.* **119**, 1049–1090 (2004).
21. J. Mork, A. Aronson, D. Demner-Fushman, *J. Biomed. Semantics* **8**, 8 (2017).
22. N. C. Lee, *Public Health Rep.* **130**, 121–122 (2015).
23. S. Wuchty, B. F. Jones, B. Uzzi, *Science* **316**, 1036–1039 (2007).
24. M. W. Nielsen, J. P. Andersen, L. Schiebinger, J. W. Schneider, *Nat. Hum. Behav.* **1**, 791–796 (2017).
25. K. B. Whittington, L. Smith-Doerr, *Gend. Soc.* **22**, 194–218 (2008).
26. K. Frances, Conley. *Walking Out on the Boys* (Macmillan, 1999).
27. K. Jensen, B. Kovács, O. Sorenson, *Nat. Biotechnol.* **36**, 307–309 (2018).
28. M. Alsan, O. Garrick, G. Graziani, *Am. Econ. Rev.* **109**, 4071–4111 (2019).
29. B. N. Greenwood, R. R. Hardeman, L. Huang, A. Sojourner, *Proc. Natl. Acad. Sci. U.S.A.* **117**, 21194–21200 (2020).
30. C.-T. Hsieh, E. Hurst, C. I. Jones, P. J. Klenow, *Econometrica* **87**, 1439–1474 (2019).
31. D. Lisa, Cook. Violence and economic activity: Evidence from African American patents, 1870–1940. *J. Econom. Growth* **19**, 221–257 (2014).
32. P. Moser, S. San, "Immigration, science, and invention. Evidence from the quota acts," technical report, working paper (2020).
33. A. W. Brooks, L. Huang, S. W. Kearney, F. E. Murray, *Proc. Natl. Acad. Sci. U.S.A.* **111**, 4427–4431 (2014).
34. M. W. Nielsen, C. W. Bloch, L. Schiebinger, *Nat. Hum. Behav.* **2**, 726–734 (2018).
35. R. Koning, S. Samila, J.-P. Ferguson, Data for "Who do we invent for? Patents by women focus more on women's health, but few women get to invent". Harvard Dataverse (2021); https://doi.org/10.7910/DVN/V8NJUV.

ACKNOWLEDGMENTS

J. Comeau, A. Fernandez, H. Ham, J. Kim, Y. Zhou, S. Neumeyer, and E. Vachlas provided invaluable research assistance. We thank audiences at the AEA, Barcelona GSE, BAIC, MIT, HBS, IESE, KU Leuven, Madrid Work and Occupations, and Queen's University along with C. Cunningham, W. Ding, K. Ferreira, D. Gross, S. Hasan, J. Krieger, M. Marx, K. Myers, R. Nanda, D. Spar, and A. Stern for helpful feedback and suggestions. **Funding:** This project has received funding from Harvard Business School and also from the European Union's Horizon 2020 research and innovation program under the Marie Sklodowska-Curie grant agreement 799330. **Author contributions:** All authors contributed equally to this work. **Competing interests:** The authors have no known competing interests. **Data and materials availability:** All data and code are available in the Harvard Dataverse at https://doi.org/10.7910/DVN/V8NJUV (*35*).

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/372/6548/1345/suppl/DC1
Supplementary Text
Figs. S1 to S8
Tables S1 to S50
References (*36*–*54*)

14 January 2020; accepted 6 May 2021
10.1126/science.aba6990

SIGNAL TRANSDUCTION

NF-κB dynamics determine the stimulus specificity of epigenomic reprogramming in macrophages

Quen J. Cheng[1,2]†, Sho Ohta[1]†, Katherine M. Sheu[1], Roberto Spreafico[1,3], Adewunmi Adelaja[1], Brooks Taylor[1], Alexander Hoffmann[1,3]*

The epigenome of macrophages can be reprogrammed by extracellular cues, but the extent to which different stimuli achieve this is unclear. Nuclear factor κB (NF-κB) is a transcription factor that is activated by all pathogen-associated stimuli and can reprogram the epigenome by activating latent enhancers. However, we show that NF-κB does so only in response to a subset of stimuli. This stimulus specificity depends on the temporal dynamics of NF-κB activity, in particular whether it is oscillatory or non-oscillatory. Non-oscillatory NF-κB opens chromatin by sustained disruption of nucleosomal histone–DNA interactions, enabling activation of latent enhancers that modulate expression of immune response genes. Thus, temporal dynamics can determine a transcription factor's capacity to reprogram the epigenome in a stimulus-specific manner.

The cellular epigenome, a regulatory network involving transcription factors, chromatin architecture, and histone modifications, contains stable, heritable information that determines cell type–specific programs of gene expression (*1*, *2*). Nevertheless, the epigenome of differentiated cells remains highly plastic, particularly in immune cells such as macrophages (*3*, *4*). These immune sentinel cells detect microenvironmental immune threats, mount appropriate gene expression responses, and reprogram their epigenomes to tailor subsequent immune responses (*5*). At a molecular level, this reprogramming is initiated by the activity of signal-dependent transcription factors (TFs) such as nuclear factor kappa-light-chain-enhancer of activated B cells (NF-κB) (*6*). In cooperation with chromatin modifiers and pioneering TFs, signal-dependent TFs increase chromatin accessibility and modify histones at previously silent regions of the genome, thus converting latent enhancers to poised or active states (*7*–*9*). NF-κB activated by bacterial lipopolysaccharide (LPS) has been a model TF in this field. However, the degree to which NF-κB or other TFs can alter the epigenome in response to different stimuli is unknown.

To investigate the stimulus specificity of epigenomic reprogramming, we stimulated bone marrow–derived macrophages (BMDMs) with five well-characterized ligands: tumor necrosis factor (TNF) and the Toll-like receptor agonists Pam3CSK, CpG, LPS, and Poly(I:C) (polyinosinic: polycytidylic acid). We performed chromatin immunoprecipitation sequencing (ChIP-seq) using antibodies recognizing the monomethylation of lysine 4 on histone H3 (H3K4me1) to identify latent enhancers that were activated upon stimulation. We found 3978 enhancer regions that segregated into two clusters by the *k*-means algorithm (Fig. 1A and fig. S1). The latent enhancers in cluster 1 were most strongly activated in response to LPS and Poly(I:C) and were enriched for interferon response factor (IRF) motifs (Fig. 1B, top), consistent with the fact that these stimuli activate IRF3 and type I interferon (*10*); in *Irf3*$^{-/-}$*Ifnar*$^{-/-}$ BMDMs, these regions no longer acquired H3K4me1 (Fig. 1C, top). Weak H3K4me1 signal was preserved in response to TNF, which activates IRF1 but not IRF3 (*11*).

In contrast, the regions in cluster 2 were highly enriched for NF-κB motifs (Fig. 1B, bottom), implying that these were latent NF-κB enhancers. We examined the contribution of other stimulus-responsive signaling pathways and found that the gain of H3K4me1 was preserved in *Irf3*$^{-/-}$*Ifnar*$^{-/-}$ BMDMs (Fig. 1C, bottom) but disrupted by pharmacologic inhibition of mitogen-activated protein kinase (MAPK) pathways (fig. S2, A and B). MAPK inhibition also blocked activation of latent enhancers in cluster 1, suggesting that the MAPK pathway is generally critical for epigenomic reprogramming (*12*) and does not specifically collaborate with NF-κB.

