13 February 2009 | $10

Science

SELENE
and the Lunar Farside

AAAS

EDITORIAL

NEWS OF THE WEEK

NEWS FOCUS

LETTERS

BOOKS *ET AL.*

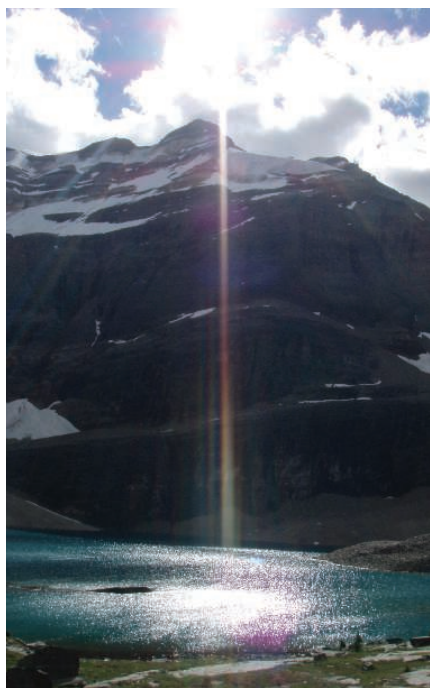
POLICY FORUM

PERSPECTIVES

REVIEW

CONTENTS continued >>

page 866



page 887

COVER

An image of Mare Moscoviense on the lunar farside, taken by the Terrain Camera aboard the SELENE (Kaguya) spacecraft on 17 January 2008. Crater frequency distributions derived from such images yield ages of about 2.5 billion years for several volcanic deposits in this mare, indicating that volcanism on the lunar farside lasted about 500 million years longer than previously thought. See page 905.

Image: Japan Aerospace Exploration Agency/SELENE

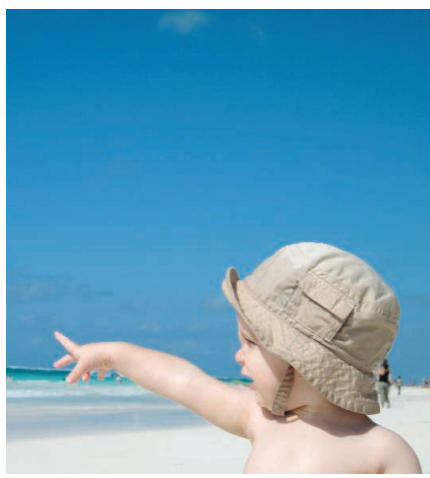
DEPARTMENTS

Published by AAAS

BREVIA

REPORTS

page 879



page 951

CONTENTS continued >>

Published by AAAS

SCIENCEONLINE

SCIENCEXPRESS
www.sciencexpress.org

Sequencing and Analyses of All Known Human Rhinovirus Genomes Reveals Structure and Evolution
A. C. Palmenberg et al.

Rhinoviruses, which are a common cause of colds, fall into three or possibly four major clades and can undergo strain recombination.
10.1126/science.1165557

The Dynamic Control of Kiss-And-Run and Vesicular Reuse Probed with Single Nanoparticles
Q. Zhang et al.

A transient form of vesicle fusion that allows vesicle reuse is more prevalent early during synaptic response to a train of stimuli.
10.1126/science.1167373

Genome-Wide Analysis in Vivo of Translation with Nucleotide Resolution Using Ribosome Profiling
N. T. Ingolia et al.

Profiling the position of ribosomes on messenger RNA allows rapid, high-precision investigation of cellular protein translation.
10.1126/science.1168978

Omnidirectional Printing of Flexible, Stretchable, and Spanning Silver Microelectrodes
B. Y. Ahn et al.

Colloidal silver particles can be formed into flexible electrodes of arbitrary shape in three dimensions.
10.1126/science.1168375

TECHNICALCOMMENTS

Comment on "Dynamic Shifts of Limited Working Memory Resources in Human Vision"
N. Cowan and J. N. Rouder

full text at www.sciencemag.org/cgi/content/full/323/5916/877c

Response to Comment on "Dynamic Shifts of Limited Working Memory Resources in Human Vision"
P. M. Bays and M. Husain

full text at www.sciencemag.org/cgi/content/full/323/5916/877d

SCIENCENOW
www.sciencenow.org
Highlights From Our Daily News Coverage

Novel HIV-Fighting Method Finally Gels
Study reports first "success" in preventing HIV infections with vaginal microbicide.

One Giant Leap for Robot-Kind
Cockroach-inspired design could help Mars rovers and other robots traverse sandy terrain.

Flu: It's the Humidity. Absolutely
Study suggests that absolute, not relative, humidity explains why influenza is seasonal.

SCIENCESIGNALING
www.sciencesignaling.org
The Signal Transduction Knowledge Environment

RESEARCH ARTICLE: Brx Mediates the Response of Lymphocytes to Osmotic Stress Through the Activation of NFAT5
T. Kino et al.

The guanine nucleotide exchange factor Brx is essential for osmotic stress–mediated expression of *nfat5* in lymphocytes.

MEETING REPORT: Emerging Roles of NAD^+ and Its Metabolites in Cell Signaling
F. Koch-Nolte et al.

Scientists discussed the molecular aspects of NAD^+ metabolism and signaling at a conference in Hamburg, Germany.

NETWATCH: Simmune Project
Build models, run simulations, and explore the dynamics of signaling pathways; in Modeling Tools.

NETWATCH: American Physiological Society Archive
Browse or search an index of peer-reviewed science education materials; in Educator Sites.

SCIENCECAREERS
www.sciencecareers.org/career_magazine
Free Career Resources for Scientists

The Man Who Wasn't There
B. L. Benderly

Douglas Prasher's story reveals a lot about the scientific labor market.

Discouraging Days for Job-Seekers
S. Carpenter

Many universities and colleges have canceled faculty searches or imposed hiring freezes.

SCIENCEPODCAST
www.sciencemag.org/multimedia/podcast
Free Weekly Show

Download the 13 February *Science* Podcast to hear about how toddler gesturing may improve children's vocabulary, a new global view of the Moon, an update on the Neandertal genome, and more.

ORIGINSBLOG
blogs.sciencemag.org/origins
A History of Beginnings

SCIENCEINSIDER
blogs.sciencemag.org/scienceinsider
Science Policy News and Analysis

SCIENCESIGNALING
Signaling researchers met in Hamburg.

SCIENCECAREERS
Bleak campus job prospects.

CREDITS: (*SCIENCE SIGNALING*) KEVIN FOREST/GETTY IMAGES; (*SCIENCE CAREERS*) MATHIEU PLOURDE/CREATIVE COMMONS

SCIENCE (ISSN 0036-8075) is published weekly on Friday, except the 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. The title SCIENCE is a registered trademark of the AAAS. Domestic individual membership and subscription (51 issues): $146 ($74 allocated to subscription). Domestic institutional subscription (51 issues): $835; Foreign postage extra: Mexico, Caribbean (surface mail) $55; other countries (air assist delivery) $85. First class, airmail, student, and emeritus rates on request. Canadian rates with GST available upon request, GST #1254 88122. 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:** $10.00 current issue, $15.00 back issue prepaid includes surface postage; bulk rates on request. **Authorization to photocopy** material for internal or personal use under circumstances not falling within the fair use provisions of the Copyright Act is granted by AAAS to libraries and other users registered with the Copyright Clearance Center (CCC) Transactional Reporting Service, provided that $20.00 per article is paid directly to CCC, 222 Rosewood Drive, Danvers, MA 01923. The identification code for *Science* is 0036-8075. *Science* is indexed in the *Reader's Guide to Periodical Literature* and in several specialized indexes.

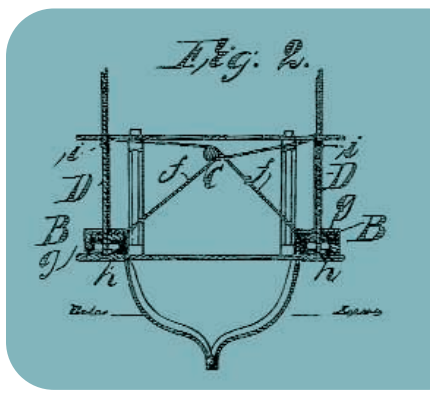
AAAS
ADVANCING SCIENCE, SERVING SOCIETY

James J. McCarthy is the Alexander Agassiz Professor of Biological Oceanography at Harvard University and the president of AAAS.

Presidents Who Value Science

THE OPENING DAY, 12 FEBRUARY, OF THE 2009 ANNUAL MEETING OF THE AMERICAN ASSOCIATION for the Advancement of Science (AAAS) in Chicago is by coincidence the 200th anniversary of the births of Charles Darwin and Abraham Lincoln. New understanding of the underlying processes and mechanisms that fascinated Darwin continues to be the focus of intense research, and many of the papers and symposia prepared for this meeting will celebrate the enduring legacy of Darwin's thesis. It is noteworthy that this meeting is being convened in the land of Lincoln and within a month of the inauguration of another son of Illinois, Barack Obama, as president of the United States.

President Obama assumes his office at a time when exceptional leadership is desperately needed to address urgent domestic and international problems. Unfortunately, the urgency of these and other national priorities does not allow for sequential planning and action, with an expectation that some can be put off for a later time. Over the past few years, scientists have expressed dismay at the lack of opportunity to bring the very best science into the decision-making processes of the U.S. government. The government's scientific integrity was publicly called into question. But the good news is that as the Obama administration coalesces, there are signs of a dramatic shift in the position of science advice. Much like Lincoln, President Obama exhibits intense curiosity and a willingness to listen. Perhaps never before has a president successfully recruited so many scientific stars to his cabinet and other executive positions. We can now hope that their energy and talents will be applied effectively to matters relating to national security and nuclear proliferation, energy security and greenhouse gas reductions, medical research and health care delivery, and more.



Lincoln's presidency was so dominated by the Civil War that his accomplishments in promoting science and advancing the findings of science for the betterment of society are sometimes overlooked. Some of his interest in technology was related to the improvement of weaponry, but his fascination with science and technology had deep roots. In 1849, for example, he patented an invention to lift riverboats over shoals. In 1858, he gave a lecture on discoveries and inventions in which he pondered greater use of wind power, predicting that "quite possibly one of the greatest discoveries hereafter to be made, will be the taming, and harnessing of the wind."* He also brought the telegraph into the White House and became the first national leader to have nearly instantaneous communication with his field commanders.

Lincoln was also keenly interested in improving the efficiency of agriculture and knew that in addition to mechanized systems for planting and harvesting, a better understanding of what we today refer to as soil science and plant ecology was also essential. He developed these ideas in a long address to the Wisconsin State Agricultural Society a year before he was elected president. Recognizing that agricultural productivity could be enhanced by research and wide dissemination of this knowledge, he signed into law a bill creating the Department of Agriculture in 1862. At that time, German scientists were leaders in the study of crop nutrition, and the first scientist hired at the new department was a student of the German school. That same year, Lincoln also signed into law the Land Grant College Act, intended to bring research and teaching on agriculture into colleges. It is also noteworthy that Lincoln's signature created the National Academy of Sciences in 1863, only hours after Congress voted in favor of the bill.

President Lincoln presided over this nation during its worst period of internal turmoil. But this did not distract him from promoting the pursuit of science and from using this knowledge in forming national policy. We can hope that when history is written for the current era, the Obama presidency will be noted as another period of great advancement for "science, engineering, and innovation throughout the world for the benefit of all people."**

– James J. McCarthy

10.1126/science.1171006

*Abraham Lincoln, *Discoveries and Inventions: A Lecture by Abraham Lincoln* (call no. L2 L7367, Lincolniana collection, Abraham Lincoln Presidential Library, Springfield, IL, 6 April 1858).

AAAS Mission Statement, *Science* **323, 856 (2009).

Published by AAAS

www.sciencemag.org

Science

1200 New York Avenue, NW
Washington, DC 20005
Editorial: 202-326-6550, FAX 202-289-7562
News: 202-326-6581, FAX 202-371-9227
Bateman House, 82-88 Hills Road
Cambridge, UK CB2 1LQ
+44 (0) 1223 326500, FAX +44 (0) 1223 326501

Subscription Services For change of address, missing issues, new orders and renewals, and payment questions: 866-434-AAAS (2227) or 202-326-6417, FAX 202-842-1065. Mailing addresses: AAAS, P.O. Box 96178, Washington, DC 20090-6178 or AAAS Member Services, 1200 New York Avenue, NW, Washington, DC 20005

Institutional Site Licenses please call 202-326-6755 for any questions or information

Reprints: Author Inquiries 800-635-7181
Commercial Inquiries 803-359-4578

Permissions 202-326-7074, FAX 202-682-0816

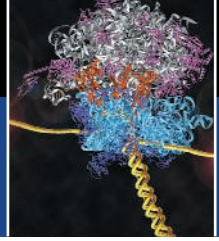
Member Benefits AAAS/Barnes&Noble.com bookstore www.aaas.org/bn; AAAS Online Store www.apisource.com/aaas/ code MKB6; AAAS Travels: Betchart Expeditions 800-252-4910; Apple Store www.apple/eppstore/aaas; Bank of America MasterCard 1-800-833-6262 priority code FAA3YU; Cold Spring Harbor Laboratory Press Publications www.cshlpress.com/affiliates/aaas.htm; GEICO Auto Insurance www.geico.com/landingpage/go51.htm?logo=17624; Hertz 800-654-2200 CDP#343457; Office Depot https://bsd.officedepot.com/portalLogin.do; Seabury & Smith Life Insurance 800-424-9883; Subaru VIP Program 202-326-6417; VIP Moving Services www.vipmayflower.com/domestic/index.html; Other Benefits: AAAS Member Services 202-326-6417 or www.aaasmember.org.

science_editors@aaas.org	(for general editorial queries)
science_letters@aaas.org	(for queries about letters)
science_reviews@aaas.org	(for returning manuscript reviews)
science_bookrevs@aaas.org	(for book review queries)

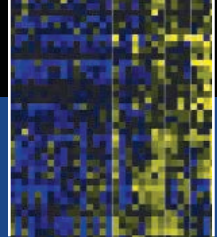
Published by the American Association for the Advancement of Science (AAAS), *Science* serves its readers as a forum for the presentation and discussion of important issues related to the advancement of science, including the presentation of minority or conflicting points of view, rather than by publishing only material on which a consensus has been reached. Accordingly, all articles published in *Science*—including editorials, news and comment, and book reviews—are signed and reflect the individual views of the authors and not official points of view adopted by AAAS or the institutions with which the authors are affiliated.

AAAS was founded in 1848 and incorporated in 1874. Its mission is to advance science, engineering, and innovation throughout the world for the benefit of all people. The goals of the association are to: enhance communication among scientists, engineers, and the public; promote and defend the integrity of science and its use; strengthen support for the science and technology enterprise; provide a voice for science on societal issues; promote the responsible use of science in public policy; strengthen and diversify the science and technology workforce; foster education in science and technology for everyone; increase public engagement with science and technology; and advance international cooperation in science.

Information for Authors
See pages 634 and 635 of the 1 February 2008 issue or access www.sciencemag.org/about/authors



Protein production in action
864



New prostate cancer marker
865

PUBLIC HEALTH

International Groups Battle Cholera in Zimbabwe

Zimbabwe's severe cholera outbreak has devolved in recent weeks into what one international health official calls "an extraordinary public health crisis." The disease is starting to spill over into adjacent nations and threatens to become endemic in Zimbabwe. The World Health Organization (WHO) has taken the unusual step of setting up a Cholera Command and Control Centre in the capital of Harare to coordinate an array of international groups, including United Nations agencies, which are working around the clock to shore up health services, distribute medications, and treat water.

Calling Zimbabwe's once-proud public health system "an empty shell," WHO's Assistant Director-General for Health Security and Environment, David L. Heymann, told *Science* that the nation's public health infrastructure is so weak after years of neglect that WHO is concerned that other infectious diseases may emerge and that the epidemics of HIV/AIDS and drug-resistant tuberculosis (TB) may worsen.

Since August, the cholera outbreak in Zimbabwe has killed some 3400 people and infected an estimated 68,000. Some experts predict that the case total will top 100,000, meaning that cholera would have struck nearly one in every 100 Zimbabweans. Lesser outbreaks are spreading across the borders into Zambia, South Africa, and Mozambique.

In Zimbabwe, outside experts say, the outbreak resulted from the breakdown of potable water and sewage systems—a symptom of the country's economic chaos. The crisis worsened in part because the public health system—devastated by the loss of so many doctors and nurses who fled the country to make a living elsewhere—was severely understaffed and underfunded. Rather than dispensing the recommended oral rehydration salts (ORS), the government initially encouraged people with cholera to rehydrate themselves at home by drinking a solution of salt and sugar—an ineffective response because many could not afford the ingredients.