We next examined the contribution of NF-κB family members. RelA:p50 is the dominant NF-κB dimer in macrophages (*13*), but cRel also plays a role (*14*). We knocked out cRel (*Rel*$^{-/-}$) and found that H3K4me1 ChIP-seq signals were unchanged (fig. S2C), including at the *Il12b* promoter (*14*). Knocking out p50 (*nfkb1*$^{-/-}$) only weakly diminished H3K4me1 signals, indicating that partial compensation by RelA:p52 or RelA homodimers was sufficient (*15*). These data indicated that RelA is the primary activator of latent NF-κB enhancers in macrophages.

To focus on latent NF-κB enhancers, we used RelA ChIP-seq data (*16*) to identify 1071 regions in cluster 2 that contained a RelA binding event. Unexpectedly, these regions acquired H3K4me1 in a stimulus-specific manner, even though all five stimuli tested activate NF-κB (*17*). Within these regions, the H3K4me1 signal was strongly induced by Pam3CSK, CpG, and LPS, with median $\log_2$ fold changes of 1.07, 1.16, and 1.33, respectively. TNF and Poly(I:C) produced less H3K4me1, with median $\log_2$ fold changes of 0.60 and 0.70, respectively (Fig. 1D, top). A pairwise comparison of samples quantitatively confirmed the stimulus specificity of these NF-κB enhancers (Fig. 1D, bottom).

This stimulus specificity would be difficult to explain if NF-κB acted as a binary on-off switch, but NF-κB is activated with complex, stimulus-specific temporal dynamics (*17*–*19*). In response to various stimuli, NF-κB enters the nucleus with distinct speeds, amplitudes, and durations and may oscillate between the nucleus and cytoplasm. To determine whether stimulus-specific NF-κB dynamics play a role in stimulus-specific activation of latent enhancers, we used live-cell microscopy of BMDMs expressing NF-κB–RelA fused with the mVenus fluorophore (mVenus-RelA) (*20*) to measure the single-cell dynamics of NF-κB–RelA in response to each of the five ligands (Fig. 1E). We quantified the six NF-κB dynamic features that function as signaling codons to encode ligand identity and dose (*20*) and correlated them to mean H3K4me1 counts in the NF-κB–activated latent enhancers (fig. S3). Oscillatory power [correlation coefficient (r) = −0.95], total activity (r = 0.77), and peak amplitude (r = 0.78) were highly correlated with the capacity of a given stimulus to activate latent enhancers (Fig. 1F).

We hypothesized that temporal dynamics of NF-κB activity might affect its interaction with chromatin. Crystallographic studies imply that stable NF-κB–DNA binding requires the DNA to be nucleosome-free, because NF-κB dimers embrace the DNA double helix circumferentially (*21*, *22*) (Fig. 2A). However, NF-κB can interact with nucleosomal DNA, particularly when its binding site is distal to the nucleosome dyad (*23*). Indeed, the DNA-histone interface is composed of low-affinity interactions that allow spontaneous disassociation or "breathing" (*24*). Thus, successive disruptions of DNA-histone contacts by NF-κB, in collaboration with remodeling complexes such as SWI/SNF (*25*), chaperone proteins such as FACT (*26*, *27*), and/or pioneer factors such as Pu.1 or CEB/Pα (*28*), may displace the nucleosome (Fig. 2B). This may be followed by the deposition of histone modifications on neighboring nucleosomes, resulting in a poised or active enhancer (*7*).

We created a multistep model describing how dynamical NF-κB activity might interact

[1]Department of Microbiology, Immunology, and Molecular Genetics, University of California, Los Angeles, CA 90095, USA. [2]Division of Infectious, Diseases Department of Medicine, David Geffen School of Medicine, University of California, Los Angeles, CA 90095, USA. [3]Institute for Quantitative and Computational Biosciences, University of California, Los Angeles, CA 90095, USA.
*Corresponding author. Email: ahoffmann@ucla.edu
†These authors contributed equally to this work.