Zimbabwe's health ministry has refocused its efforts, and WHO and other international groups are helping with several key activities: distributing ORS and chlorine tablets, finding funds to pay thousands of Zimbabwean health care workers, and providing better services in remote areas. "We are sending more teams into the provinces now," says Dominique Legros, a medical epidemiologist with WHO's Disease Control in Humanitarian Emergencies program, who helped put together the cholera partnership in Zimbabwe. Although the overall case mortality rate remains an unacceptable 5%, Legros says the weekly mortality rate has begun to fall, recently dropping below 4%. The target in cholera outbreaks is to get below 1%.

Breakdown. Its health system in tatters, Zimbabwe has been hit with a devastating cholera epidemic that has taken a heavy toll in remote areas.

CREDIT: REUTERS/PHILIMON BULAWAYO

Vaccines have not been a factor. The only approved cholera vaccine available—the Swedish-produced Dukoral oral vaccine—must be kept cool and administered twice with at least a week's delay in between, which would be difficult in remote areas. There are two other cholera vaccines. But the newest, made in Vietnam, has not yet gained WHO approval, and an older one, developed by James B. Kaper of the University of Maryland's medical school and Myron M. Levine, director of the university's Center for Vaccine Development, went out of production in 2004. Levine says, "We are now trying to get the vaccine back into production."

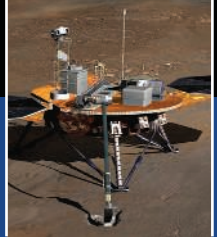
It is highly unusual to find "cholera breaking out so widespread in a country, involving more than 85% of the districts," says Pradip Bardhan, a cholera specialist who led a team from the Bangladesh-based International Centre for Diarrhoeal Disease Research to investigate the Zimbabwe outbreak. Heymann said it is ironic that cholera could take hold in a country that once had one of Africa's best health infrastructures, including a first-rate public health research center, the University of Zimbabwe's (UZ's) Blair Research Institute. Heymann says the center "has pretty much dried up."

UZ pharmacologist Chiedza Maponga agrees that there is a "shortage of health personnel, not just physicians but environmental health practitioners and nurses." The number of doctors may have dropped to fewer than half the official level of about 2000 (*Science*, 4 May 2007, p. 684) in a nation of 12 million, which may also have been reduced to about 10 million due to a mass exodus.

Epidemiologist Chris Beyrer of the Johns Hopkins Bloomberg School of Public Health

Published by AAAS

A draft Neandertal genome
866



Phoenix's mixed record
872

in Baltimore, Maryland, who recently visited Zimbabwe with a Physicians for Human Rights delegation, says he was appalled by the country's public health system and deeply concerned about the nation's HIV/AIDS and TB epidemics, as well as the cholera crisis. "One clinic we visited had not seen a physician since 2006," Beyrer says, "and many health workers had not been paid in months." He believes the international community should consider taking over the public health system "in a kind of receivership."

WHO estimates that about 120,000 people died of cholera worldwide in 2007 and that millions more were infected. These figures are much higher than the officially reported numbers, which in 2007 were 4031 deaths and 178,000 cholera cases. The current Zimbabwean outbreak is one of the most severe in recent southern African history. An outbreak in Angola in 2006–07 led to 85,000 cases, with a mortality rate of about 4%. A cholera outbreak in South Africa in 2002 registered 116,000 cases, with a mortality rate of less than 1%.

Cholera had virtually disappeared from Zimbabwe over a decades-long period, even though *Vibrio cholerae* has been found in plankton living in the Zambezi River system. Although the pathogen has emerged several times since 1998 to infect people, health officials were able to control the minor outbreaks within a few weeks. But Legros says, "I am afraid this will not be the case this time and that it will remain endemic" in the population.

–ROBERT KOENIG

BIODEFENSE

Army Halts Work at Lab After Finding Untracked Material

The U.S. Army Medical Research Institute of Infectious Diseases (USAMRIID) has suspended research involving select agents and toxins after a spot inspection of the Frederick, Maryland, facility found four vials of Venezuelan equine encephalitis (VEE) that weren't in its electronic database. Some 350 researchers and technicians are affected by the stoppage, which began on 6 February and could last 3 months or longer as officials open up every freezer and refrigerator to take stock of all hazardous biomaterials at the institute.

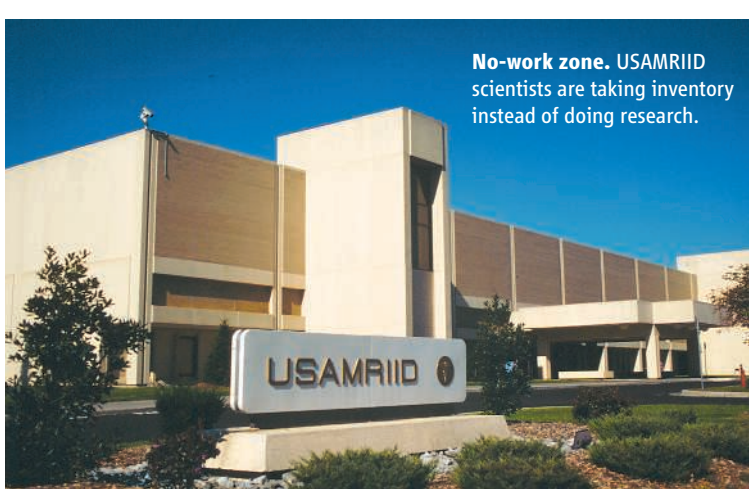
The lab, the largest U.S. defense facility for work on deadly pathogens, has been under intense scrutiny since the Federal Bureau of Investigation (FBI) named former USAMRIID researcher Bruce Ivins as the perpetrator of the 2001 anthrax letter attacks (*Science*, 8 August 2008, p. 754). Ivins committed suicide on 29 July 2008, but FBI officials have argued that the evidence against him is beyond doubt and that he carried out the mailings using anthrax stolen from a flask at USAMRIID. A special task force has spent the past few months considering what new measures are needed to improve security at USAMRIID and other Army biodefense labs.

Last month, all lab administrators were ordered to file a Serious Incident Report if they discovered materials that did not have a corresponding record in the lab's computerized inventory. Traditionally, such materials would have been added to the inventory without any disruption in the workflow. "In the past, we would have entered the [four VEE] vials into our database and moved on," USAMRIID deputy commander Mark Kortepeter told *Science*. But the new requirement prompted institute managers to "reexamine how we do inventory," he says. Officials realized that "if we can find four vials in a spot check, it is possible there are other vials in our freezers that are not accounted for in our database."

In a 4 February memo to employees, USAMRIID commander John Skvorak explained that all work with select agents would be suspended until he received "certification that the full contents of each freezer and refrigerator have been evaluated and that all BSAT [biological select agents and toxins] is included in our inventory." Some critical animal experiments will be allowed to continue.

Kortepeter says officials expect to find materials in addition to those listed in the database because of human errors made when the lab switched from a paper to a computerized system in 2005. He says the four vials of VEE were part of a set of 20 and that 16 had been entered into the system. "It was simply an oversight," he says.

Greg Koblenz, a biosecurity expert at George Mason University in Fairfax, Virginia, believes that the work stoppage is justified. "Not having an accurate and complete baseline inventory makes it difficult to ensure that no material has been diverted," he says. Koblenz says it's reassuring that high-level Army officials are trying to improve the system.



No-work zone. USAMRIID scientists are taking inventory instead of doing research.

CREDIT: USAMRIID/GETTY IMAGES

Kortepeter says there are no plans for USAMRIID to investigate the possible loss or theft of previously undocumented materials. Other security measures at the institute would have detected any such incident, he says, even with an incomplete inventory.

Researchers affected by the new rules are understandably unhappy, says Kortepeter, but they "realize the importance" of carrying out the inventory. Many of them will help with the task, he says.

–YUDHIJIT BHATTACHARJEE

Published by AAAS

FORENSIC SCIENCE

Journal Flinches as Article on Voice Analyzer Sparks Lawsuit Threat

A peer-reviewed journal recently yanked a long-since published paper from its Web site after the makers of a voice-analysis system—which is sold as a device to detect emotional stress and help ferret out liars—complained that the article contained inaccuracies and defamed them. The authors of the review article, two scientists from Sweden who normally study the sounds of speech, complain that the company is attempting to stifle free inquiry. The company founder counters that the paper was less a scientific analysis of his product than a personal attack. Meanwhile, 25 local governments in the United Kingdom are already using the controversial technology to try to weed out fraud among people applying for public assistance, and its use may be extended nationwide.

Penned by phoneticians Anders Eriksson of the University of Gothenburg and Francisco Lacerda of Stockholm University, the controversial paper appeared in the December 2007 issue of *The International Journal of Speech, Language and the Law* with the title "Charlatanry in forensic speech science: A problem to be taken seriously." The paper argues that there is no scientific basis for techniques that aim to determine emotional stress by analyzing the sound of a person's voice. In particular, Eriksson and Lacerda take aim at the Layered Voice Analysis (LVA) systems made by Nemesysco Ltd. of Netanya, Israel, of which they write, "The ideas on which the products are based are simply complete nonsense."

Amir Liberman, the founder and CEO of Nemesysco, took exception to the paper. On 3 November, his lawyers wrote to the journal's publisher, Equinox Publishing Ltd. of London, threatening to sue for defamation if the article was not retracted. Liberman says he acted not because the researchers questioned his technology but because they targeted him personally by, for example, noting in a section titled "Who is Mr. Liberman?" that he has no university degree. "The objection was not in the publication of their study results, it was in their calling us charlatans," Liberman says.

On 4 December, the journal removed the article from its subscription-based Web site and posted a notice that acknowledged that "Mr. Liberman and Nemesysco Limited ... were not invited to comment on the content of the article prior to its publication where, in view of the content of the article, it would have been appropriate to invite them to do so." Janet Joyce, managing director at Equinox, declined to discuss the specifics of the case but says the journal—which is published biannually, has a circulation of fewer than 500, and employs no full-time staff—simply lacks the resources to put up a legal fight. The journal has agreed to publish a rebuttal letter from Lieberman and the company, but Joyce notes that "we didn't withdraw the article. It's still in print."

In this corner. Nemesysco founder Amir Liberman (*left*) says phonetician Francisco Lacerda defamed him.

In the paper, Eriksson and Lacerda cite two studies that question the scientific validity of LVA technology, which analyzes the pressure waves that make up an utterance and tallies spikes, or "thorns," and plateaus in order to deduce the speaker's state of mind. The phoneticians also tried to reproduce the software for the LVA system using computer code in Liberman's 2003 patent as a guide. Lacerda claims that the spikes and plateaus are largely artifacts of the digitization of the sound and that the system makes arbitrary connections between the rates at which they occur and emotional states such as "high stress" or "untruthfulness." "The measurements themselves … don't contain any meaningful information," Lacerda asserts.

Liberman counters that Eriksson and Lacerda used information from only one of three patents and that they never used one of Nemesysco's systems itself. "This attack is being made by people who never saw our technology, never touched the equipment," he says. Yossi Pinkas, vice president for sales and marketing at Nemesysco, says that the company, which employs 11 people, has rigorous studies that show how the technology works, but "we don't necessarily publish everything that we have because we've been ripped off a couple of times."

The debate notwithstanding, the Nemesysco system is already in use. In an effort to stamp out fraud, 25 local government councils in the United Kingdom are using it in conjunction with scripted questions to evaluate phone callers applying for housing and other benefits. The national government will decide in spring 2010 whether to deploy the technology nationwide, says John Stevenson, a spokesperson for the Department of Work and Pensions. Officials use the system to decide whether further investigation is warranted, he says. "They wouldn't just say from one phone call we're stopping the benefit," Stevenson says.

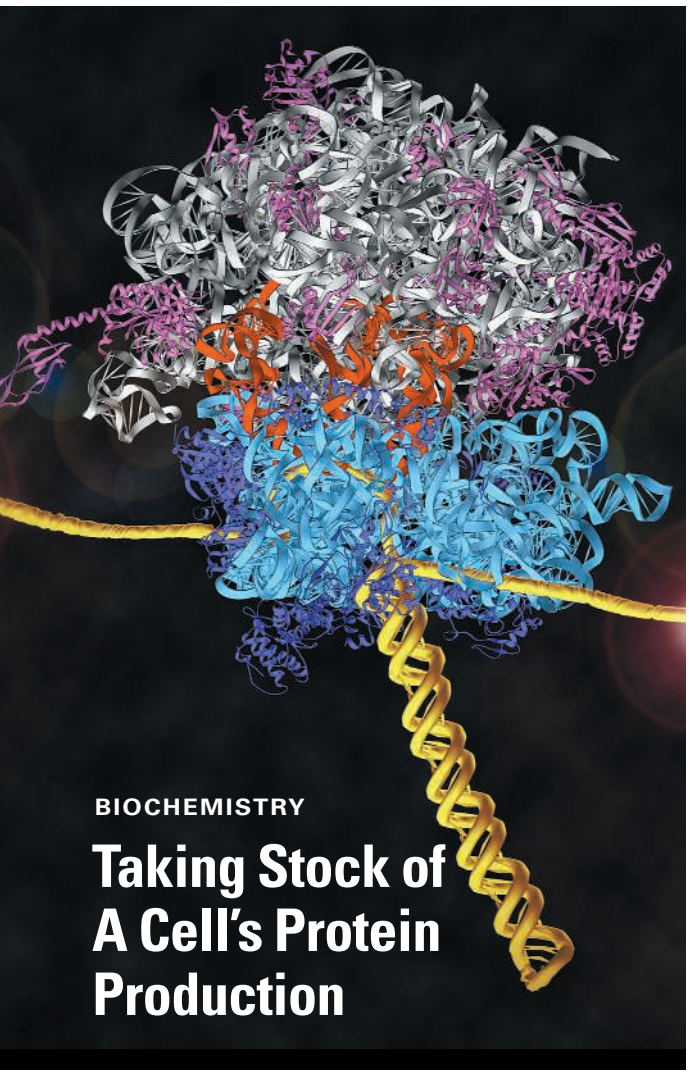
The London Borough of Harrow has used the technology since May 2007. "In our view, it's been a great success," says Harrow spokesperson Fergus Sheppard. "We estimate that we've saved £520,000." From May 2007 to January this year, the system flagged 189 callers as "high risk," and 84 eventually had their benefits reduced, he says. In addition, of 1948 callers flagged "no risk," 474 voluntarily admitted that their original claims were inflated when they were told that they were being evaluated by the system. "They self-policed," Sheppard says. "The system has a deterrent effect."

That's all it has going for it, Lacerda argues. He contends it's unethical for the government to use technology that, in his view, has no scientific basis and essentially deceives people—even if it only tricks would-be cheaters into going straight.

–ADRIAN CHO

CREDITS (LEFT TO RIGHT): A. LIBERMAN/NEMESYSCO LTD.; STOCKHOLM UNIVERSITY/ORASIS FOTO

Published by AAAS



BIOCHEMISTRY

Taking Stock of A Cell's Protein Production

Freeze frame. New technique shows just how efficiently ribosomes translate messenger RNA into proteins.

High-school students learn that cells build themselves in a sort of waltz: Genes encoded in DNA are transcribed into fragments of messenger RNA (mRNA), which are then translated into proteins. But what exactly are they building? Microchip-based RNA arrays can show just which genes have been switched on to produce mRNAs, but not all RNAs get translated into proteins. Researchers would love to have a fast way of tallying all the proteins being synthesized at a given time and how far along each one is. Now that wish may be coming true.

In a paper published online by *Science* this week (www.sciencemag.org/cgi/content/ abstract/1168978), researchers led by Jonathan Weissman, a Howard Hughes Medical Institute investigator at the University of California, San Francisco, report a technique that takes a snapshot of a cell's protein production: each protein's identity, how far along it is in synthesis, and how quickly and efficiently it's being made. The information is far better than RNA arrays at predicting how many proteins will wind up in the cell—information that could reveal which proteins are switched on during cancer and other diseases.

"I'm excited about the work because it gets at the dynamics of gene expression," says Rachel Green, a molecular biologist at Johns Hopkins University in Baltimore, Maryland. The new work "is at a resolution you just couldn't get" with coarser-grained older methods, Green says. In addition to offering insights into diseases, Green and others say the new tool could help developmental biologists understand the earliest stages in the development of life. "It allows you to ask a different set of questions in a global way," says John Yates, a proteomics expert at the Scripps Research Institute in San Diego, California.

Tracking protein synthesis has long been an inexact science. Several years ago, researchers led by Patrick Brown at Stanford School of Medicine in California improved matters by developing a technique for looking at mRNAs as the cellular machines called ribosomes translate them into proteins. The technique, known as polysomal profiling, takes advantage of the fact that multiple ribosomes often work simultaneously along one RNA strand. Researchers lyse cells in the presence of a chemical that stops protein synthesis and glues the ribosomes in place on the mRNAs. Then they sort mRNAs by how many ribosomes are stuck on each one, to estimate how many copies of the protein were being made. Finally, they identify the mRNAs present using RNA microarrays, thereby revealing the genes—and thus the proteins—they encoded.

Polysomal profiling has its limits. For starters, ribosomes sometimes get stuck on regions of the mRNA that don't become part of the protein, or hit sites where they pause. Such regulatory switches can shut down a protein's production before it even starts. To get at these details, Weissman's team, working with yeast cells, first halted protein production as before. The researchers then used a standard enzyme to digest all of the mRNA except snippets buried and protected inside each ribosome. They collected those RNA snippets, converted them to DNA, and ran them through a high-throughput gene-sequencing machine. By comparing the resulting sequences with a database of yeast gene sequences, the team could identify which protein each ribosome was producing and where on the larger mRNA strand it had stopped in its tracks.

The technique is like taking aerial pictures of cars on a highway, Weissman says. Whereas polysomal profiling counts just how many cars are on a stretch of road, the new technique shows exactly where each car is, making it possible to see whether traffic is moving smoothly or is piled up at an accident.

Those details revealed some surprises, Weissman says. For starters, by looking at millions of ribosomes, the researchers found that they could track how efficiently mRNAs were translated into proteins. That's because they knew how many ribosomes were present on each message, how many proteins they wound up producing in the end, and exactly where ribosomal traffic jams were holding things up. The efficiency of protein production, it turns out, is far better than the number of copies of mRNA for predicting how many copies of a protein will be produced.

Weissman's team could also see sharp differences in yeast cells under different conditions. For example, yeast cells that were starved for nutrients started producing proteins at mRNA sites that were different from those of normal cells—a way the protein translation apparatus helps regulate the amount of proteins inside each cell. "We can now look directly at [protein] translation and see a great deal of regulation not seen by microarrays," Weissman says.

The ability to compare protein translation among different cells and tissues could prove a boon to everyone from developmental biologists to diagnosticians, Weissman says. By combining the new technique with known methods for tagging ribosomes in specific tissues, researchers could look for changes in protein translation between animal models with and without breast cancer, for example. It may also prove useful for proteomics researchers by revealing new proteins produced in low abundance, a long-standing challenge for conventional techniques. Certainly, watching the skips and shuffles in biology's three-step is going to be a lot easier.

–ROBERT F. SERVICE

CREDIT: COURTESY OF BERKELEY LAB

BIOMARKERS

Metabolite in Urine May Point To High-Risk Prostate Cancer

Nearly 30,000 men die of prostate cancer in the United States each year, but millions of others who have the disease are not even aware of it. This duality has long frustrated oncologists: They overtreat many men with indolent disease and at the same time fail to catch aggressive cases. "There's been a lot of screening, and people still die," says Ian Thompson, a urologic oncologist at the University of Texas Health Science Center, San Antonio. The community is "desperate," he says, for ways to identify life-threatening tumors.

This week, a team based mainly at the University of Michigan, Ann Arbor, claims to have hit on a novel approach using metabolomics. That mouthful refers to the study of small molecules known as metabolites, chemical products present throughout the body. By screening samples of men's urine, blood, and tissues, the scientists identified 1126 metabolites, then winnowed the list down to six whose levels were higher in samples linked to localized prostate cancer and higher still in metastatic disease. "We were literally going in and measuring everything we could see," says Christopher Beecher, a chemist at the University of Michigan and co-author of the study, led by Arul Chinnaiyan and published this week in *Nature*.

The Michigan group focused on one metabolite in particular, called sarcosine, which may influence cell mobility and which, when added to cells in the lab, transformed normal prostate cells into invasive ones. They hope to devise a simple urine test to pick up the worst prostate cancers at an early stage. Some of the authors have a stake in a company, Metabolon, which is working to commercialize the findings.

It will be years before that can happen, however. The findings need to be extended to a much larger group. This study included just 110 samples each of urine and blood matched with tissue that had been removed from patients. Sixteen had no sign of disease, 12 had localized prostate cancer, and 14 had metastatic cancer.

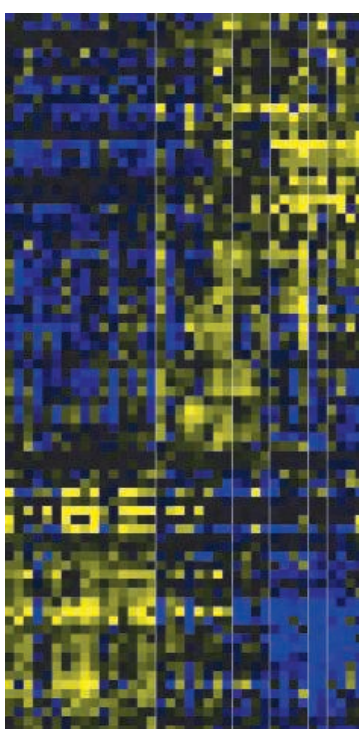
In this case, sarcosine was also "not much better" at picking up aggressive disease than prostate-specific antigen, the protein now widely used in screening, says Cory Abate-Shen, a cancer biologist at Columbia University. Seventy-nine percent of metastatic samples contained high levels of sarcosine, compared with 42% in localized disease and none in healthy tissue. Still, says Abate-Shen, the strategy is unusually promising: "This hasn't been done before for any cancer."

Researchers have poured tremendous energy into early cancer detection and forecasting a cancer's path. Hundreds of studies have been published on the use of gene expression or protein analysis. But few tests have reached the clinic. Even those that have, such as an early-detection test for ovarian cancer called OvaSure, which picks up protein patterns, engender worries that they haven't been adequately validated.

The Michigan work shows potential because it explores a whole new class of particles—metabolites—and because it seeks to distinguish between localized tumors and the more dangerous kind that spread. "It's a terrific way to conduct discovery," says Thompson. But he believes the results must be validated by following many men going forward, tracking those that develop disease and comparing them with those who are healthy. And that will take years.

Patients, Thompson believes, will have little interest in waiting. When the paper's published, "I'll get the usual 100 phone calls saying, 'My biopsy was negative,' " he predicts. " 'Can I come in and get my sarcosine measured?' "

–JENNIFER COUZIN



Chemical clues. Metabolites in three classes of tissue—healthy (blue), localized cancer (yellow), and metastatic disease (red)—show distinct patterns.

CREDIT: SREEKUMAR *ET AL.*, *NATURE* **457**, 912 (2009)

*Science*Insider

From the *Science* Policy Blog

The big news this week was the massive economic stimulus plan as it wended its way through the U.S. Congress. The Senate added $6.5 billion for the National Institutes of Health (NIH) but pared some of the billions the House of Representatives had approved for other science agencies, setting up a round of horse-trading as the House and Senate tried to come up with a final bill. Here's a roundup of other important developments, courtesy of *Science*'s new policy blog, *Science*Insider.

In **European news**, the blog returned this week to escalating tension between the French government and its scientists. Taking their cue from an angry Iraqi journalist, several hundred researchers hurled shoes at the Department of Higher Education and Research in Paris to protest hotly contested reforms. Meanwhile, France will be selling India two nuclear reactors, ending the Asian country's long history of nuclear ostracism from the West.

Elsewhere in Europe, the U.K.'s Royal Society called for the creation of a National Institute of Infectious Diseases to protect the public from bird flu, mad cow, and other maladies. Europe as a whole wants to protect sharks from people. The European Commission tightened regulations on catch limits and shark finning on 5 February.

In Japan, researchers are sounding an alarm about the burdens of noncommunicable diseases, particularly cancer, in Asia. They are working on plans to bring together labs and hospitals throughout Asia into a cancer-research network, and they are looking to the 20th Asia Pacific Cancer Conference, to be held in Tsukuba, Japan, from 12 to 14 November, to give the effort a boost.

In the United States, Senator Charles Grassley (R–IA) has introduced an amendment that would force NIH to keep a tighter leash on its grantees. He wants anyone getting more than $250,000 in grant money to be up front about potential conflicts of interest.

For the full postings and more, go to **blogs.sciencemag.org/scienceinsider**.

Coming alive. Two Neandertal women, perhaps resembling the one reconstructed here, lent their ancient DNA to the genome project.

Tales of a Prehistoric Human Genome

After a mad scramble, researchers have completed a rough draft of a female Neandertal genome, which will offer a new view of *Homo sapiens* as well as our extinct cousins

A half a gram barely tips the postal scale. But from a Neandertal fossil, it's a whopping big chunk of priceless material. Yet Croatian geologist Ivan Gušić readily gave that up on the gamble that 38,000-year-old bones from a cave in northwestern Croatia might help provide a glimpse of the Neandertal genome. This week, researchers announced that Gušić's gamble paid off: They have gotten their first peek at 3 billion bases of Neandertal DNA, and the view, although still hazy, is spectacular. Paleogeneticist Svante Pääbo of the Max Planck Institute for Evolutionary Anthropology in Leipzig, Germany, and his colleagues have compiled a very rough draft of this genome, they reported in a press conference in Leipzig and in a talk at the annual meeting of the American Association for the Advancement of Science (*Science*'s publisher) in Chicago, Illinois, this week.

Because *Homo neanderthalensis* is a member of the human family, much closer kin to us than are chimpanzees, matching Neandertal DNA against our own will reveal genetic changes that define who we are, as well as a different way to be human. Pääbo says he can't wait to finish crunching the sequence through their computers. "We will have in hand hard data on what nucleotide changes occurred in our [lineage]." Those changes are the genetic basis of what makes our species unique, or "what really made modern humans 'modern,' " as paleoanthropologist Jean-Jacques Hublin of Max Planck puts it. Initial comparisons with our own 3 billion bases indicate that a mere 1000 to 2000 amino acid differences, as well as a yet-unknown number of noncoding changes, do that job. For comparison, about 50,000 amino acid differences separate us and chimpanzees.

More than just another genome to add to those of the chimpanzee, macaque, and several modern humans, this ancient genome marks "a philosophical turning point in our understanding of ourselves and our evolution," says Carles Lalueza-Fox, a paleogeneticist at the University of Barcelona in Spain. A feat that 5 years ago would have required many grams of material and many more times the $6.4 million it cost to achieve, the sequencing project "calls attention to the almost unlimited opportunities that modern sequencing methods are creating," adds population geneticist Montgomery Slatkin of the University of California (UC), Berkeley.

The announcement comes more than a decade after Pääbo first demonstrated it was possible to get DNA from fossils of the extinct Neandertals, who lived in Europe from at least 350,000 to about 30,000 years ago. "Only a decade ago, we were hoping to perhaps see the chimpanzee genome someday. At that time, we did not even imagine the possibility of seeing a Neandertal genome," says Ajit Varki of UC San Diego. "It's fascinating to get the first view of the genome of our closest extinct evolutionary cousins, who were so similar and yet so different from us."

The genome is compiled from three shards of limb bone from Vindija Cave that turned out to be from two females. With publication and release of the data expected in the next 6 months, our ability to examine the molecular details of human evolution is poised to explode. So far, the new data suggest that the human and Neandertal

Online
sciencemag.org
Podcast interview with author Elizabeth Pennisi.

Published by AAAS

lineages began to diverge some 800,000 years ago, in line with the most recent estimates from partial genomic data, Pääbo reported at the press conference. Early analyses have yielded no sign of introgression of modern genes into the Neandertal sequence, supporting the idea that Neandertals did not interbreed with modern humans during the thousands of years the two species shared territory in Europe (see p. 870).

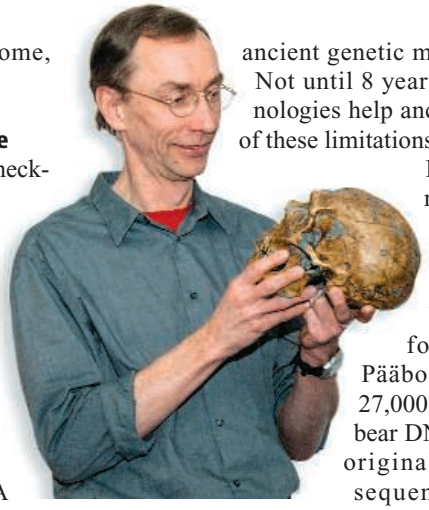
For now, there's not enough sequence to do more than make a rough sketch of Neandertals. "Many gaps and errors are expected," says Lalueza-Fox. To be certain about the order of the bases, sequencers need to determine each base at each location multiple times—and Pääbo has only enough sequence to cover the genome one time over. In total, the team has sequenced 3 billion bases, not quite the length of the whole genome. But some bases were sequenced repeatedly, and some spots were missed entirely. All together, Pääbo estimates they have covered 60% of the entire genome.

"There are substantial limitations to the utility of a low-coverage genome," says Chris Ponting, a genomicist at the University of Oxford in the United Kingdom. "[It] provides only tantalizing research leads that will always need to be validated by more in-depth sequencing, often from other fossils." That's why other groups are seeking Neandertal DNA for their own projects (see p. 868). There are also lingering concerns about contamination by modern human DNA, which can easily masquerade as Neandertal material because the two genomes are so similar.

Pääbo says he's just getting started. "1× is really just a milestone," he notes. "Our ambition is to go on and produce a Neandertal genome of a quality comparable to, for example, the chimpanzee genome, over the next few years."

Inching toward a genome

Ancient DNA has had a checkered past. Once organisms die, their genetic material begins to degrade. Chromosomes splinter, and some bases transform. Microbes infiltrate the decaying organism, leaving their own genomes to posterity. Humans collecting and cleaning fossils add their own DNA smudges.



Neandertal in hand. Pääbo's dogged pursuit of this ancient human's DNA has paid off.

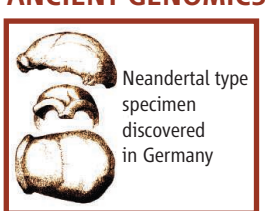
These molecular booby traps have foiled many a sequencing effort, yielding sequences consisting mainly of contaminant. But over the past 20 years, Pääbo and others have worked to avoid the traps. In 1997, he and student Matthias Krings, then at the University of Munich in Germany, decoded about 400 bases from Neandertal mitochondrial DNA (mtDNA) (*Science*, 11 July 1997, p. 176). These were different enough from *H. sapiens* mtDNA to confirm for many that Neandertals were a separate species from us. That feat was possible because mitochondria are plentiful in cells, providing a relative mother lode of genetic material, some 500 to 1000 times more than the DNA in the nucleus.

But in the years following, the field of ancient DNA was at an impasse. Much-heralded fossil bee DNA encased in amber and even dinosaur DNA turned out to be modern human contamination. It was almost impossible to reliably isolate enough ancient genetic material to work with. Not until 8 years later did new technologies help ancient DNA break free of these limitations.

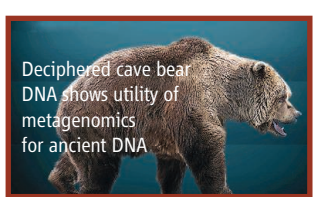
In 2005, James Noonan and Edward Rubin of the U.S. Department of Energy Joint Genome Institute in Walnut Creek, California, working with Pääbo, managed to extract 27,000 bases of ancient cave bear DNA using an approach originally developed for sequencing all microbial DNA in an environment (*Science*, 3 June 2005, p. 1401; 22 July 2005, p. 597). Six months later, Hendrik Poinar of McMaster University in Hamilton, Canada, and his colleagues used so-called next-generation sequencing technology to generate a whopping 28 million bases of DNA from a mammoth bone. "The paper opened the possibility of doing a Neandertal genomics [project] as well," says ancient DNA expert Eske Willerslev of the University of Copenhagen. (*Science*, 20 January 2006, p. 392.)

Pääbo was more than ready to put these advances to use. He and his colleagues had already collected 70 samples from bones from throughout Europe, Central Asia, and Southern Siberia. Three rather nondescript shards, collected around 1980 from the Vindija Cave in Croatia, passed initial tests for the presence of true ancient human DNA. One had a whopping 4% Neandertal DNA, compared with less than 1% in most samples tested. At 38,000 years old, these are some of the youngest Neandertal bones

ANCIENT GENOMICS TAKES OFF



1856 Neandertal type specimen discovered in Germany

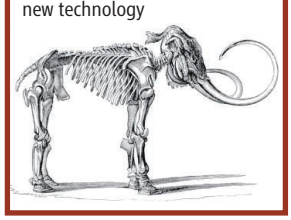


1997 Deciphered cave bear DNA shows utility of metagenomics for ancient DNA

2005 Costly technology and contamination concerns stymie the quest for nuclear DNA

2005 First Neandertal mitochondrial DNA (400 bases) sequenced

2005 Neandertal genome project established



2006 Mammoth DNA sequenced with new technology

2007 65,000 Neandertal bases deciphered by Noonan *et al.*

2007 1 million Neandertal bases deciphered by Green *et al.*

2008 Contamination found in Green *et al.* data

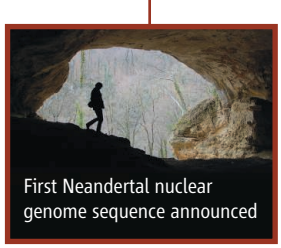
2008 Neandertals found to have red hair, fair skin

2008 Full mammoth genome published

2008 Neandertal *FOXP2* proves same as modern humans

2009 Complete Neandertal mitochondrial genome published

2009 Neandertals found with Type O blood



2009 First Neandertal nuclear genome sequence announced

CREDITS (TOP TO BOTTOM): SILVIO TUEPKE/MPI-EVA; JOHANNES KRAUSE/MPI-EVA

Published by AAAS

Priceless bone. This fossil fragment and two others provided the DNA for the Neandertal genome.

known, and they were buried in a cool, dry cave. And anatomically, they were uninformative, so they were unlikely to have been handled extensively.