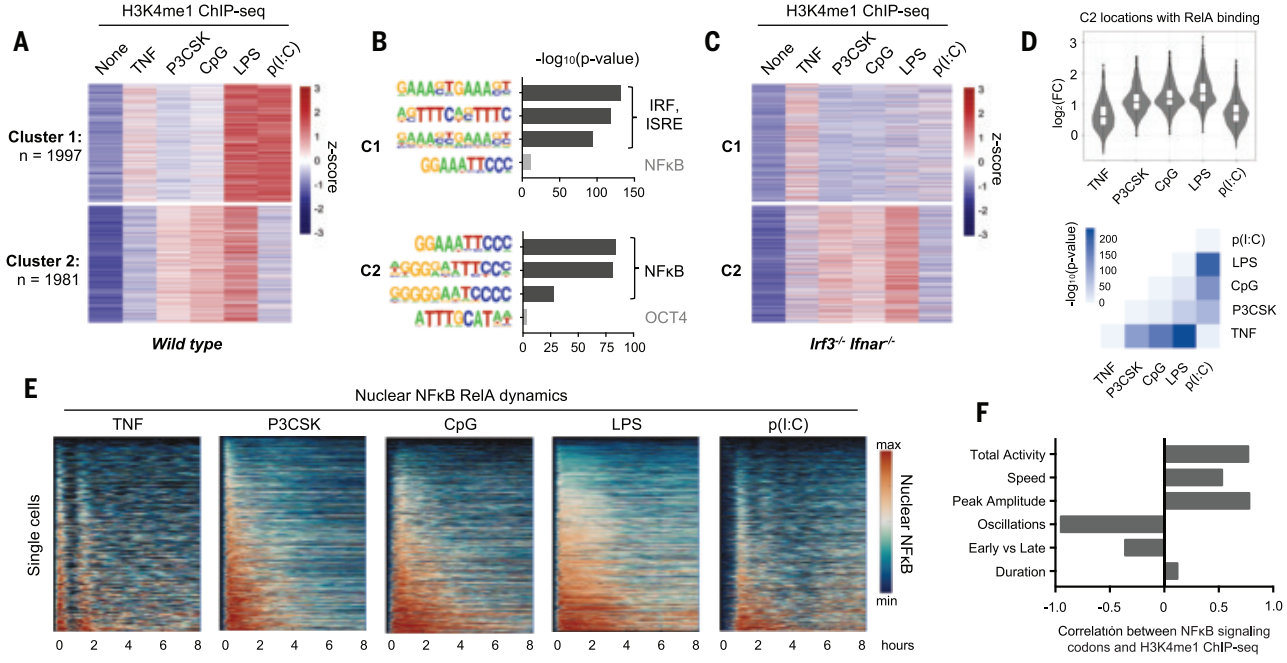


Fig. 1. NF-κB–activated latent enhancers are stimulus-specific and correlate to dynamic features of NF-κB activity. (**A**) Heatmap of H3K4me1 ChIP-seq inducible peaks from BMDMs stimulated with five ligands for 8 hours, unsupervised *k*-means clustering. Average of two biological replicates. (**B**) Known transcription factor motifs with greatest enrichment in cluster 1 and cluster 2 peaks. ISRE, interferon-stimulated response element; OCT4, organic cation/carnitine transporter 4. (**C**) Heatmap of H3K4me1 ChIP-seq in *Irf3*$^{-/-}$*Ifnar*$^{-/-}$ BMDMs, using the same clusters as in (A). (**D**) Violin and box plots of $\log_2$ fold change in H3K4me1 signal of 1071 NF-κB enhancers from cluster 2 that also contain an NF-κB–RelA binding event. Corresponding matrix of *P* values of H3K4me1 ChIP-seq fold change, by two-tailed *t* test between pairs of conditions. (**E**) Heatmaps of NF-κB activity in single cells by live cell microscopy of *mVenus-RelA* BMDMs, showing nuclear abundance of NF-κB in response to five stimuli over 8 hours. (**F**) Bar graph of correlation coefficients between mean H3K4me1 ChIP-seq counts of NF-κB enhancers and the six signaling codons of NF-κB dynamics (*20*); see also fig. S3.

with nucleosomal DNA. A series of 14 Hill equations described the competition between NF-κB and histone for interaction with DNA (Fig. 2C), reflecting the number of contact points in the histone octamer–DNA crystal structure (*29*). Relative rates of nucleosome wrapping and unwrapping were based on available biophysical data (*30*). With measured single-cell NF-κB activities (Fig. 1E) as inputs, the model simulations reproduced the differences in experimental H3K4me1 ChIP-seq data (Fig. 2, D and E, and fig. S4, A and B).

We used the model to investigate which features of NF-κB dynamics affect chromatin accessibility. We examined the three features most highly correlated with the H3K4me1 ChIP-seq data (Fig. 1F): oscillations, amplitude, and total activity. The model indicated that a non-oscillatory dynamic produces a twofold greater chromatin accessibility than an oscillatory dynamic (Fig. 2F). The model also indicated that NF-κB activity must have a minimal amplitude (Fig. 2G and fig. S4C) and extend for a minimal duration (Fig. 2H and fig. S4D) to open chromatin; but above these thresholds, non-oscillatory NF-κB always has greater capacity to open chromatin than does oscillatory NF-κB. This was consistent across a range of parameter values (fig. S5). These simulations predicted that the presence or absence of oscillations, not the maximum amplitude or duration of activity, is the key determinant of whether NF-κB preserves or alters the chromatin state.

To test this prediction, we generated a mouse in which NF-κB dynamics are perturbed. When activated, NF-κB rapidly induces expression of *Nfkbia*, whose gene product is the negative regulator IκBα (Fig. 3A) (*31*). IκBα knockout alone is perinatal lethal owing to persistent inflammation (*32*), but we rescued this lethality by genetically ablating endogenous TNF expression (*33*). We then crossed the composite knockout strain with *mVenus-RelA* knock-in mice to examine the dynamics of NF-κB by live-cell microscopy. IκBα$^{-/-}$ BMDMs responded to TNF with altered NF-κB dynamics compared with wild-type (WT) controls (Fig. 3B). We quantified the differences in the distribution of single-cell dynamic features by the non-parametric Kolmogorov-Smirnov (K-S) test and found that the greatest dynamic difference between IκBα$^{-/-}$ and WT was a loss of oscillations, with a test statistic (*D*) of 0.85, corresponding to $P < 10^{-16}$ (Fig. 3C and fig. S6A). Other dynamic features were either unaffected or favored WT cells, as in the case of activation speed ($D = 0.66$) and early-versus-late activity ($D = 0.52$). The area under the NF-κB activity curve slightly favored WT cells at all time points (Fig. 3C and fig. S6B). We concluded that loss of IκBα abolished NF-κB oscillations without increasing its total activity.

We examined the chromatin state by stimulating BMDMs from IκBα$^{-/-}$ and littermate control mice with TNF and performed the assay for transposase-accessible chromatin using sequencing (ATAC-seq) at 2, 4, and 8 hours. This was followed by a 16-hour period without TNF, after which a final ATAC-seq sample was collected (fig. S7A). We identified 1443 genomic regions that demonstrated TNF-inducible chromatin accessibility in either genotype. Of these, 332 were differentially inducible between control and IκBα$^{-/-}$, and 97% of those 332 regions ($n = 322$) had greater chromatin accessibility in the knockout than in the control (Fig. 3D), despite the slight reduction in total NF-κB activity (fig. S6B). These differentially inducible regions were enriched for NF-κB motifs (Fig. 3E), and 311 of 322 regions showed RelA binding by ChIP-seq (fig. S7B). Ninety-six percent were located in intergenic or intronic portions of the genome (fig. S7C), suggesting that they function as cis-acting enhancers of immune genes such as *Ccl5* (Fig. 3F), which requires chromatin remodeling for maximal expression (*16*).

Our model predicted that chromatin accessibility is primarily determined by whether