Pääbo shared some of the DNA from these bones with Rubin, and the two groups sequenced it independently, using different approaches. Both succeeded: Rubin's team extracted 65,000 bases, Pääbo, 1 million bases (*Science*, 17 November 2006, p. 1113). Their successes prompted Pääbo, the German government, and the sequencing company 454 Life Sciences in Branford, Connecticut, to team up to do the entire Neandertal genome at 1× coverage, while the collaboration between Rubin and Pääbo ended.

But a few observers noted that the two groups, working from the same Neandertal bone, reported different results. "The first time we looked at the two data sets, it was quite clear to us that there were large discrepancies," Rubin recalls. For example, his group had concluded that there was no sign of modern human DNA infiltrating the Neandertal genome, whereas Pääbo's data at that time suggested that our ancestors had intermingled with Neandertals. This and other discrepancies raised suspicions that the Leipzig group's sample may have included DNA from living people. But Pääbo was convinced that contamination wasn't a problem, and given the prestige of his group, few were willing to speak out publicly at the time.

Eleven months later, Jeffrey Wall and Sung Kim of UC San Francisco sounded an alarm. They had reanalyzed both groups' data and recalculated the divergence date of the two species. Rubin's data suggested a final split at about 325,000 years ago, which roughly concurs with fossil evidence, but Pääbo's results gave a date of only 35,000 years ago. Wall and Kim also noticed that the sequenced fragments Pääbo's group reported were relatively long, as might be expected from modern rather than ancient DNA (http://sciencenow.sciencemag.org/cgi/content/full/2007/829/4).

Pääbo was also finding problems with the early results. His team changed its opinion on interbreeding in May 2007 and eventually realized that 11% of their sample was modern human DNA, as they reported in the 8 August 2008 issue of *Cell*. "Contamination was indeed an issue," Pääbo says.

Microbial and nonhuman sequences among the deciphered DNA were easy to spot and cull thanks to comparative genomics, but telling Neandertal from modern human contamination was next to impossible.

So Pääbo and colleagues, including postdoc Richard "Ed" Green, who has spearheaded the Neandertal effort, continued to improve their procedures. They did more of the sample preparation in a clean room to minimize exposure to modern human DNA. They also began adding 4-base tags to DNA extracted from the Neandertal samples so that they could detect modern human contamination introduced after the sample left the clean room. Any completed sequence lacking these tags was discarded. Says Beth Shapiro of Pennsylvania State University (PSU), University Park, "I think this tagging technique is one of the most exciting advances in using ancient DNA and provides a big step forward for the field."

These and other technical challenges put the team behind schedule, so they bought sequencing machines made by Illumina, a

Siberian search. Edward Rubin (*right*) and Institute of Archaeology and Ethnography Director Anatoly Derevianko are still seeking DNA-rich Neandertals.

WANTED: CLEAN NEANDERTAL DNA

It usually pays to have more than one team working on a scientific problem. Take the early work on sequencing Neandertal nuclear DNA, for example. In 2005, Svante Pääbo and his crew at the Max Planck Institute for Evolutionary Anthropology in Leipzig, Germany, gave Neandertal DNA samples to Edward Rubin's group at the U.S. Department of Energy Joint Genome Institute (JGI) in Walnut Creek, California. Together, the two groups helped revitalize the field of ancient DNA by producing more than a million bases of Neandertal nuclear genome sequence (*Science,* 17 November 2006, p. 1068). Equally important: Differences in their results alerted the field to what later proved to be contaminating DNA in Pääbo's analysis.

But since 2006, when the company 454 Life Sciences, with its faster, cheaper sequencing technology, stepped up to do the entire genome with Pääbo (see main text), the Leipzig group has been pretty much on its own. Rubin and others are hoping, however, to remain in the game—if they can get their hands on enough Neandertal DNA.

Rubin plans to sequence target regions from several different Neandertals to complement Pääbo's approach of massively sequencing one sample. The goal is to determine whether differences noted between Neandertal and *Homo sapiens* genomes are real or simply reflect variation among individual Neandertals. Alan Cooper of the University of Adelaide in Australia is also working toward scanning DNA from several Neandertals, and Pääbo himself has sequenced a million bases from additional samples.

But Neandertal DNA preserved and excavated under the right conditions is proving elusive. Over the past 2 years, Rubin and his postdoctoral fellows have traveled the world, striking up relationships with paleontologists in hopes of finding relatively uncontaminated, undegraded Neandertal fossils. So far, they have had little luck with 20 Neandertal specimens from Israel, Spain, France, or Siberia. "None of these have had significant fractions of uncontaminated Neandertal DNA," Rubin laments. "As it turns out, the sequencing of a Neandertal is now much easier than finding a good uncontaminated [sample]."

In the meantime, Rubin's team has been developing methods to extract ever more DNA from each sample. Rubin hopes they will soon get to try these approaches on Neandertals. "We do have some other Neandertal specimens coming in that we are optimistic about," he says. Ancient DNA expert Eske Willerslev of the University of Copenhagen agrees that it's important to have more than one group producing data. "It's definitely worthwhile to have independent replication of results," he says. **–E.P**

CREDITS (TOP TO BOTTOM): CHRISTINE VERNA/MPI-EVA; COURTESY OF EDWARD RUBIN

different kind of "next generation" technology that rapidly and cheaply churns out small stretches of sequence (*Science*, 7 November 2008, p. 838). About one-third of the sequencing was done at the 454 Life Sciences facility in Connecticut, and much of the rest rolled off the Illumina machines.

Ancient panorama

Pääbo's crew harvested one of the first fruits of the Neandertal genome this past summer, deciphering all 16,500 bases making up Neandertal mtDNA. About 200 of those bases differ from those of modern humans, supporting the idea that the two lineages were separate species, they reported in the 8 August issue of *Cell*.

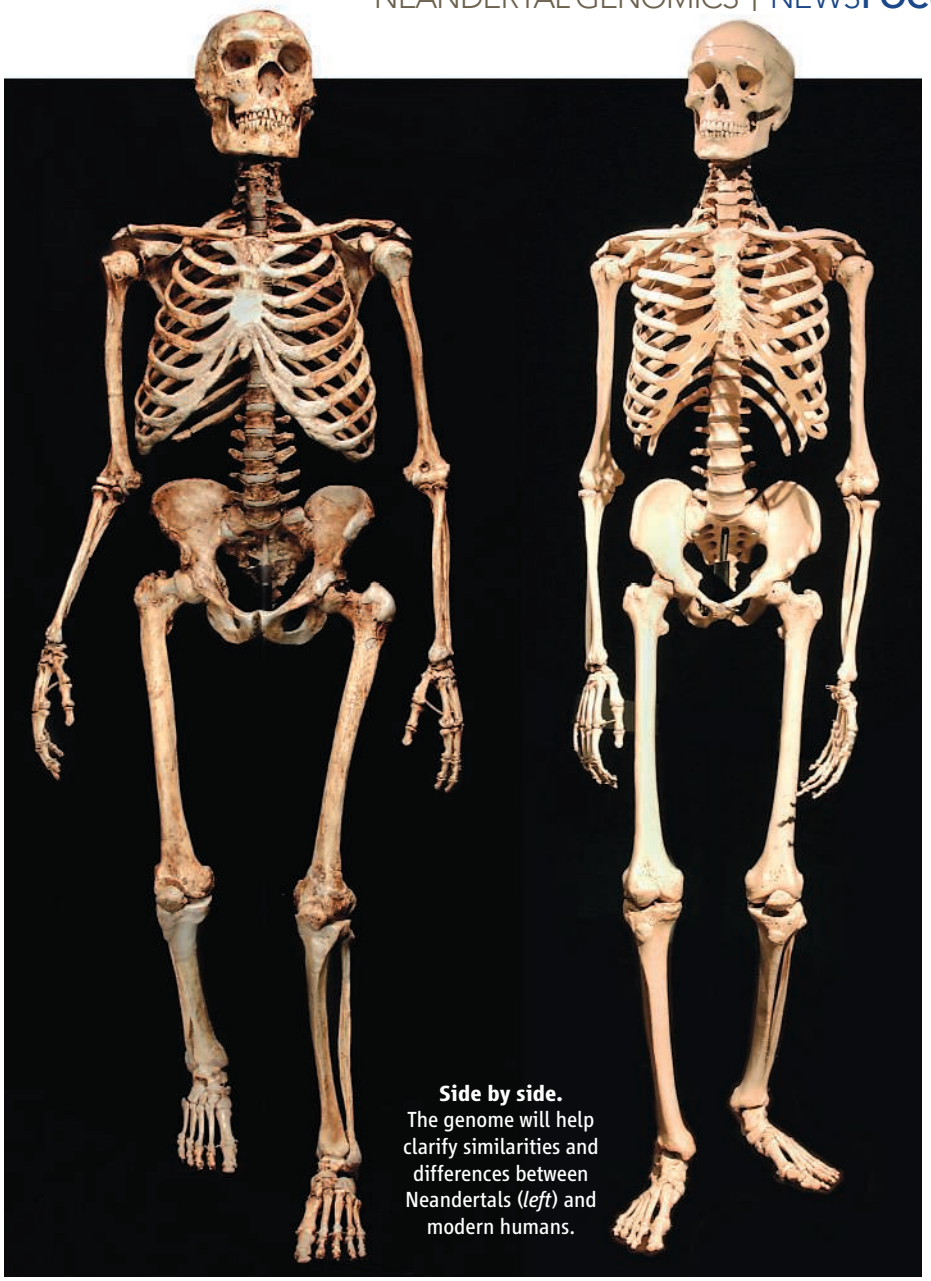
The mtDNA genome, plentiful in cells, was low-hanging fruit compared with the nuclear genome. It took 68.9 billion bases, 96% of which were microbial in origin, to cull the 3 billion human bases of the 1× Neandertal genome.

One of the first results was a surprise: The indeterminate shards of bone turned out to be from two females. So this first genome carries no information on Neandertals' Y chromosome, although the group is now deciphering DNA from additional samples that may include a male.

Now the team is matching up their 3 billion bases to those of the modern human genome to compare the two. "We hope to identify regions where the Neandertal is ancestral at positions where current humans vary," Pääbo explains. "These will then be candidates for having undergone positive selection in recent human history."

The team's preliminary analysis has pinpointed many thousands of such differences. But they have yet to prove that the differences don't vary among living people. Nor have they been able to discern any patterns in the types of regions, or the genes themselves, affected by these changes.

Until now, paleogeneticists have had to take what's called a candidate gene approach: They first identify a gene of interest and then attempt to fish it out of Neandertal DNA. In the past 18 months, those efforts have yielded intriguing insights. For example, the University of Barcelona's Lalueza-Fox found a unique Neandertal variant of a pigmentation gene, *mc1r*, which can result in pale skin and red hair (*Science*, 30 November 2007, p. 1453). More recently, Lalueza-Fox found that leg bone flakes from two adult males from the El Sidron cave in Spain had type O blood, as he and his colleagues reported 24 December 2008 in *BMC Evolutionary Biology*. And with Pääbo, they found that the *FOXP2* gene, which has been linked to language ability, is the same in modern humans and in Neandertals (*Science*, 26 October 2007, p. 546.) Subsequent work suggested that the *FOXP2* finding might be due to contamination, but the new genome indicates that the Croatian Neandertals had the same modern human *FOXP2* variant, bolstering the notion that Neandertals may have had some linguistic facility.



Side by side. The genome will help clarify similarities and differences between Neandertals (*left*) and modern humans.

CREDIT: D. FINNIN/AMNH

With the whole genome data now in hand, researchers can scan the entire genetic landscape for species-defining variants. The whole-genome approach "makes no assumptions about what sequences are interesting," says Rubin. The resulting catalog of differences will be a great resource. "I don't think there is a human geneticist that has a gene or noncoding region of interest who is not going to be interested in comparing their human sequences with that of the Neandertal," Rubin adds.

Varki has already asked Pääbo to keep an eye out for genes for proteins that recognize sialic acids, which in chimp-human comparisons seem to be an evolutionary "hot spot," although no one knows how the changes affect the immune and inflammatory functions of the proteins. Meanwhile, Noonan, now at Yale University, is looking at a particular stretch of regulatory DNA that has a variant unique to humans. If he substitutes the human version of this so-called enhancer for the mouse version, the mouse hand develops as if it's trying to grow a thumb. "We'd like to know when did these sequence changes arise," says Noonan. If Neandertals—who have thumbs—don't have the modern human sequence, for example, then

it's unlikely that the enhancer played a role in thumb evolution.

Did they or didn't they?

Everyone is eager to learn more about the perennial question of whether Neandertals and modern humans interbred. So far, early analysis of the complete nuclear genome shows no sign of admixture, says Pääbo. Working with David Reich of Harvard University and Jim Mullikin from the National Human Genome Research Institute in Rockville, Maryland, the team has compared about one-third of the Neandertal data with sequence from an African and from a European. If Neandertals and ancient Europeans had mixed genes, their genomes should resemble each other more than African and Neandertal genomes do. Thus far, the team has found no contributions by Neandertals to modern humans.

Nor do the sequences of two genes whose histories suggested interbreeding support admixture. Researchers have suggested that a supposedly 1-million-year-old variant of *Microcephalin*, which is associated with brain growth and appeared in modern humans only 37,000 years ago, evolved first in the big-brained Neandertals, then spread to modern humans through interbreeding. But Pääbo's team did not find that variant in the Neandertal nuclear sequence. A gene called *MAPT* also has a very old variant, found primarily in Europeans, that some think may have originally come from Neandertals via interbreeding. Although the variant's function is uncertain, it today provides some female Icelanders with a reproductive advantage, and so researchers speculated that selection might have favored its spread into modern humans. But the Neandertal nuclear genome didn't show this *MAPT* variant, either.

For some, the question of mixing genes was all but settled by the complete mitochondrial genome, which showed no sign of it. Still, some researchers point out that mtDNA sequence comparisons may not catch low levels of ancient interbreeding. "The problem with mtDNA is that … it offers just a single glimpse into the gene-flow question. An absence of gene flow at the mtDNA does not rule out significant levels of gene flow at other genes," says population geneticist Jody Hey of Rutgers University in Piscataway, New Jersey. And the mitochondrial genome is inherited through only the maternal line and therefore would not preserve any sign of mixing between modern females and Neandertal males.

With the full nuclear genome in hand—albeit from females only—population geneticist Daniel Garrigan of the University of Rochester in New York state thinks researchers should be able to detect low levels of gene flow between Neandertals and moderns. He recently used a mathematical model to test just that question. When he looked for evidence of admixture in a stretch of 1 million bases of simulated DNA code, he found that he could detect levels of admixture as low as 1% if the two species split from a common ancestor more than 600,000 years ago. "Even if admixture is low, you still have good power to detect it," says Garrigan.

In practice, however, even if Pääbo's team did see signs of mixing, they would have to be on guard against contamination, which could mimic the signature of ancient interbreeding by putting a modern human variant into the Neandertal genome. Pääbo recognizes the challenge: "It will never be possible to completely rule out minute amounts of contribution from them to us, but it becomes less and less biologically relevant the smaller that contribution could have been."

Hands off. Clean-room procedures for preparing fossil samples (*inset*) helped reduce contamination by modern human DNA.

CREDITS: WOLFGANG STEINMETZ; JOHANNES KRAUSE/MPI-EVA

Reality check

Indeed, even the most ardent Neandertal gene hunters recognize they need to temper their enthusiasm. "It will be dangerous to interpret any of the data when the coverage is 1×," says Steve Scherer, a geneticist at the Hospital for Sick Children in Toronto, Canada. It takes sequencing a base multiple times at each spot to guarantee accuracy—and 1× doesn't allow for much redundancy. With ancient DNA, there's always a chance that the sequence is wrong because DNA changes as it ages. "Discerning what those [genetic] changes mean or meant is going to be tricky," warns anthropologist Pat Shipman of PSU. "There are bound to be dead-ends and false leads." In addition, this scant coverage will make finding duplicated genes and other so-called structural changes, many of which are proving important in separating mammalian species, next to impossible.