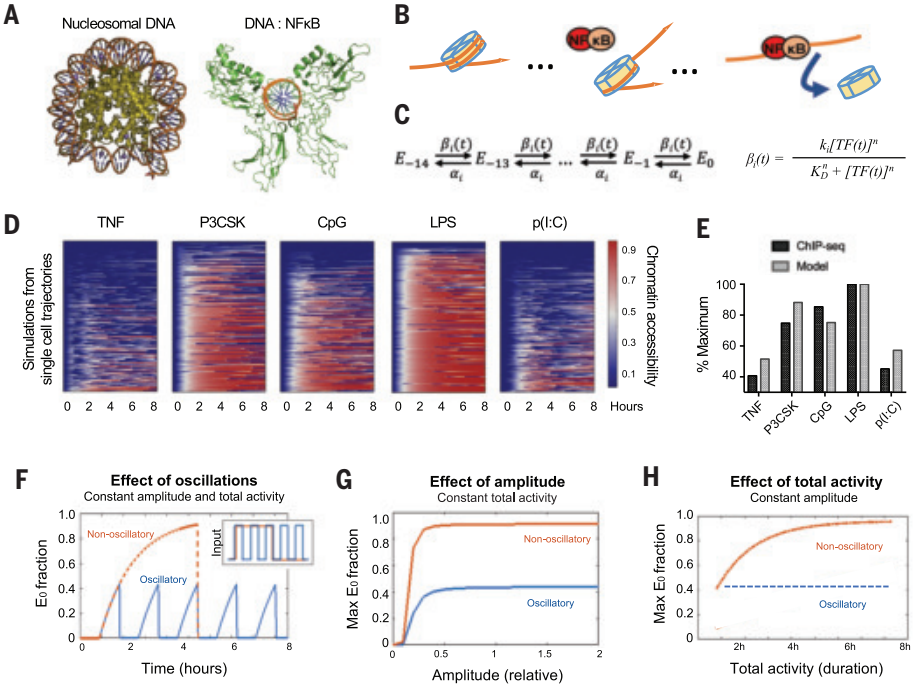


Fig. 2. Mathematical model predicts epigenetic response to distinct dynamic features of NF-κB. (**A**) Crystal structures of nucleosomal DNA [Protein Data Bank (PDB) ID 1F66] versus NF-κB–bound DNA (PDB ID 1VKX), where the RelA:p50 NF-κB dimer is in green. (**B**) Schematic of model illustrating NF-κB–driven nucleosome displacement. (**C**) Multistep model with 14 steps to complete nucleosome unwrapping, each expressed as a Hill function (for further explanation, see supplementary materials). E_{-14}, fully closed state; E_0, fully open state. (**D**) Heatmaps of simulations of chromatin opening in response to different stimuli, using single-cell trajectories from NF-κB microscopy data as input. (**E**) Model simulation versus ChIP-seq data. Mean ChIP-seq counts from 1071 latent NF-κB enhancers (Fig. 1D), background-subtracted and scaled to maximum signal (LPS stimulation). Model simulations are mean of maximum E_0 fraction per cell (compare with fig. S3A), scaled to LPS condition. (**F**) Model simulation of predicted chromatin accessibility comparing oscillatory versus non-oscillatory input activities. (**G** and **H**) Model simulation of predicted chromatin opening across a range of amplitudes and durations.

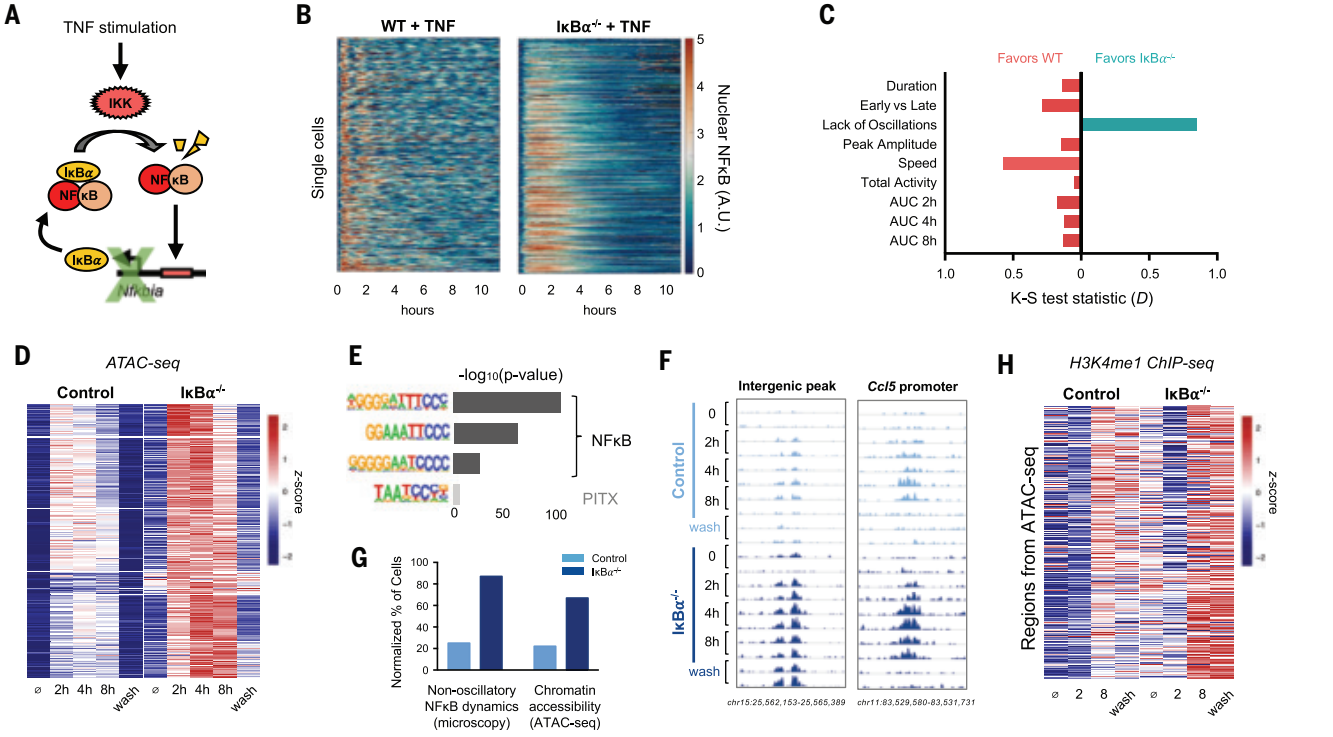


Fig. 3. IκBα knockout abolishes NF-κB oscillations, increases chromatin accessibility, and activates latent enhancers. (**A**) Schematic of IκBα as inducible negative regulator of NF-κB. IKK, IκB kinase. (**B**) Heatmap of single-cell NF-κB activity by live cell microscopy of mVenus-RelA BMDMs, comparing TNF response in WT versus IκBα$^{-/-}$ macrophages. A.U., arbitrary units. (**C**) Bar graph of K-S test statistic for difference in distribution of six signaling codons and areas under NF-κB activity curve (AUC), comparing IκBα$^{-/-}$ and WT. (**D**) Heatmap of ATAC-seq signal after TNF stimulation at 322 genomic regions that are TNF-inducible and differential between IκBα$^{-/-}$ and control. Average of two biological replicates. The term "wash" indicates 8 hours with followed by 16 hours without TNF stimulation. (**E**) Known transcription factor motifs with greatest enrichment in differentially inducible ATAC-seq regions. (**F**) Genome browser tracks for representative differentially inducible ATAC-seq regions, two replicates per time point. (**G**) Percentage of cells with non-oscillatory NF-κB trajectories by microscopy, compared with relative percentage of cells with accessible chromatin by ATAC-seq at *Chr15* intergenic peak (F). (**H**) Heatmap of H3K4me1 ChIP-seq signal after TNF stimulation over the 322 regions defined as differentially inducible by ATAC-seq. Average of two biological replicates. The term "wash" indicates 8 hours with followed by 16 hours without TNF stimulation.

NF-κB is oscillatory or non-oscillatory at the single-cell level. We therefore considered that the magnitude of ATAC-seq signal can be interpreted as the proportion of cells in a sample in which a particular region of DNA is accessible. As determined by microscopy, 87% of IκBα$^{-/-}$ cells had non-oscillatory NF-κB, compared with 25% in WT cells. This was similar to the magnitude of ATAC-seq differences between IκBα$^{-/-}$ and control. For example, at an intergenic peak on chromosome 15, 67% of IκBα$^{-/-}$ cells showed accessible chromatin, compared with 22% of control cells (Fig. 3G).