Contamination remains a concern. "Although the criteria have improved, it still remains one of the biggest challenges to the Neandertal genome project," says the University of Copenhagen's Willerslev.

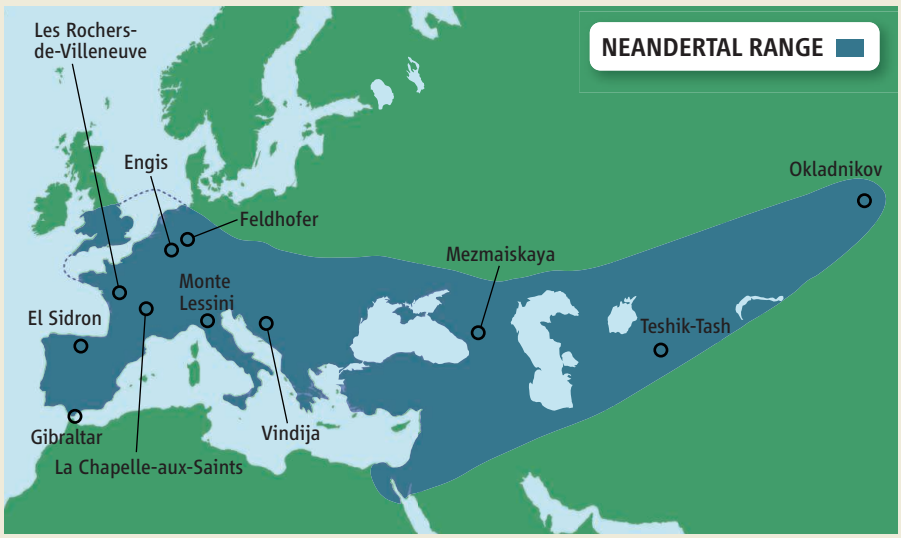
Also, the best way to understand a difference between, say, an *H. sapiens* and a Neandertal gene is to look at that base in multiple individuals. We can do that for our species, but to date, this Neandertal genome represents a consensus; individual variation and sequence validity are not yet established for many regions of the genome. It will be especially hard to trust any Neandertal base that is different from both the modern human's and the chimp's, Pääbo points out because the difference could be due to degradation.

Much legwork needs to be done to begin to pin down genetic differences that might explain the Neandertal's bigger nose, wider body, and short limbs, as few gene variants have been tied to these traits. Lalueza-Fox is beginning to look at skeletal variation in human populations with an eye toward eventually identifying the gene variants that underlie them. Then he plans to look at those same genes in the Neandertal. "But the first step is to do the work in humans," he says.

Moreover, there are limits to what genes can say about how Neandertals lived their lives. "Issues of behavior and ecology will not have strictly genetic answers," says Shipman. "In the end, it is the fossils and the artifacts that are the final arbiters about what the genes produced or did not produce."

But for now, the genes have turned a few hundred milligrams of bone into a gold mine. The Neandertal women who left their bones in the cave likely led undistinguished lives, yet now "they are the genetic spokespeople of their kind," says Green. "Through the DNA preserved in their bones, they may wind up making a profound, posthumous contribution to our species' understanding of its own evolutionary history." **–ELIZABETH PENNISI**

With reporting by Ann Gibbons.



At home in Eurasia. Neandertals ranged from Europe to southern Siberia.

A NEANDERTAL PRIMER

Why are they called Neandertals? The first Neandertal fossil was found in 1830 in Belgium, but scholars failed to realize it was an extinct human. Then in 1848, when quarry workers in Gibraltar came across a strange-looking skull, local scientists again did not know what to make of it. Only in 1856, after miners working in Germany's Neander Valley (*Neander Tal* in German) uncovered a skull cap and other bones, did scientists figure out what they had.

When and where did they live? Bones of full-fledged Neandertals show up in the fossil record at least 130,000 years ago, from Spain to Uzbekistan (see map). Recently, Neandertal mitochondrial DNA was found in bones in southern Siberia, extending their apparent range another 2000 kilometers east. The last known Neandertals survived until perhaps 28,000 years ago on Gibraltar.

How are they related to modern humans? Neandertals are our closest relatives. Although some researchers once thought they were our immediate ancestors in Europe, most now agree that Neandertals and modern humans shared a common ancestor that lived somewhat less than 500,000 years ago, possibly in Africa. No later than 350,000 years ago, fossils resembling this ancestor but with incipient Neandertal features show up in Spain and Britain.

Did they have sex with modern humans? If they did, the two species apparently didn't have many babies together. The latest data from the Neandertal mitochondrial genome and preliminary results from the nuclear genome show little sign of interbreeding (see main text). But many researchers say some admixture cannot be ruled out.

What did they look like? Neandertals were once portrayed as brutish creatures, but scientists now think they resembled modern humans in many ways. Indeed, the late anthropologist Carleton Coon once suggested that a Neandertal dressed in a suit and hat riding the New York City subway would go unnoticed. The next time you ride the subway, look for someone with a stocky, muscular body with short forearms and legs; a large head with bony brow ridges; a jutting face with a very big nose; and perhaps reddish hair and fair skin.

What did they wear? Clothing isn't preserved in the archaeological record, but researchers assume that Neandertals could not have survived ice-age winters without covering up with animal skins. Neandertals did not have needles, but some researchers have argued that the bone awls they routinely made were used to tie animal skins together.

What did they eat? Small and medium-sized game, mostly rabbits and deer, which they probably chased and killed with spears. Some Neandertals ate seafood, including shellfish and sea birds. And at least on occasion, they may have eaten each other. At Neandertal sites in France, Spain, and Croatia, researchers have found Neandertal bones that were apparently cut with stone tools.

Were they smart? Neandertals had big brains, ranging from 1200 to 1600 cubic centimeters, slightly larger than the modern human average. Yet judging from the artifacts they left behind, many researchers have assumed that Neandertals were not as intelligent as ancient *Homo sapiens*. Until the last 10,000 or so years of their existence, Neandertals exhibited little evidence of symbolic behavior, such as art or personal ornamentation. Their skillfully made tools were more sophisticated than those of their ancestors, but those tools changed little for 100,000 years. And although there is good evidence that Neandertals intentionally buried their dead, researchers argue about whether there are signs of burial rituals. Only after 45,000 years ago, when modern humans began colonizing Eurasia, do beads and more sophisticated bone tools crop up at a small number of Neandertal sites. That has led many researchers to suggest that they were merely imitating the modern humans, although that idea is bitterly debated. As anthropologist Leslie Aiello, president of the Wenner-Gren Foundation for Anthropological Research in New York City, put it: "The Neandertals had big brains, and they must have been using them for something."

Could they talk? They may have spoken, but probably not as precisely or inventively as modern humans do. The soft tissues most critical to human language, such as the larynx and vocal cords, do not fossilize, and there are few differences between Neandertals and modern humans in their underlying bony structures. Yet some researchers have suggested that Neandertals couldn't produce vowel sounds as well as modern humans do, and others insist that their limited symbolic behavior suggests limited language skills as well. Nor have genetic studies been able to resolve the debate.

Why did they go extinct? Several decades ago, many researchers assumed that Neandertals met a violent end at the hands of invading modern humans. That view is no longer widespread: There is no evidence of warfare between the species, and Neandertals held on for some 15,000 years after *H. sapiens* arrived. Today's debates focus primarily on whether modern humans outcompeted Neandertals for resources or whether Neandertals succumbed to the cold as the last ice age approached its peak about 25,000 years ago.

–MICHAEL BALTER

SOURCE: SVANTE PÄÄBO

Published by AAAS

PLANETARY SCIENCE

Phoenix Rose Again, But Not All Worked Out as Planned

For better and for worse, Phoenix often wandered from its scripted mission on Mars, but there was some groundbreaking science behind the often distracting headlines

At first, it was all about asparagus. Early results from the Phoenix lander released at a NASA press conference showed that the soil there would be great for raising the vegetable, if only the temperature were 100°C or so higher. Then the news was the umpteenth discovery of water on Mars, even though researchers running an orbiting spectrometer had confidently predicted Phoenix's discovery of near-surface ice. And later, the big science news was snow. A laser shot skyward had detected the white stuff overhead, although it never made it as far as the ground.

Throw in harrowing accounts of balky equipment, frustratingly sticky soil samples, and terminal hypothermia, and it was hard to remember why NASA sent Phoenix to Mars in the first place: to try to decipher the history of water and to see whether, at some point in that history, liquid water could have let martian life bloom. The $450 million mission of exploration fell well short of those ambitious goals, largely because Mars failed to cooperate. Phoenix found minerals that formed when rock interacted with liquid water, but no one knows for sure where or when that happened.

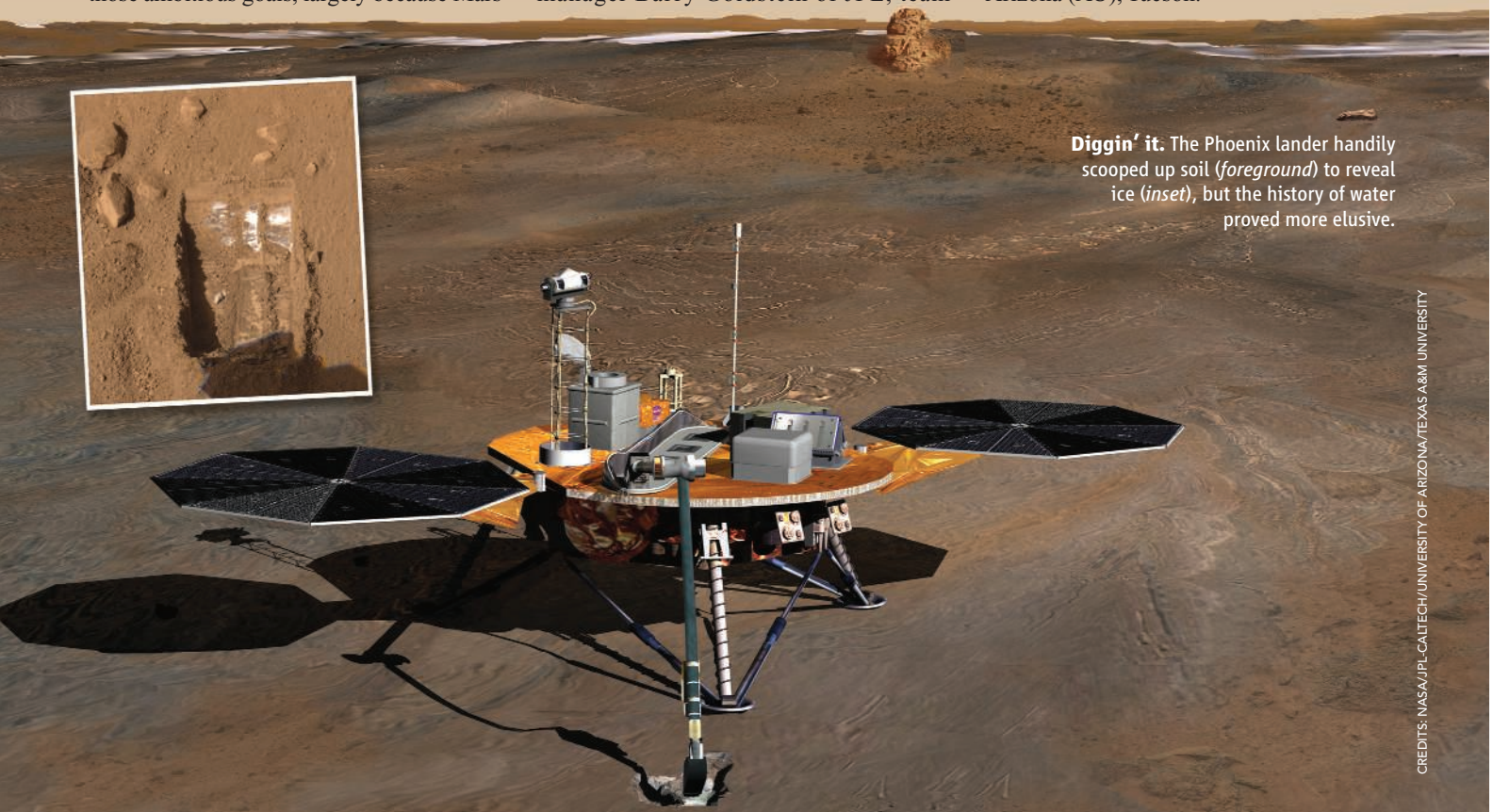
"If we were looking for answers, we'd be disappointed," says applied physicist and instrument lead Michael Hecht of NASA's Jet Propulsion Laboratory (JPL) in Pasadena, California. Even so, planetary scientists see Phoenix's open-ended exploration as well worthwhile. "It reveals that Mars is complex and heterogeneous and has a far more interesting history than thought," says soil scientist Ronald Amundson of the University of California, Berkeley, who is not on the Phoenix team.

Ever onward

Before it could uncover intriguing complexities, Phoenix had to surmount some daunting problems, some of them self-inflicted. According to the plan, said Phoenix project manager Barry Goldstein of JPL, team members' "7 minutes of terror" between Phoenix blazing into the atmosphere and gently settling onto the surface was "going to be followed by 3 months of joy."

Not quite. Mission engineers and scientists had only 151 martian days between landing and being snuffed out by the frigid grip of approaching martian winter, and they spent them grappling with a series of time-consuming mechanical and environmental problems. Clumpy polar soil got stuck when they dumped it onto millimeter-size screening that covered instrument inlets (*Science*, 8 August 2008, p. 758). Sprinkling the soil worked better than dumping it, but no one is sure what caused its maddening stickiness.

Going after ice or even icy soil—a prime target for the mission—proved nearly disastrous. Icy samples refused to fall out of an overturned scoop, like ice cubes refreezing in a drink glass. Unsuccessful efforts to work around the ice problem consumed 3 weeks mid-mission. And then there were the inlet doors of the Thermal and Evolved Gas Analyzer (TEGA), one of Phoenix's two principal analytical instruments. They wouldn't open, at least not fully. The TEGA team had caught the problem early but missed several opportunities to ensure that it had been corrected, according to TEGA lead William Boynton of the University of Arizona (AU), Tucson.



Diggin' it. The Phoenix lander handily scooped up soil (*foreground*) to reveal ice (*inset*), but the history of water proved more elusive.

CREDITS: NASA/JPL-CALTECH/UNIVERSITY OF ARIZONA/TEXAS A&M UNIVERSITY

Published by AAAS

In the end, Phoenix operators managed to fill only five of TEGA's eight single-use analysis chambers, none with the desired amount of ice. One of the four chambers of the Microscopy, Electrochemistry, and Conductivity Analyzer remained empty (though it proved to be a useful calibration blank), and MECA examined eight samples out of a possible 10 under its optical microscope. "Clearly, we would have liked to move more quickly," says Phoenix principal investigator Peter Smith of UA. The icy soil, he says, "was just nasty. We lost part of the clues that would make interpretation easier." Nevertheless, he adds, "we've got a lot of pieces of the puzzle."

A habitable world

Many of the pieces to the water puzzle that Phoenix sent back fit surprisingly well into a picture of a habitable planet. The mission began with the assumption that the most likely place on Mars to find signs of recently liquid water—and therefore a place where microbes could have thrived—was a few centimeters beneath the surface of the northern polar region. Ice was almost certainly there, and it may have periodically melted a bit. On time scales of tens of thousands of years, the planet tilts farther over on its axis, bringing a warming sun higher in the sky of the martian arctic. That might have melted the very top of the ice, wetted the soil, altered its minerals, and—not incidentally—revived any microbes in the martian deep-freeze.

Scientists believed that evidence of such habitable conditions would be preserved in the chemistry, mineralogy, and small-scale geology of the landing site. TEGA, they hoped, would sniff out minerals by progressively heating samples and identifying the volatile products. MECA would add its own water to the soil and see what went into solution, just as if the ice had melted.

Phoenix certainly delivered on the habitability question. Nothing it found was inimical to life. The moderate pH of 8.3 found by MECA and the tiny amount of salts detected by both TEGA and MECA bode well for the evolution of martian life, not just for growing asparagus, notes paleontologist Andrew Knoll of Harvard University. And, much to everyone's surprise, MECA detected relatively abundant perchlorate, possibly magnesium perchlorate. Although used to fuel rockets, perchlorate can also fuel life; certain bacteria are known to use perchlorate as an energy source. "If you warmed that site up, it would be entirely habitable for some types of organisms," says astrobiologist Dawn Sumner of the University of California, Davis.

Phoenix also found signs of once-liquid water, although the news media paid more attention to the abortive snowstorm. MECA found at least 4% calcium carbonate and, less definitively, at least 1% sulfate, says MECA team member Samuel Kounaves of Tufts University in Medford, Massachusetts. Water plus rock typically produces those compounds. "These salts were dissolved in the past," Kounaves says. "They have been wet before."