To confirm that the negative feedback function of IκBα was indeed critical for the observed effects, we used an IκBα$^{κB/κB}$ mutant in which NF-κB binding sites in the promoter of the *Nfkbia* gene are disrupted (*34*) (fig. S8A). In this model, basal IκBα expression is preserved, and the mice live into adulthood without requiring compound suppressor mutations. We found that upon TNF stimulation, IκBα$^{κB/κB}$ BMDMs activated NF-κB in a non-oscillatory manner with minimal disruption of other dynamic features (fig. S8, B to E). ATAC-seq analysis of TNF-stimulated WT versus IκBα$^{κB/κB}$ BMDMs recapitulated our findings in the IκBα$^{-/-}$ system, with 131 genomic regions demonstrating greater gain of chromatin accessibility in the mutant compared with WT (fig. S8F). These regions were enriched for NF-κB motifs, and 90% showed RelA binding by ChIP-seq (fig. S8, G and H). Taken together, the ATAC-seq data from both IκBα$^{-/-}$ and IκBα$^{κB/κB}$ experimental models indicated that loss of NF-κB oscillations results in greater chromatin accessibility at NF-κB binding sites.

We examined whether regions with differentially inducible chromatin accessibility acquired the corresponding histone mark of enhancers. H3K4me1 ChIP-seq in TNF-stimulated BMDMs showed that in the 322 differentially inducible ATAC-seq regions there was also a greater gain of H3K4me1 signal in IκBα$^{-/-}$ than in littermate controls (Fig. 3H). These histone marks persisted for 16 hours after TNF was removed, suggesting that chromatin opening facilitated by NF-κB may be transient but leads to durable H3K4 methylation even after the stimulus is removed, thus activating a latent enhancer and reprogramming the epigenome.

Because histone methylation is more durable and more indicative of enhancer function, we analyzed the H3K4me1 ChIP-seq data independently and identified 2081 regions that acquired more H3K4 methylation in IκBα$^{-/-}$ cells than did controls (Fig. 4A). These differentially induced, dynamics-dependent enhancers persisted after the TNF stimulus was removed, were enriched for NF-κB motifs (Fig. 4B), and showed significant overlap with the set of stimulus-specific NF-κB enhancers identified

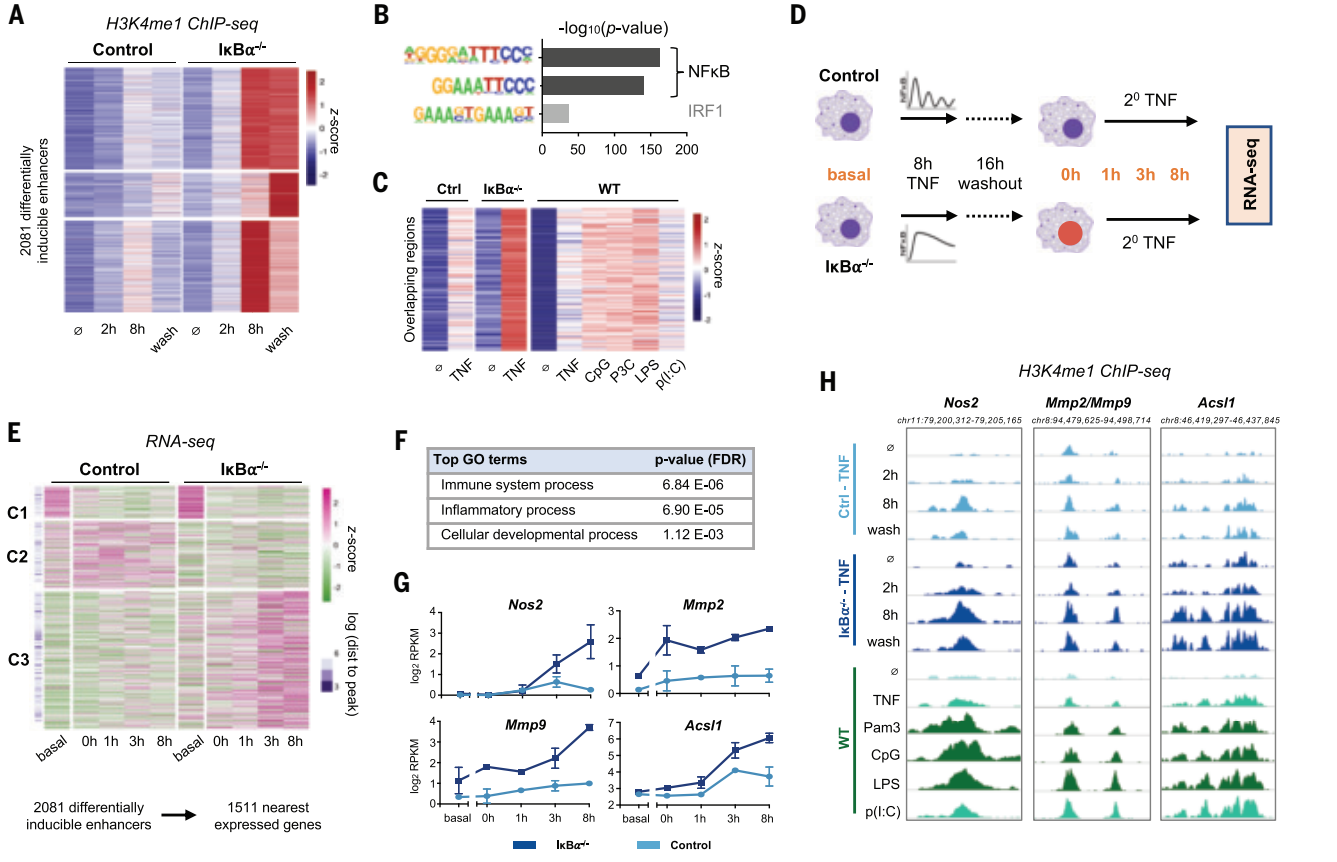


Fig. 4. NF-κB dynamics-dependent enhancers are associated with dynamics-dependent gene expression. (**A**) Heatmap of H3K4me1 ChIP-seq signal after TNF stimulation at 2081 dynamics-dependent enhancers that are differentially induced by TNF between IκBα$^{-/-}$ and control. Average of two biological replicates. The term "wash" indicates 8 hours with followed by 16 hours without TNF. (**B**) Known transcription factor motifs with greatest enrichment in dynamics-dependent enhancers. (**C**) Heatmap of H3K4me1 signal after 8-hour stimulation at regions that overlap between (A) and Fig. 1D (n = 211 regions, P for overlap = 3.0 × 10^{-135}). (**D**) Schematic of RNA-seq experiment. (**E**) Heatmap showing expression of genes closest to dynamics-dependent enhancers, where cluster 3 exhibits differential gene expression between IκBα$^{-/-}$ and control. Average of two biological replicates. (**F**) Top biological process ontology terms for genes in cluster 3 of (E). FDR, false discovery rate. (**G**) Examples of genes differentially induced between IκBα$^{-/-}$ and control, average and standard deviation of two replicates. RPKM, reads per kilobase per million mapped reads. (**H**) Genome browser tracks of differentially inducible H3K4me1 peaks near differentially inducible genes, showing TNF-stimulated IκBα$^{-/-}$ versus control and stimulus-specific response in WT BMDMs. More darkly shaded tracks indicate non-oscillatory NF-κB conditions. Average of two biological replicates.

previously (Fig. 1D, 211 genomic regions, $P = 3.0 \times 10^{-135}$). The inducible ChIP-seq signal was consistently greater when NF-κB dynamics were non-oscillatory rather than oscillatory, whether by genetic perturbation or by stimulus-specific signaling mechanisms (Fig. 4C).

We tested whether NF-κB dynamics-dependent enhancers alter macrophage transcriptional responses to subsequent stimulation. We treated IκBα$^{-/-}$ and littermate control BMDMs with TNF for 8 hours, let cells rest for 16 hours, and then restimulated with TNF for up to 8 hours and collected samples for mRNA sequencing (mRNA-seq) (Fig. 4D). We explored the relation between differentially inducible enhancers and gene expression with two approaches. First, we identified the nearest expressed genes to the 2081 enhancers, removed duplicates, and found three distinct patterns of expression for the resulting 1511 genes (Fig. 4E). Cluster 1 and 2 genes were not TNF-responsive in either condition, reflecting an intrinsic limitation of this approach in which enhancers often do not regulate their nearest genes (*35*). Despite this limitation, cluster 3 genes (58% of total) were both TNF-responsive and more strongly induced in IκBα$^{-/-}$ BMDMs. Of these genes, 88% were not induced in controls at all (using a twofold threshold). These differentially regulated genes were enriched for ontology terms "immune system process" and "inflammatory process" (Fig. 4F), indicating that non-oscillatory NF-κB epigenetically reprograms macrophages to enhance their immune response.

We also examined our data with a gene-centric approach. From the RNA-seq dataset, we identified 1958 TNF-inducible genes, 482 of which were differentially induced in IκBα$^{-/-}$ versus control (fig. S9, A and B). For each gene, we annotated the genomic distance to the nearest differentially inducible H3K4me1 ChIP-seq peak. Differentially induced genes were closer to differentially induced ChIP-seq peaks ($P = 1.13 \times 10^{-9}$) than genes that were not differentially induced (fig. S9, C and D). Thus, both analyses indicated that NF-κB dynamics-dependent enhancers regulate gene expression responses to a subsequent stimulus.

The dynamics-dependent gene expression program included *Nos2*, *Mmp2*, and *Mmp9*, which are well-defined markers of classical macrophage activation (*36*), as well as *Acsl1*, which plays a role in the pathogenesis of atherosclerosis (*37*) (Fig. 4G). Each of these genes had a nearby enhancer that acquired more H3K4me1 signal in the presence of non-oscillatory NF-κB, whether in the IκBα$^{-/-}$ system or in WT BMDMs stimulated with various ligands (Fig. 4H). These specific examples further suggested that latent enhancers activated by non-oscillatory NF-κB regulate genes involved in macrophage activation.

Our results indicate that the dynamics of NF-κB activity, particularly whether it is oscillatory or non-oscillatory, determine NF-κB's capacity to reprogram the macrophage epigenome. We show with a mathematical model how biophysical principles governing nucleosome dynamics might decode stimulus-specific NF-κB dynamical features. The role of temporal dynamics may thus complement the structure-function model in which distance from the nucleosome core determines accessibility to partially exposed DNA motifs (*38*). Together, TF dynamics and motif accessibility may regulate the sensitivity of a particular nucleosome to eviction. To date, the function of NF-κB oscillations has been unclear given that there is little difference in the expression of poised inflammatory-response genes induced by oscillatory versus non-oscillatory NF-κB (*39*, *40*). We propose that in response to some stimuli, the role of oscillations is to maintain the epigenomic state while exploiting existing poised enhancers for inflammatory gene activation. However, in response to other immune threats, non-oscillatory NF-κB induces a comparable gene expression program while also activating latent enhancers, thus changing the epigenomic state of the cell and its response to subsequent stimuli. Although further work will be needed to determine the physiological implications of NF-κB dynamics-dependent enhancers and to identify the proteins that collaborate with NF-κB to evict nucleosomes, our study establishes TF temporal dynamics as a key mechanistic determinant of epigenomic reprogramming.

REFERENCES AND NOTES

1. S. Heinz *et al*., *Mol. Cell* **38**, 576–589 (2010).
2. C. D. Allis, T. Jenuwein, *Nat. Rev. Genet.* **17**, 487–500 (2016).
3. C. K. Glass, G. Natoli, *Nat. Immunol.* **17**, 26–33 (2016).
4. L. B. Ivashkiv, *Trends Immunol.* **34**, 216–223 (2013).
5. P. J. Murray, T. A. Wynn, *Nat. Rev. Immunol.* **11**, 723–737 (2011).
6. T. Lawrence, *Cold Spring Harb. Perspect. Biol.* **1**, a001651 (2009).
7. R. Ostuni *et al*., *Cell* **152**, 157–171 (2013).
8. M. U. Kaikkonen *et al*., *Mol. Cell* **51**, 310–325 (2013).
9. S. Heinz, C. E. Romanoski, C. Benner, C. K. Glass, *Nat. Rev. Mol. Cell Biol.* **16**, 144–154 (2015).
10. K. Honda, A. Takaoka, T. Taniguchi, *Immunity* **25**, 349–360 (2006).
11. A. Yarilina, K.-H. Park-Min, T. Antoniv, X. Hu, L. B. Ivashkiv, *Nat. Immunol.* **9**, 378–387 (2008).
12. A. M. Klein, E. Zaganjor, M. H. Cobb, *Curr. Opin. Cell Biol.* **25**, 272–277 (2013).
13. S. Mitchell, J. Vargas, A. Hoffmann, *Wiley Interdiscip. Rev. Syst. Biol. Med.* **8**, 227–241 (2016).
14. S. Sanjabi, A. Hoffmann, H.-C. Liou, D. Baltimore, S. T. Smale, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 12705–12710 (2000).
15. A. Hoffmann, T. H. Leung, D. Baltimore, *EMBO J.* **22**, 5530–5539 (2003).
16. A.-J. Tong *et al*., *Cell* **165**, 165–179 (2016).
17. S. L. Werner, D. Barken, A. Hoffmann, *Science* **309**, 1857–1861 (2005).
18. M. Behar, A. Hoffmann, *Curr. Opin. Genet. Dev.* **20**, 684–693 (2010).
19. M. W. Covert, T. H. Leung, J. E. Gaston, D. Baltimore, *Science* **309**, 1854–1857 (2005).
20. A. Adelaja *et al*., *Immunity* **54**, 916–930.e7 (2021).
21. F. E. Chen, D. B. Huang, Y. Q. Chen, G. Ghosh, *Nature* **391**, 410–413 (1998).
22. R. K. Suto, M. J. Clarkson, D. J. Tremethick, K. Luger, *Nat. Struct. Biol.* **7**, 1121–1124 (2000).
23. I. N. Lone *et al*., *PLOS Genet.* **9**, e1003830 (2013).
24. G. Li, M. Levitus, C. Bustamante, J. Widom, *Nat. Struct. Mol. Biol.* **12**, 46–53 (2005).
25. K. Kobayashi *et al*., *Sci. Rep.* **7**, 11772 (2017).
26. A. V. Gasparian *et al*., *Sci. Transl. Med.* **3**, 95ra74 (2011).
27. D. D. Winkler, K. Luger, *J. Biol. Chem.* **286**, 18369–18374 (2011).
28. F. Jin, Y. Li, B. Ren, R. Natarajan, *Proc. Natl. Acad. Sci. U.S.A.* **108**, 5290–5295 (2011).
29. C. A. Davey, D. F. Sargent, K. Luger, A. W. Maeder, T. J. Richmond, *J. Mol. Biol.* **319**, 1097–1113 (2002).
30. H. S. Tims, K. Gurunathan, M. Levitus, J. Widom, *J. Mol. Biol.* **411**, 430–448 (2011).
31. A. Hoffmann, A. Levchenko, M. L. Scott, D. Baltimore, *Science* **298**, 1241–1245 (2002).
32. A. A. Beg, W. C. Sha, R. T. Bronson, D. Baltimore, *Genes Dev.* **9**, 2736–2746 (1995).
33. V. F.-S. Shih *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **106**, 9619–9624 (2009).
34. B. Peng *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **107**, 15193–15198 (2010).
35. M. R. Corces *et al*., *Science* **362**, eaav1898 (2018).
36. P. J. Murray *et al*., *Immunity* **41**, 14–20 (2014).
37. J. E. Kanter *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **109**, E715–E724 (2012).
38. A. Soufi *et al*., *Cell* **161**, 555–568 (2015).
39. D. Barken *et al*., *Science* **308**, 52 (2005).
40. C. S. Cheng *et al*., *Cell Syst.* **4**, 330–343.e5 (2017).
41. Signaling Systems Lab, signalingsystemslab/NFkB_enhancer_dynamics: First release bioinfo analysis and nucleosome model, Version 1.0, Zenodo (2021); http://doi.org/10.5281/zenodo.4698447.