Sticky situation. Martian arctic soil proved to be sticky, clumpy stuff (*top*, in scoop) that passed through inlet screens—and *partially* opened doors—only with difficulty.

Missing pages

As promising as the habitability front looked, the other mission goal—getting a history of water on Mars—was looking tougher. "It's quite tricky," says Smith. To start with, Phoenix dug through only about 5 centimeters of loose soil before it hit concrete-hard ice. "I don't know that we'll have much to say about profiles in 2 inches of soil," Smith said at the meeting of the Division for Planetary Sciences last October. "That's challenging."

Team members had hoped to read a story about water in that soil layer. In one possible scenario, the ice would have melted tens of thousands of years ago. The meltwater would have wicked upward, leaving a thinner and thinner trail of salts behind it. Such a gradient in salt content or other soil properties from bottom to top would tell the hoped-for story. Using Phoenix's camera and the analytical instruments, "we looked for gradients," Hecht said at the fall meeting of the American Geophysical Union last December. "We didn't see any."

Researchers can offer various explanations for the uninformative uniformity of the Phoenix site. "I am not surprised," says Mars researcher John Mustard of Brown University. "You want to go down several meters," not centimeters. Mustard sees the subsurface ice as a thick remnant of glacial ice deposited during a past martian ice age rather than a thin, soil-cementing frost formed in recent millennia. If that's true, he says, the story of water would lie far deeper than Phoenix reached.

It's also possible that any record the thin soil held may have been wiped out. Planetary scientist Michael Mellon of both the University of Colorado, Boulder, and the Phoenix team notes that, as expected, the icy terrain around Phoenix shows signs of ever so slowly churning like a pot of boiling water as the ice's temperature swings up and down with the seasons. Such "cryoturbation" could be running fast enough to smear out any record of melting until it is unrecognizable, Mellon says. Alternatively, the wind may be to blame. Under the microscope, much of the soil appears to be either fine, windblown dust or larger, angular mineral particles that the wind bounced across the surface. That means the water-altered minerals "could come from the other side of the planet," says geochemist Nicholas Tosca of Harvard.

Mixed results or not, planetary scientists are more than content with Phoenix. "There are chemical processes at work we're just beginning to piece together," says Mustard. "It's a first step." But that first step did yield a pleasant surprise. "The demonstration of diverse chemical processes is really important for people studying Mars," says Sumner. Observations from five previous landers and rovers at lower latitudes had painted a picture of an entire planet drenched billions of years ago in acid and brine (*Science*, 5 January 2007, p. 37). That was not particularly inviting for life and perhaps prohibitive for the origin of life. The Phoenix polar site, on the other hand, with its moderate pH and low saltiness, "says a huge amount in a positive way about habitability," says Sumner. Now planetary scientists can see that, like Earth, Mars has a diversity of environments and therefore chemical and physical transitions between environments that help maintain a healthy biosphere. That, Sumner adds, "is a lot of science for the buck."

–RICHARD A. KERR

CREDITS: NASA/JPL-CALTECH/UNIVERSITY OF ARIZONA/MAX PLANCK INSTITUTE

Can Mathematics Map the Way Toward Less-Bizarre Elections?

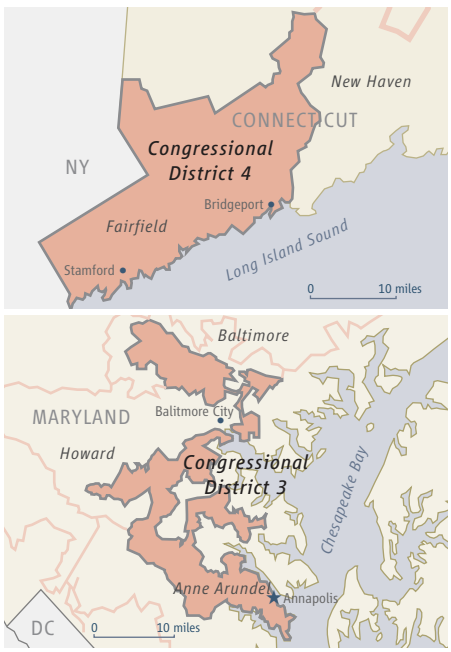
With the 2010 census looming, U.S. politicians and their legal teams are gearing up for another round of wrangling over the spoils of redistricting: the process of deciding which voters get to reelect which members of the House of Representatives and other legislative bodies. Parties in power like to carve up voters to their own advantage, a practice known as gerrymandering. Some reformers, however, hope to limit the mischief—and are turning to mathematics for tools to do so. In a marathon 6-hour session at the Joint Meetings, speakers discussed ideas ranging from pie-in-the-sky theoretical to crust-on-the-ground practical.

The term "gerrymandering" dates back to 1812, when Massachusetts Governor Elbridge Gerry signed into law a tortuous districting map that favored his Democratic-Republican Party over the rival Federalists. But given the fine-grain demographic detail of modern political databases, "the problem is much worse than it used to be," says Richard Pildes, an expert on election law at the New York University School of Law in New York City. Gerrymandering "gives people the sense that they're not really in control of their democracy," Pildes says. "It's part of what contributes to an alienation and cynicism about democracy."

The mathematics of redistricting starts with arithmetic and geometry. Ideally, every district in a state would have an equal population and would be, in some sense, both "contiguous" and "compact." Socioeconomic, political, and racial demographics also come into play. "You can have equipopulous districts and still have whoppingly biased gerrymanders," notes Sam Hirsch, a lawyer at Jenner & Block in Washington, D.C., who specializes in election law and voting rights.

To a mathematician, contiguous means connected—i.e., you can travel from any point in it to any other without leaving the region. Compactness is trickier. Various definitions have been proposed, including one presented at the session by Alan Miller, a graduate student in social science at the California Institute of Technology (Caltech) in Pasadena, California.

Miller's method, developed with Caltech economist Christopher Chambers, quantifies the "bizarreness" of geometric shapes. (The word "bizarre" traces to a 1993 ruling in which the U.S. Supreme Court struck down several oddly shaped congressional districts. Politicians' attempts to handpick their constituents invariably create convolutions in district lines.) In essence, bizarreness is the probability that the most direct path between two randomly chosen voters within a district crosses district lines. The higher the probability, the more bizarre the district is. (The path is required to stay within the state, to avoid penalizing districts that sit on ragged state boundaries.)



How bizarre. Researchers can rank the shapes of Congressional districts ranging from highly compact (top) to convoluted.

Using block data from the 2000 census, Miller and Chambers have computed bizarreness for the congressional districts of Connecticut, Maryland, and New Hampshire. Most compact was Connecticut's 4th District, with bizarreness 0.023; most oddly shaped: Maryland's 3rd district, at 0.860 (see figure).

Bizarreness could be used as a threshold criterion in producing redistricting maps or comparing alternatives, Miller says. "You can use it to reject districts that are badly shaped."

In his own proposal, Hirsch took the idea of thresholds and added a dose of high-octane competition. Rival factions—or anyone else interested in entering the fray—would be able to counter one another's maps, as long as each new submission improved on at least one of three criteria and matched the other two. The goals of the three criteria are to minimize the number of counties cut up by district lines, equalize as much as possible the number of districts leaning toward each of the two major parties, and maximize the number of "competitive" districts, in which neither major-party candidate in a recent statewide contest would have won by more than 7% of the vote.

Hirsch's proposal "is a great idea," says Charles Hampton, a mathematician at the College of Wooster in Ohio, who has been involved in redistricting since the early 1980s. (He drew maps in 1991 for the governor of California's Independent Redistricting Panel.) "We quibble on some of the details," Hampton says, but "I think [it] has some real prospect of producing a much better situation."

No one expects mathematics to solve the problem to everyone's satisfaction. "It's ultimately a political problem," Hirsch says. Kimball Brace, head of Election Data Services in Manassas, Virginia, and a member of the 2010 Census Advisory Committee, agrees. "Redistricting is contradictions out the wazoo," Brace says. "One person's equality is another person's gerrymander." Nonetheless, a growing group of practitioners believe mathematics can play a key role. Says Pildes, "Math can give you tools for creating processes that are likely to lead people to feel that the process is fair and that the outcome is therefore something to be respected."

Taking a Cue From Infinite Kinkiness

"Clouds are not spheres, mountains are not cones, coastlines are not circles," fractal pioneer Benoit Mandelbrot famously wrote. Pool tables, on the other hand, *are* polygons. But mathematicians can ask, "What if they weren't?"

Robert Niemeyer, a graduate student at the University of California, Riverside, and his adviser, Michel Lapidus, are exploring the mathematics of fractal billiards, in which a point-mass cue ball rattles around inside a shape whose boundary seemingly

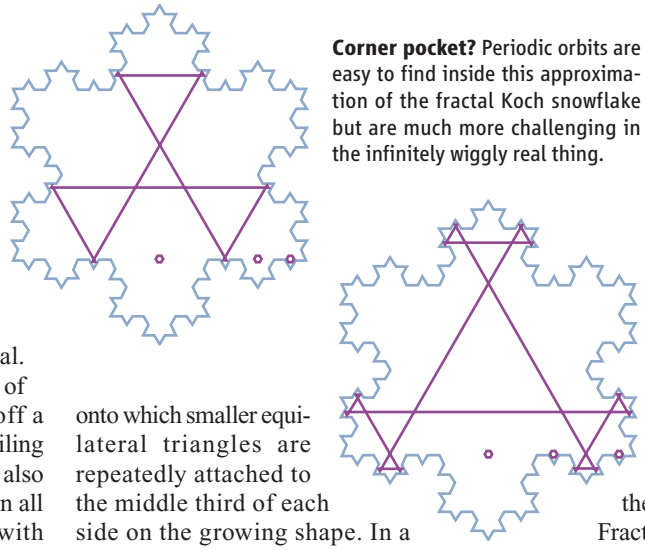
SOURCE: ADAPTED FROM NATIONALATLAS.GOV

consists of nothing but corners. Niemeyer displayed a few trick shots for future fractal hustlers at the Joint Meetings in a special session on experimental mathematics.

"It's very cool," says Victor Moll of Tulane University in New Orleans, Louisiana, who helped organize the session. "It's a new angle on a very classical, well-studied problem."

The research isn't purely recreational. Fractal billiards could account for details of what happens when sound bounces off a rough surface such as a seabed or the ceiling of a concert hall. Quantum physicists are also keen to understand billiard trajectories on all sorts of shapes for their connections with solutions to wave equations in a branch of physics called quantum chaos.

Such applications are far in the future. For now, Niemeyer and Lapidus would be happy to find examples of periodic orbits—trajectories of finite length that wind up retracing themselves endlessly—on a familiar fractal known as the Koch snowflake. The Koch snowflake is formed from an initial equilateral triangle, onto which smaller equilateral triangles are repeatedly attached to the middle third of each side on the growing shape. In a paper published last year in *The American Mathematical Monthly*, Andrew Baxter of Rutgers University, New Brunswick, and Ronald Umble of Millersville University in Pennsylvania classified the periodic orbits on the equilateral triangle itself. Niemeyer hopes to parlay their results to classifications for successive approximations of the snowflake and ultimately to the limiting fractal.



Corner pocket? Periodic orbits are easy to find inside this approximation of the fractal Koch snowflake but are much more challenging in the infinitely wiggly real thing.

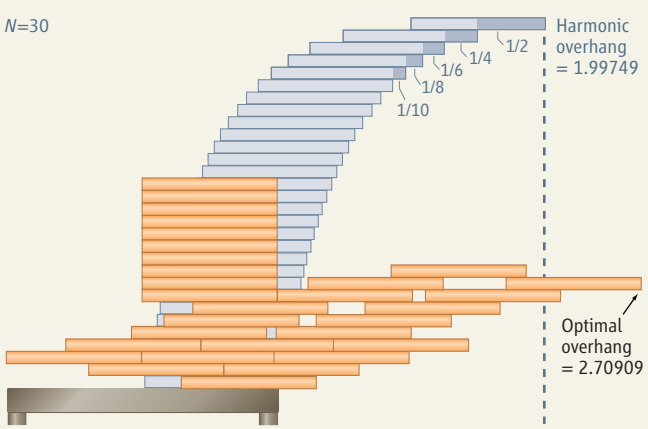
"There's a lot of interesting behavior when we look at these things analytically," Niemeyer says. In particular, a fractal technique known as iterated function systems produces a promising set of trajectories that generalize the "Fagnano orbit" on the equilateral triangle, in which the cue ball simply bounces from midpoint to midpoint to midpoint. Some trajectories display a pleasing pattern of self-similarity, whereas others change radically on successive approximations of the snowflake (see figure).

Fractal billiards fits well with the experimental approach to mathematics in which researchers use computers to systematically explore complicated phenomena and report the findings even when proofs remain elusive, Moll says. "I really like those pictures," he adds. If the concept of bouncing off a boundary that's all kinks and sharp turns seems baffling, Niemeyer agrees: "It makes my head hurt, too." **–BARRY CIPRA**

The Joys of Longer Hangovers >>

One of the joys of calculus is learning how to build a bridge to nowhere, by stacking bricks so that they extend as far out over the edge of a tabletop as you please. The key to the classic problem is that the "harmonic" series $1 + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \ldots$ sums logarithmically to infinity (see figure). A "harmonic" stack of N bricks extends approximately $\frac{1}{2}\ln N$ (ln being the natural logarithm). But a group of mathematicians has now left logarithms far behind, stacking bricks about as far as they can go.

In an invited address at the Joint Meetings, Peter Winkler of Dartmouth College described results that he and colleagues Mike Paterson of the University of Warwick, U.K., Uri Zwick of Tel Aviv University, Yuval Peres of Microsoft Research in Redmond, Washington, and Mikkel Thorup of AT&T Labs Research in Florham Park, New Jersey, have obtained on the best way to stack bricks. The work began with a "remarkable" and "completely new" result by Paterson and Zwick, Winkler says. In a paper just published in *The American Mathematical Monthly*, they showed that a better way to get an overhang with a large set of bricks is not to try extending things at each level but rather to build what looks like a parabolic brick wall with jagged edges. Even though it looks like a lot of wasted masonry, stacking the bricks that way boosts the overhang from a multiple of the logarithm of N to a multiple of the cube root of N—an unexpectedly huge increase.



Balancing act. A "harmonic" stack of N bricks, each of unit length, overhangs a tabletop by half the harmonic sum ($1 + \frac{1}{2} + \frac{1}{3} + \ldots + 1/N$). Clever bricklayers can do better without ever getting their trowels dirty.

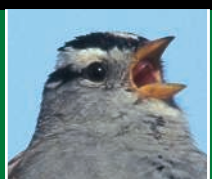
Building a wall to act as a counterweight for the overhang "hadn't occurred to people," says *Monthly* editor Daniel Velleman of Amherst College in Massachusetts. "It's not obvious it's going to help. And the fact that it helps so much is surprising."

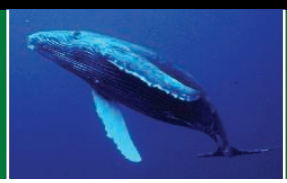
The question Paterson and Zwick left unresolved was how much better one might do. They were able to prove that no amount of cleverness can arrange N bricks to overhang by the square root of N. But that left a lot of room for possible improvement over the cube-root construction.

Enter Winkler, Peres, and Thorup. In a follow-up paper slated for the *Monthly*, the five-man crew has nailed down $6N^{1/3}$ as an upper limit on overhang. They did so by translating the problem of stacking bricks into a problem about random walks. Adding a brick, it turns out, spreads force in essentially the same way that taking a random step spreads the walker's probability of being at a given location equally in each direction. How close to the multiple 6 it's possible to get, however, is unclear. Paterson and Zwick's parabolic construction achieves an overhang of $0.57N^{4/5}$, and other constructions they have found suggest overhangs of about $1.02N^{4/5}$.

Other questions linger as well—the role of friction, for example. "Friction makes a real difference," Winkler says. "And real bricks have a lot of that—not to mention mortar!" **–B.C.**

CREDITS (TOP TO BOTTOM): ADAPTED FROM R. NIEMEYER/UC RIVERSIDE; ADAPTED FROM M. PATERSON, U. ZWICK, *ET AL.*





LETTERS

edited by Jennifer Sills

Robot Rights

IN HIS PERSPECTIVE ("THE ETHICAL FRONTIERS OF ROBOTICS," 19 DECEMBER 2008, P. 1800), N. Sharkey regrets that there are no international guidelines for robot use from the U.N. Convention on the Rights of Children. Although the international community has not officially released such a code of ethics, several sets of guidelines are under way, including Japan's Draft Guidelines to Secure the Safe Performance of Next Generation Robots by the Ministry of Economy, Trade, and Industry; the South Korean Robot Ethics Charter by the Ministry of Commerce, Industry, and Energy; and the guidelines for robots by the European Robotics Research Network (*1*–*3*).

These guidelines differ from each other when it comes to social problems, robot service, and robot political rights. Lack of consistency in ethical guidelines can be attributed to some degree to cultural differences in human beliefs and practices. Differences, for example, in the value placed on the development of independence in infants and toddlers could lead to totally divergent views of the use of robots as caregivers for children. Because different cultures may disagree on the most appropriate uses for robots, it is unrealistic and impractical to make an internationally unified code of ethics.

SHESEN GUO* AND GANZHOU ZHANG

Qianjiang College, Hangzhou Normal University, Hangzhou, Zhejiang 310012, People's Republic of China.

*To whom correspondence should be addressed. E-mail: guoshesen@126.com

References

1. K. Yoon-mi, *AI Magazine* **2**, 144 (2007).
2. L. Lewis, "The robots are running riot! Quick, bring out the red tape," *The Times*, 6 April 2007 (www.timesonline.co.uk/tol/news/world/asia/article1620558.ece).
3. R. J. Sawyer, *Science* **318**, 1037 (2007).

Adapting to Climate Change

IN THEIR PERSPECTIVE "PHYSIOLOGY AND climate change" (31 October 2008, p. 690), H. O. Pörtner and A. P. Farrell stress the importance of understanding organisms' physiological responses to climate change but fail to consider thermally induced phenotypic plasticity (*1*, *2*).

An organism with thermal plasticity can compensate for changing environmental temperatures. For example, some Antarctic fish can reverse the negative effects of rising water temperatures. *Pagothenia borchgrevinki* usually experience water temperatures of –1.8°C, but exposing the fish to +4°C for 4 weeks induces compensatory responses at the level of cellular metabolism, in the cardiovascular system, and in whole-animal swimming performance (*3*, *4*). These phenotypic changes allow *P. borchgrevinki* to function at +6°C, which is 8°C above their usual water temperature, far warmer than any projections of human-induced warming. There are many more examples of similar responses in temperate and tropical ectotherms (*1*, *2*, *5*).

We do not advocate that the potential effect of climate change be taken lightly. However, to respond adequately to climate change, it is crucial to consider the full breadth of physiological responses to temperature. The most appropriate approach is not to assume that all animals are specialized to their environment—which is clearly not the case—but to determine which animals have the capacity for phenotypic plasticity and then concentrate conservation efforts on those organisms that are most likely to be negatively affected.

CRAIG E. FRANKLIN[1]* AND FRANK SEEBACHER[2]

[1]School of Biological Sciences, The University of Queensland, Brisbane, QLD 4072, Australia. [2]Department of Biological Sciences, The University of Sydney, Sydney, NSW 2006, Australia.

*To whom correspondence should be addressed. E-mail: c.franklin@uq.edu.au

References

1. F. E. J. Fry, *Annu. Rev. Physiol.* **20**, 207 (1958).
2. H. Guderley, *Biol. Rev.* **79**, 409 (2004).
3. F. Seebacher *et al.*, *Biol. Lett.* **1**, 151 (2005).
4. C. E. Franklin, W. Davison, F. Seebacher, *J. Exp. Biol.* **210**, 3068 (2007).
5. E. J. Glanville, F. Seebacher, *J. Exp. Biol.* **209**, 4869 (2006).

Response

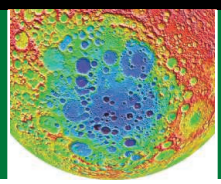
WE THANK C. E. FRANKLIN AND F. SEEBACHER for elaborating on the importance of thermal acclimatization (thermally induced phenotypic plasticity) as a mitigation strategy for climate change. Our figure did in fact note that acclimatization would shift the thermal window along the temperature axis. However, acclimatization is limited to the thermal niche of a species, and these limits reflect thermal specialization. Such a niche becomes visible in analyses of temperature-dependent growth. The niche differs between species or even populations of the same species in various climates.

Although some Antarctic fishes have conserved limited warm acclimatization capabilities beyond present habitat temperatures (*1*–*3*) and thus live on the cold side of their thermal niche, it is clear that not all species do (*4*, *5*). Acclimatization capacity is minor among Antarctic marine invertebrates. It is likely also minimal among Antarctic icefishes

Published by AAAS

Earth and Moon under the microscope

882



Lunar asymmetry

885

that lost their hemoglobin when specializing on the high oxygen content of cold waters. Continued warming will thus cause species losses and, together with invasions by species presently at the doorstep of the marine Antarctic (*6*), will elicit progressive restructuring and functional shifts in marine Antarctic ecosystems.

The hypothesis of variable specialization according to climate variability requires further testing at the level of crucial life functions like growth, reproduction, and development in aquatic and terrestrial species and ecosystems (*7*).

HANS O. PÖRTNER,[1]* ANTHONY P. FARRELL,[2] RAINER KNUST,[1] GISELA LANNIG,[1] FELIX C. MARK,[1] DANIELA STORCH[1]

[1]Department of Integrative Ecophysiology, Alfred Wegener Institute, Bremerhaven, D-27570, Germany. [2]Department of Zoology, University of British Columbia, Vancouver, BC V6T 1Z4, Canada.

*To whom correspondence should be addressed. E-mail: hans.poertner@awi.de

References and Notes

1. G. Lannig, D. Storch, H.O. Pörtner, *Polar Biol*. **28**, 575 (2005).
2. E. Brodte, R. Knust, H.O. Pörtner, *Polar Biol.* **30**, 95 (2006).
3. H. A. Hudson, P. R. Brauer, M. A. Scofield, D. H. Petzel, *Polar Biol.* **31**, 991 (2008).
4. L. S. Peck, *Antarctic. Sci.* **17**, 497 (2005).
5. H. O. Pörtner, L. S. Peck, G. N. Somero, *Philos. Trans. R. Soc. London Ser. B* **362**, 2233 (2007).
6. S. Thatje *et al., Ecology* **86**, 619 (2005).
7. B. Huntley *et al., Ecol. Lett.* **7**, 426 (2004).

8. Supported by the MarCoPoli research program of the Alfred Wegener Institute and NSERC Canada.

LIFE IN SCIENCE

No Restroom for the Weary

OUR LAB IS IN A HISTORIC BUILDING AT THE University of Tokyo. Unfortunately, in this case, the word "historic" is synonymous with "very old" and "shabby." The poor condition of the electric power supply makes our electroencephalography (EEG) experiments a challenge—the electric signals we obtain are contaminated with every sort of noise.

We decided to send a postdoc, Yosuke Morishima, on a journey in search of a good experimental room. Yosuke carried an EEG amplifier and monitor and a colleague wore an EEG electrode cap. They visited every room in the seven-story building, including those that belonged to other labs, and checked the noise level of the EEG. Finally, they found the best room for the experiment: the men's restroom on the east wing of the building. We invited a subject to the new experimental room and started an EEG experiment. The recording was fantastic—we could see beautiful brain signals. However, after running several experimental sessions, we began to receive complaints from people who visited the room for the purpose for which it was originally designed. Thus, our search for the ideal laboratory continues.

KATSUYUKI SAKAI

Department of Cognitive Neuroscience, University of Tokyo 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan. E-mail: ksakai@m.u-tokyo.ac.jp

EDITOR'S NOTE

This is an occasional feature highlighting some of the day-to-day humorous realities that face our readers. Can you top this? Submit your best stories at www.submit2science.org.

CREDIT: PETER HOEY

Social Science Evolves to Include Biology

THE SPECIAL SECTION ON GENETICS OF Behavior (7 November 2008) helps put flesh on the evolutionary skeleton that biologists, and some social scientists, have been developing for many years. In the Perspective "Biology, politics, and the emerging science of human nature" (p. 912), J. H. Fowler and D. Schreiber argue that "biologists and political scientists must work together to advance a new science of human nature." The study of politics is not the only social science that would profit by such collaboration. Indeed, efforts to integrate evolutionary biology have recently been under way in every facet of the social sciences, including anthropology, economics, history, law, linguistics, psychology, sociology, and even literary criticism and aesthetics, in addition to political science. Finally, belatedly, but with increasing empirical and theoretical validity, social scientists representing all disciplines are collaborating with biologists to advance a much-needed, new science of human nature.

DAVID P. BARASH

Department of Psychology, University of Washington, Seattle, WA 98195, USA. E-mail: dpbarash@u.washington.edu

TECHNICAL COMMENT ABSTRACTS

Comment on "Dynamic Shifts of Limited Working Memory Resources in Human Vision"

Nelson Cowan and Jeffrey N. Rouder

Bays and Husain (Reports, 8 August 2008, p. 851) reported that human working memory, the limited information currently in mind, reflects resources distributed across all items in an array. In an alternative interpretation, memory is limited to several well-represented items. We argue that this item-limit model fits the extant data better than the distributed-resources model and is more interpretable theoretically.

Full text at www.sciencemag.org/cgi/content/full/323/5916/877c

Response to Comment on "Dynamic Shifts of Limited Working Memory Resources in Human Vision"

Paul M. Bays and Masud Husain

Cowan and Rouder suggest that a modification to the four-slot model of visual working memory fits the available data better than our distributed-resource model. However, their comparisons of statistical fit are biased in favor of the slot model. Here, we compare the predictions of the two models and present further evidence against the division of visual memory into slots.

Full text at www.sciencemag.org/cgi/content/full/323/5916/877d

Comment on "Dynamic Shifts of Limited Working Memory Resources in Human Vision"

Nelson Cowan* and Jeffrey N. Rouder

Bays and Husain (Reports, 8 August 2008, p. 851) reported that human working memory, the limited information currently in mind, reflects resources distributed across all items in an array. In an alternative interpretation, memory is limited to several well-represented items. We argue that this item-limit model fits the extant data better than the distributed-resources model and is more interpretable theoretically.

Human working memory temporarily represents limited information, but the nature of the information limit remains controversial. Previous research (*1*, *2*) supported a model in which working memory is limited to a fixed number of items, typically three or four. Bays and Husain (*3*) argued that, instead, working memory resources are distributed among all items, as sparsely as is necessary. We argue that Bays and Husain incorrectly assessed the item-limit model and that the evidence favors it over the distributed-resources model. Bays and Husain (*3*) presented haphazard arrays of items, each followed by a probe item displaced from an array item. The task was to indicate the direction of probe displacement. In their model, distribution of memory resources among all items results in greater variance with increasing array size, diminishing the resources available for each item (see Fig. 1A). The probability of indicating one displacement direction (in one condition, outward as opposed to inward) is

$$p_{ij} = \Phi[(d_j - \mu_i)/\sigma_0 N_i^r)] \quad (1)$$

where i indexes the number of array items (array size), j indexes the level of displacement of the probe, Φ is the standard normal cumulative distribution function, d_j is the displacement of the probe, μ_i is a shift parameter, N_i is the array size, σ_0 is the standard deviation for $N_i = 1$, and r is a power constant.

Bays and Husain (*3*) claimed their model outperformed the item-limit model of Zhang and Luck (*1*), but the comparison was unfair. The latter model holds that a fixed number of working memory slots is available, each representing a single item to fixed precision. When there are too few slots and the probed array item is not represented, the participant guesses randomly; consequently,

$$p_{ij} = (k/N_i)\Phi[(d_j - \mu_i)/\sigma_0] + (1 - k/N_i)g \quad (2)$$

where g is the probability of guessing outward. What Bays and Husain overlooked was another postulate of the model. If the number of available slots, k, exceeds the number of array items, N_i, more than one slot represents an item concurrently. (Note that slots doubling up on an item like this may help to exclude distractions by protecting the extra slot.) Consequently,

$$p_{ij} = \Phi[((d_j - \mu_i)/\sigma_0)(k/N_i)^{0.5}] \quad (3)$$

The item-limit model makes predictions much like the distributed-resources model, given a moderate range of probe displacement (see Fig. 1B). The best fit for this model was with $k = 3.31$ items, in the expected range (*1*, *2*, *4*). The increase in response variance across array sizes for $N < k$ is due to the doubling up of slots, contradicting the claim that "The item-limit model of visual working memory predicts that discrimination performance will begin to decline only once the limiting number of items is exceeded" (*3*). The continued increases in variance for $N > k$ are due to increased guessing.

The item-limit model slightly bested the distributed-resources model for three of the four conditions in (*3*) and tied in the fourth (displacement judgments with eye movements). The two models diverge in predictions when the displacements are extreme (beyond about $4\sigma_0$, or about 3 degrees according to our model shown in Fig. 1A) and the array size is large (≥4). Although there is relatively little evidence in this range, the available evidence does not favor the distributed-resources model, as one can see from figure 3D and supplementary figure S2 in (*3*). The item-limit model, in contrast, fit extremes in a study requiring item reproduction (*1*).

The models can also be compared in another paradigm (*2*). In the task (Fig. 2A), participants decided whether the probe's color was the same as the corresponding array item at study or differed categorically from all items in the array. A key manipulation was base rate, the proportion of trials in which color changed. There were three base rates (0.3, 0.5, and 0.7) in different blocks of trials, with the rate always known to the participant. If working memory has discrete slots and one slot per item suffices for categorical colors, then base-rate information can affect guessing processes only when the probed item has no slot. The predictions from this paradigm are expressed in terms of hits,

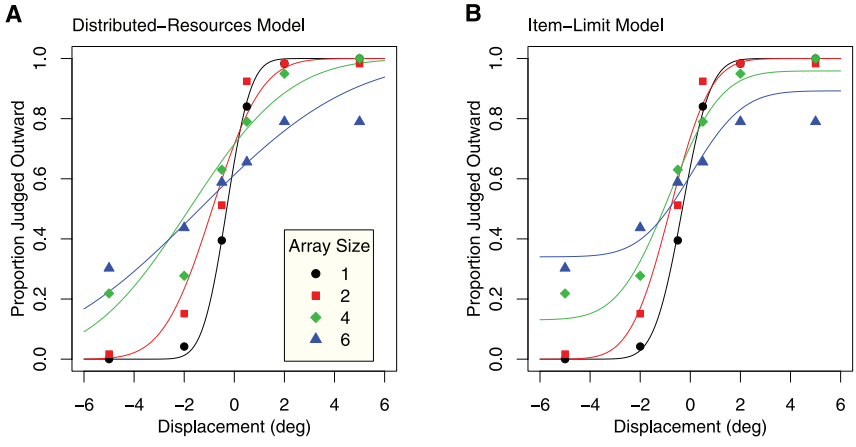


Fig. 1. Two models of performance in an array-item displacement judgment task in which gaze remained at fixation (*3*). Data points in each panel are the proportions of trials in which the displacement was judged to be outward, as a function of the displacement in degrees of visual angle. The graph parameter indicated by point shapes is array size. To fit the models, the function SSE = $\Sigma_{i,j}[\sin^{-1}(\hat{p}_{ij}^{0.5}) - \sin^{-1}(p_{ij}^{0.5})]^2$ was minimized, where the arcsin-square-root transform stabilizes variance across the range of proportions. Root mean square error was higher for the distributed-resources model (0.092) than for the item-limit model (0.078). (**A**) Lines represent the best-fitting distributed-resources model of Bays and Husain (*3*). Best-fitting parameters were $\sigma_0 = 0.75$, $r = 1.03$, and $\mu_i = (0.30, 0.88, 1.79,$ and $1.38)$ degrees for the four array sizes, respectively. (**B**) Lines represent the best-fitting item-limit model. Best-fitting parameters were $\sigma_0 = 1.62$, $k = 3.31$, $g = 0.76$, and $\mu_i = (0.32, 0.80, 0.84,$ and $-0.04)$ degrees for the four array sizes, respectively. The model asymptotes at greater than zero on the low end and less than one on the high end.

Department of Psychological Sciences, University of Missouri, Columbia, MO 65211, USA.