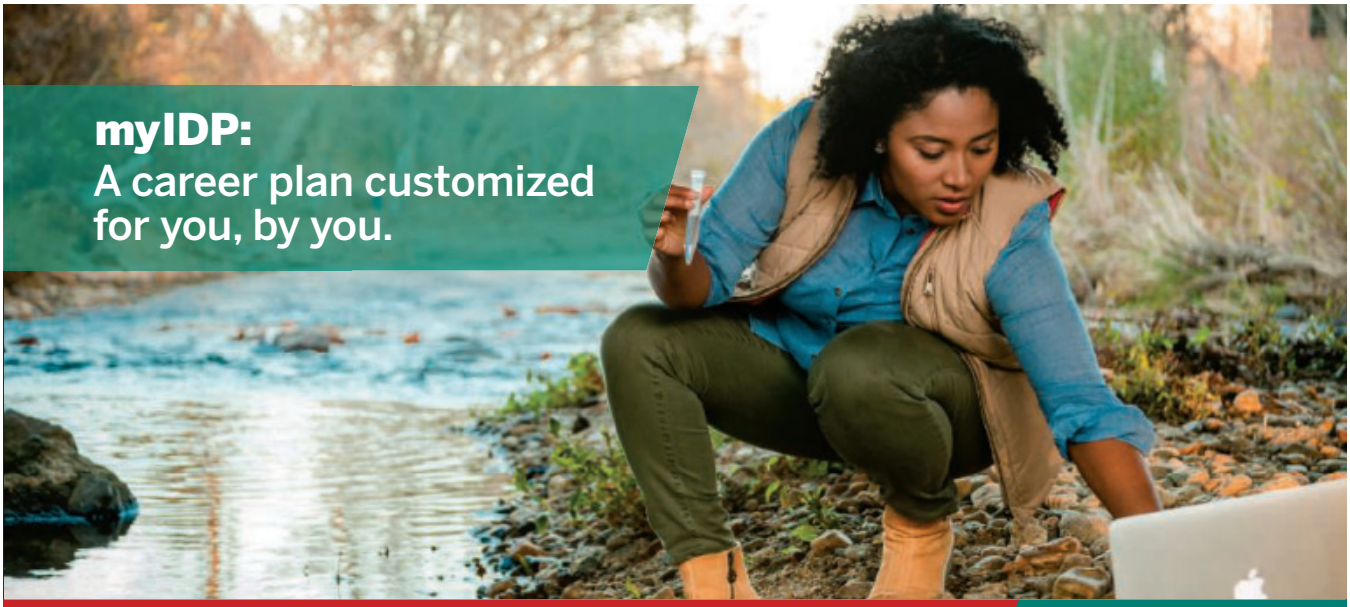
ACKNOWLEDGMENTS

We thank K. Kishimoto, D. Lefaudeux, X.-F. Lin, and A. Mazumber for their experimental and analytical contributions, and S. Kurdistani, M. Carey, E. Lin, and Y. Tang for their insights and critical reading of the manuscript. **Funding:** This work was supported by NIH grants R01-AI127864 (to A.H.), R01-GM117134 (to A.H.), F31-AI138450 (to A.A.), T32-GM008042 (to K.M.S.), T32-HL069766 (to A.A.), and T32-AI089398 (to Q.J.C.); the Specialty Training and Advanced Research (STAR) program of the UCLA Department of Medicine (to Q.J.C.); and the QCB Collaboratory for a postdoctoral fellowship (to R.S.). Sequencing was performed at the UCLA Broad Stem Cell Center Sequencing Core. **Author contributions:** Q.J.C., S.O., A.A., and B.T. performed the experiments. Q.J.C., S.O., K.M.S., R.S., and A.A. analyzed the data. K.S. and B.T. developed the mathematical model. Q.J.C., K.M.S., and A.H. wrote the manuscript. All authors reviewed the manuscript. A.H. coordinated and funded the work. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** Raw data and count tables for ChIP-seq, ATAC-seq, and RNA-seq data have been submitted to the NCBI GEO repository under accession number GSE146068. Code is available at https://github.com/signalingsystemslab/NFkB_enhancer_dynamics and Zenodo (*41*).

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/372/6548/1349/suppl/DC1
Materials and Methods
Figs. S1 to S9
References (*42*–*60*)
MDAR Reproducibility Checklist

1 April 2020; accepted 22 April 2021
10.1126/science.abc0269



Director of the Institut Pasteur

The Search Committee for Director of the Institut Pasteur invites expressions of interest from potential candidates for the position of Director of the institute

Since its creation in 1887, the Institut Pasteur has been a world leader in fundamental biology and biomedical research, as well as a renowned teaching center. The Institut Pasteur is an independent private non-profit foundation for basic biological research, research into biomedicine and public health.
With a history of excellence in research on infectious diseases and major discoveries in microbiology, virology, immunology and vaccines, it also has made a significant commitment to cross-disciplinary approaches and has extended its focus to other biomedical fields.
The Director will join an exciting and innovative academic environment of over 2,800 people, on the campus in the heart of Paris, and at the center of an international network.

The Director should be a scientific personality with a world-class reputation. She/he will provide the executive leadership for the institute. He/she will develop an ambitious and strategic vision for the scientific activities to take the Institut Pasteur into the 21st century. She/he will be responsible for the management of the institute's scientific mission management, the successful translation of basic science into improvement of human health; for fund-raising and diversification of resources; for overall management of the institute; and for forging and overseeing university and industrial partnerships. The Director is also responsible for developing and reinforcing the cooperation among the 33 members comprising its unique international network.
Both French and English language skills are required. The Institut Pasteur is committed to internationality, diversity and equality, and encourages applications from women and minorities. The starting date of the position is January 2024.

The committee, chaired by Edith Heard, General Director of the EMBL, encourages expressions of interest from scientists of the highest international standing irrespective of nationality or experience working in France.

The Search Committee for Director of the Institut Pasteur welcomes letters of interest accompanied by curriculum vitae. Please send all letters and other inquiries by e-mail to **searchcommittee2021@pasteur.fr.**

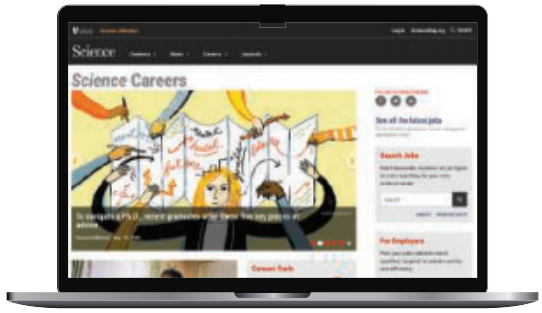
All submissions will be kept strictly confidential. Additional information regarding the recruiting process can be found at **pasteur.fr/en/searchcommittee2021/**

INSTITUT PASTEUR

Science Careers helps you advance your career. Learn how!

- Register for a free online account on **ScienceCareers.org**.
- Search hundreds of job postings and find your perfect job.
- Sign up to receive e-mail alerts about job postings that match your criteria.
- Upload your resume into our database and connect with employers.
- Watch one of our many webinars on different career topics such as job searching, networking, and more.
- Download our career booklets, including Career Basics, Careers Beyond the Bench, and Developing Your Skills.
- Complete an interactive, personalized career plan at "my IDP."
- Visit our Employer Profiles to learn more about prospective employers.
- Read relevant career advice articles from our library of thousands.

Visit **ScienceCareers.org** today — all resources are free



SCIENCECAREERS.ORG

ScienceCareers
FROM THE JOURNAL SCIENCE AAAS

By **Brandon Brown** *and* **Annie Lu Nguyen**

In case of death

It started with an attempt at humor during a team meeting in April. Brandon, the principal investigator (PI), expressed gratitude for the work of a team member, Chris, then bemoaned the big shoes that would be left to fill if Chris was hit by a bus. Not long before, a friend and colleague had died suddenly, and we had seen the people who worked most closely with him worry and wonder, in the midst of their grief, who would continue the research, advise the graduate students, track down the data files, and lead the national coordinating center. We had both taken steps toward personal end-of-life planning since the onset of the pandemic, and we realized it could be relevant to our work lives, too. But it wasn't until our team meeting that the reality really hit home. "As the PI, I have details and documents no one else has access to," Brandon realized. "What happens if I get hit by a bus?"

After that team meeting, we sprang into action. Annie, a co-investigator on the project who studies end-of-life planning, looked into the literature for best practices to plan for a PI's sudden death—and came up empty. So we had to develop our own plans. Brandon made sure all the study documents, including the grant application, institutional review board protocols, contracts, and budgets, were backed up and stored in a way that is secure but accessible in case of his absence. He asked the financial administrators at his university and the grantor about what happens to the funding if he dies. His inquiries often yielded more questions than answers, as the administrators had not thought much about this question before, but we hoped to set the stage for some resolution down the line. Brandon even mapped out the percent effort paid to each team member on the project and how his own payment would be distributed in his absence.

"It is important and worthwhile to lay this groundwork ... should the unexpected happen."

At some points, wrangling these details was frustrating and almost felt like a waste of time to Brandon, when he could have been working on something more immediately "productive" related to his scholarly work—especially given that he is a midcareer PI and hopefully far from death. But he reminded himself it is important and worthwhile to lay this groundwork to help his colleagues should the unexpected happen.

Finally, we had to address perhaps the most important question: Who would lead the study in Brandon's absence? This would be a significant request, almost an academic marriage proposal. Still, Brandon approached it with a bit of levity, sending his request to Annie in an email headed "in case of ☠." On one level, this was banter between friends, and Annie immediately agreed that she would step in. But it was more than just words. She felt ready to take on the responsibility if needed and honored that Brandon trusted her to carry on the project. She told Brandon that if something were to happen to him, she would lead the study in his honor and memory. She has started to practice for the role, running a few meetings as PI to establish a precedent and make a transition—if it proves necessary—as seamless as possible.

Every PI can enact plans to ensure continuity in their absence. Think about who will lead the project and make major decisions if you are unable to. Consider naming the proxy PI on documents, including contracts, manuscripts, funding agreements, and research ethics applications, under the title "proxy PI in case of PI death or incapacitation." Other questions include, what is your biggest wish or vision for this project? On papers published postmortem, do you want to be included as an author or acknowledged? Where should the royalties due to you from a project or discovery be directed?

Contemplating and talking about our own deaths can be uncomfortable. But by planning ahead, we find peace in knowing that we are doing our best to help our life's work continue beyond our physical presence. ■

Brandon Brown is an associate professor at the University of California, Riverside, School of Medicine. Annie Lu Nguyen is an assistant professor at the University of Southern California Keck School of Medicine. Send your career story to SciCareerEditor@aaas.org.

ILLUSTRATION: ROBERT NEUBECKER

CALL FOR PAPERS

SPJ
SCIENCE PARTNER JOURNALS

北京理工大學
BEIJING INSTITUTE OF TECHNOLOGY

中国空间技术研究院
China Academy of Space Technology
中国航天

Space: Science & Technology

Space: Science & Technology is an online-only, Open Access journal published in affiliation with **Beijing Institute of Technology (BIT)** and distributed by the **American Association for the Advancement of Science (AAAS)**. BIT cooperates with China Academy of Space Technology (CAST) in managing the journal. The journal aims to promote the exploration and research of space worldwide, to lead the rapid integration and technological breakthroughs of interdisciplinary sciences in the space field, and to build a high-level academic platform for discussion, cooperation, technological progress and information dissemination among professional researchers, engineers, scientists and scholars.

Submit your research to *Space: Science & Technology* today!

Learn more at **spj.sciencemag.org/space**

The Science Partner Journal (SPJ) program was established by the American Association for the Advancement of Science (AAAS), the nonprofit publisher of the *Science* family of journals. The SPJ program features high-quality, online-only, Open Access publications produced in collaboration with international research institutions, foundations, funders and societies. Through these collaborations, AAAS furthers its mission to communicate science broadly and for the benefit of all people by providing top-tier international research organizations with the technology, visibility, and publishing expertise that AAAS is uniquely positioned to offer as the world's largest general science membership society.

Visit us at **spj.sciencemag.org**

@SPJournals **@SPJournals**

OPEN ACCESS

ARTICLE PROCESSING CHARGES WAIVED UNTIL JULY 2023