*To whom correspondence should be addressed. E-mail: CowanN@missouri.edu

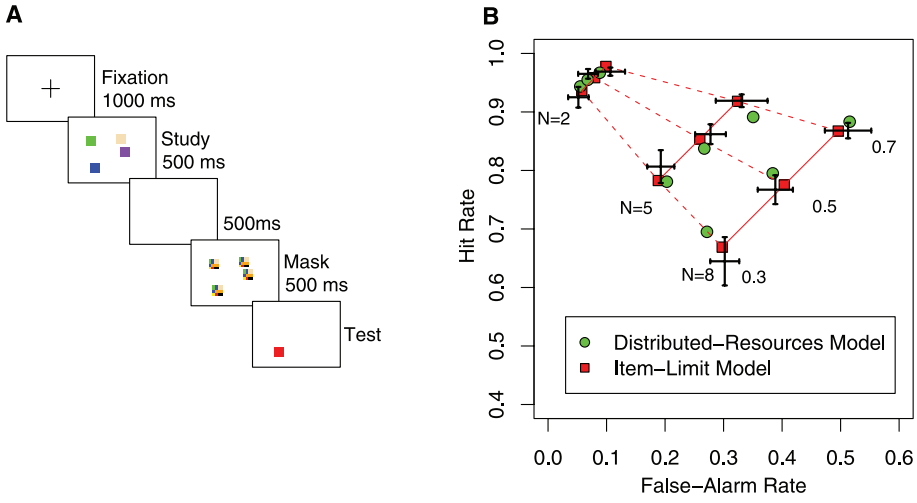


Fig. 2. Procedure and data from Rouder *et al*. (*2*), with model fits from the item-limit and distributed-resources models. (**A**) On each trial, an array of differently colored squares was followed by a blank interval and then a post-perceptual mask. This mask was followed by a single-item probe, which the participants classified as the same as or different from the corresponding study square. The number of squares in the study array was 2, 5, or 8, and the base rate for a change (altered across trial blocks and known to participants) was 0.3, 0.5, or 0.7. Drawing is not to scale. (**B**) Receiver operating characteristics. Group results from (*2*) and averaged predictions for two models. Crosses reflect the data ± SEM on both axes. The averaged predictions of the item-limit model are denoted by the red squares. In this model, the predictions for each array size fall along a different solid line (with slope 1), and the predictions for each base rate fall along a different dashed line. The averaged predictions of the distributed-resources model results are denoted by the green circles. With five parameters per model, the results favored the item-limit model, –2 log-likelihood difference = 7.05.

the probability that changed-color probes are identified as such, and false alarms, the probability that unchanged-color probes are identified as changed. A receiver operating characteristic plot of these measures is shown in Fig. 2B. Hit and false-alarm rates for the item-limit model, denoted H_{im} and F_{im}, where m is the base-rate level, are

$$H_{im} = a[\min(1,k/N_i)] + g_m(1 - a[\min(1,k/N_i)]) \quad (4)$$

$$F_{im} = g_m(1 - a[\min(1,k/N_i)]) \quad (5)$$

N_i is the array size, a is the probability the array was attended, k is item capacity, and g_m is the rate of guessing that there was a change (*2*). The slope of the receiver operating characteristic lines in Fig. 2B is 1.0, with an intercept dependent only on set size (solid lines), whereas the placement of points on these lines reflects only base rates (dashed lines). We implemented the distributed-resources model for this paradigm in a signal detection framework in which the participant perceives an amount of apparent change along a linear dimension and must judge whether this amount arose from one normal distribution, for no-change trials, or from another shifted rightward by a unit amount, for color-change trials. The standard deviation of the distributions depends on the array size: $\sigma_i = (\sigma_0 N_i^r)$. Hit and false alarm rates are

$$H_{im} = \Phi[(d'_i/2)-(c_m/d'_i)] \quad (6)$$

$$F_{im} = \Phi[(-d'_i/2)-(c_m/d'_i)] \quad (7)$$

where $d'_i = 1/\sigma_i$. For the experiment in which three base rates were crossed with three array sizes (*2*), the item-limit and distributed-resource models each have five free parameters. Figure 2B shows that the item-limit model outperformed the distributed-resources model. The models were fit to individuals and a difference in summed deviance (–2 log-likelihood) of 7.05 favored the item-limit model, fitting the means closely.

Extant evidence favors item-limit models (*1*, *2*, *4*–*10*). Moreover, advocates of the distributed-resources model (*3*) left unstated a psychological or neural interpretation of their power law. Item limits apparently hold regardless of item precision (*6*), have known neural bases (*7*), and may occur because of a limited time for items to be reactivated in succession within a processing cycle (*8*) without confusion between features of different items (*9*).

References and Notes

1. W. Zhang, S. J. Luck, *Nature* **453**, 233 (2008).
2. J. N. Rouder *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **105**, 5975 (2008).
3. P. M. Bays, M. Husain, *Science* **321**, 851 (2008).
4. N. Cowan, *Behav. Brain Sci.* **24**, 87 (2001).
5. J. Jonides *et al*., *Annu. Rev. Psychol.* **59**, 193 (2008).
6. E. Awh, B. Barton, E. K. Vogel, *Psychol. Sci.* **18**, 622 (2007).
7. Y. Xu, M. M. Chun, *Nature* **440**, 91 (2006).
8. J. E. Lisman, M. A. P. Idiart, *Science* **267**, 1512 (1995).
9. S. J. Luck, E. K. Vogel, *Trends Cogn. Sci.* **2**, 78 (1998).
10. N. Cowan, *Working Memory Capacity* (Psychology Press, New York, 2005).
11. This research was supported by NIH grants R01 HD-21338 and R01-MH071418 and National Science Foundation grant SES-035-1523. The authors contributed equally to the article.

26 September 2008; accepted 15 January 2009
10.1126/science.1166478

Response to Comment on "Dynamic Shifts of Limited Working Memory Resources in Human Vision"

Paul M. Bays[1,2]* and Masud Husain[1,2]

Cowan and Rouder suggest that a modification to the four-slot model of visual working memory fits the available data better than our distributed resource model. However, their comparisons of statistical fit are biased in favor of the slot model. Here, we compare the predictions of the two models and present further evidence against the division of visual memory into slots.

A long-standing and influential model of visual working memory proposes a fixed number of discrete memory slots, each storing one visual item (*1*–*3*). Recently, this "item-limit" model has been challenged by studies showing that the resolution with which items are held in memory depends on the number of items stored, even when this number is below the proposed limit (*4*–*7*). Rather than being stored in separate slots, these results suggest that discrete visual items share a common memory resource that must be distributed between them (Fig. 1, left). We have demonstrated (*7*) that performance on memory tasks can be captured by a power law describing how the proportion of resources allocated to an item determines the precision with which it is stored. Crucially, we have shown that this "resource" model also predicts the apparent discontinuity in change detection performance that provided the original incentive for dividing memory into slots.

Several attempts have been made to modify the item-limit model to see if it too can be made to account for changes in precision (*5*, *6*, *8*). One proposal by Zhang and Luck (*8*) allows slots to "double up" and store the same item, combined with an averaging process to obtain a single estimate per item. By taking a quote from our article out of context, Cowan and Rouder (*9*) might be misunderstood to suggest that we overlooked this hypothesis; in fact, we addressed this proposal in our supplementary text (*7*). Nonetheless, it is worth examining in more detail.

By allowing multiple slots to combine and thereby represent items with greater precision, this "slots + averaging" model behaves like a quantized version of the resource model, and so is almost indistinguishable from the resource model for small numbers of items (Fig. 1, right). Exceptions occur only when the number of slots is not divisible by the number of items, in which case some items must be allocated more slots than others. No evidence for such unequal allocation has been reported, and indeed this consideration appears to have been overlooked by both (*8*) and (*9*).

Once the number of items equals the number of slots, Zhang and Luck's model makes the strong prediction that any further increases in set size cannot affect the precision with which items are stored, only the probability that an item enters memory. In comparison, the resource model does not predict a "hard limit" on the number of items stored (although neither do we claim that resources will always be distributed equally among all items: outside of the laboratory, this will rarely be an optimal strategy).

If we accept Zhang and Luck's model, their results indicate that working memory capacity is limited to about two items (0.38 probability of storing any individual item in a six-item array) [figure 2A and supplementary figure 3 in (*8*)]. This claim is inconsistent with our previous finding that precision continues to decrease with set size up to at least six items, even when we allow for the possibility that some items are not stored [supplementary text and figure S3 in (*7*)]. Cowan and Rouder now present a reanalysis of our data, which they claim shows that the Zhang and Luck model fits the data slightly better than our resource model. However, their formulation of the slot model [equations 2 and 3 in (*9*)] has more free parameters than the resource model [equation 1 in (*9*)], rendering this comparison invalid (*10*). Indeed, their equation 1 provides a more than adequate fit to our full data set, as shown in Fig. 2A.

Nonetheless, Cowan and Rouder are correct to point out that the two models make the most divergent predictions when larger changes to stimuli are used. We have therefore repeated our experiment with a wider range of stimulus displacements (Fig. 2B). The "slots + averaging" model predicts that the response function will asymptote to a nonmaximal value once the number of slots is exceeded (blue dashed line). However, we find no evidence of this even for six items (blue symbols). As discussed previously [supplementary text in (*7*)], the true error distribution for a feature such as color or shape will be observed only if the tested parameter space corresponds to the parameter space in which the feature is stored in the brain. This may explain why Zhang and Luck (*8*), testing in an arbitrary color space, did not observe a Gaussian distribution at extreme values.

Cowan and Rouder also present an analysis from a previous study (*11*), in which the authors attempted to differentiate between item-limit and resource models without examining precision. Instead, they constructed mathematical models of performance in a change-detection task based on the competing theories and attempted to fit them to experimental data. The authors again claim that the item-limit model provides a better fit than the resource model. In fact, the item-limit model

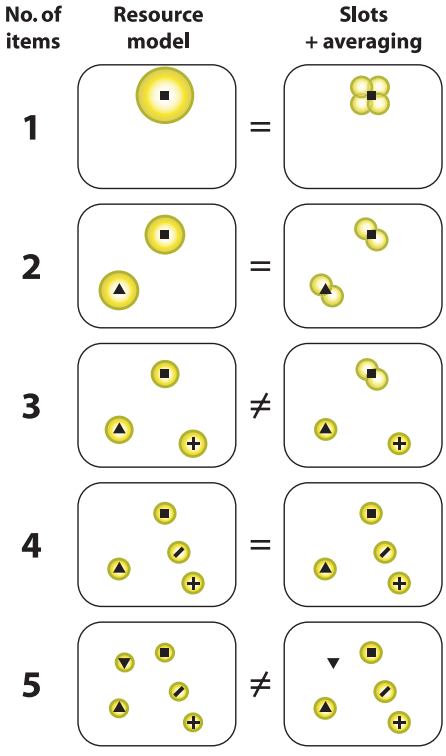


Fig. 1. Allocation of visual working memory under the resource model (*7*) and the slots + averaging model (*8*). In the resource model (left), a common memory resource (yellow) is shared out between multiple visual items (black symbols). Larger circles indicate greater resources dedicated to representing an item in memory, and so greater precision on subsequent recall. In the slots + averaging model (right), memory is divided into discrete slots (here, four) of equal resolution (yellow circles), and multiple slots can combine to represent an item with increased precision. The two models make equivalent predictions for one, two, and four items. For three items, slots must be allocated unevenly in the slots + averaging model, so that one item is stored with greater precision than the others. When the number of items exceeds the number of slots, the slots + averaging model predicts that no information will be stored about items without a slot.

[1]Institute of Cognitive Neuroscience, University College London, 17 Queen Square, London WC1N 3AR, UK. [2]Institute of Neurology, University College London, Queen Square, London WC1N 3BG, UK.

*To whom correspondence should be addressed. E-mail: p.bays@ion.ucl.ac.uk

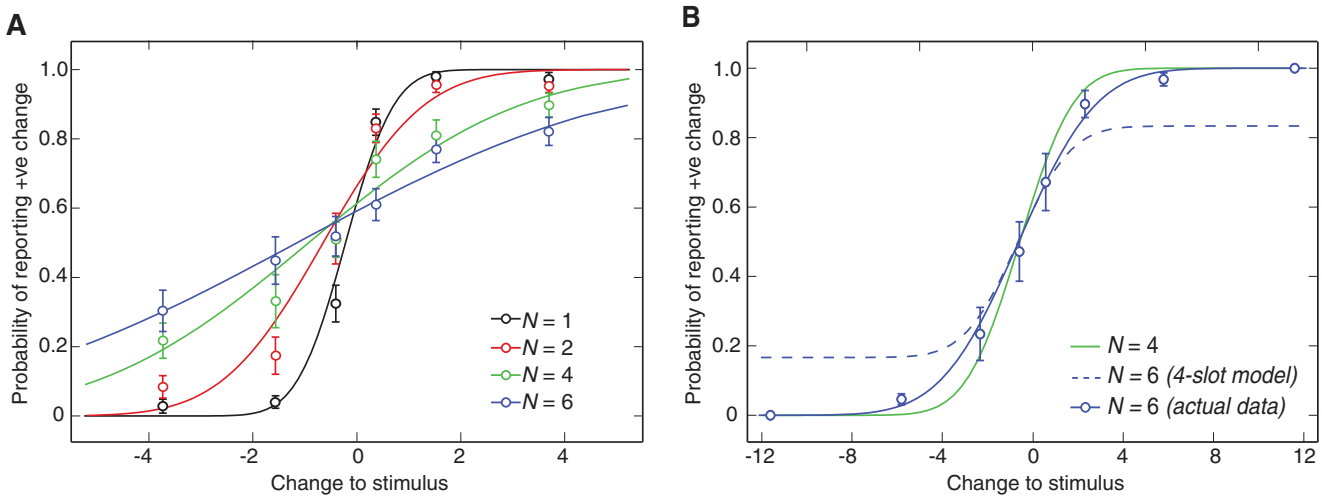


Fig. 2. (**A**) Combined performance on location and orientation memory tasks as reported in (*7*). On each trial, *N* items were presented, then briefly blanked; one item reappeared, changed in either location or orientation, and subjects had to indicate the direction of change. Gaze was monitored to ensure that fixation was maintained. Stimulus change is plotted relative to the standard deviation of the $N = 1$ response function [σ_0 in (*9*)]. Curves represent the maximum-likelihood fit of the resource model defined by equation 1 in (*9*). (**B**) New data from four subjects on the location memory task with an extended range of displacements (up to 10° of visual angle, corresponding to >11 σ_0). The green line indicates the slope of the mean response function for four items. The blue dashed line represents the $N = 6$ response function predicted by an item-limit model with four slots. Blue solid line and symbols indicate the actual $N = 6$ response function (error bars, ±1 SE).

fits their experimental data poorly (*11*), but rather than accepting this as evidence against the theory, the authors introduce an additional component: an attention parameter (*a*) intended to account for trials when subjects do not pay attention [see equations 4 and 5 in (*9*)]. Their comparison of statistical fit is clearly biased in favor of this extended slot model, because no attention component is included in their formulation of the resource model [equations 6 and 7 in (*9*)].

In conclusion, the item-limit model of visual memory cannot be brought into agreement with current results except by introducing new mechanisms, such as combining and averaging of memory slots, for which there is no behavioral evidence or known neurophysiological basis. In contrast, a resource model represents a highly parsimonious account of visual memory, able to explain both old and new results, and with a plausible neural basis in population coding (*12*).

References and Notes

1. H. Pashler, *Percept. Psychophys.* **44**, 369 (1988).
2. N. Cowan, *Behav. Brain Sci.* **24**, 87 (2001).
3. S. J. Luck, E. K. Vogel, *Nature* **390**, 279 (1997).
4. P. Wilken, W. J. Ma, *J. Vis.* **4**, 1120 (2004).
5. G. A. Alvarez, P. Cavanagh, *Psychol. Sci.* **15**, 106 (2004).
6. E. Awh, B. Barton, E. K. Vogel, *Psychol. Sci.* **18**, 622 (2007).
7. P. M. Bays, M. Husain, *Science* **321**, 851 (2008).
8. W. Zhang, S. J. Luck, *Nature* **453**, 233 (2008).
9. N. Cowan, J. N. Rouder, *Science* **323**, 877 (2009); www.sciencemag.org/cgi/content/full/323/5916/877c.
10. W. Zucchini, *J. Math. Psychol.* **44**, 41 (2000).
11. J. N. Rouder *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **105**, 5975 (2008).
12. A. Pouget, P. Dayan, R. Zemel, *Nat. Rev. Neurosci.* **1**, 125 (2000).
13. This research was supported by the Wellcome Trust and the National Institute for Health Research Comprehensive Biomedical Research Centre at University College London Hospitals/University College London.

22 October 2008; accepted 15 January 2009
10.1126/science.1166794

NEUROSCIENCE

Singing in the Brain

Lucie H. Salwiczek

"Birds' song tradition appears to be the truest in the world." –Kant (*1*)

It might come as a surprise to some, but the ability to reproduce vocalizations is rare in nature. All birds and mammals can probably memorize new sounds and may learn to respond appropriately to them, but very few species also produce sounds they have previously memorized. Among the 9700 known species of extant birds, three groups (songbirds, parrots, and hummingbirds), comprising roughly 5200 species, have such vocal learning abilities. Among mammals, only four taxa (humans, cetaceans, bats, and elephants) learn their vocalizations, and humans are by far the most sophisticated vocal learners. In order to produce normal, species-typical vocalizations, a vocal learner needs to hear a conspecific; learning to produce the appropriate sound is directed by auditory feedback control of the learner's voice.

Only vocal learners have developed a special neuronal circuit in the brain devoted to learning and producing vocalizations. These brain regions have not yet been investigated in mammals other than primates. However, the song-control areas of avian brains have received considerable attention from neuroscientists. *Neuroscience of Birdsong* provides a fascinating, thorough (though not complete) overview of their findings, which have been made possible by increasingly sophisticated technology. The introductory section, an easy read even for non-experts, covers song learning, discusses parallels between birdsong and speech, and briefly outlines the relations between avian and mammalian brains. In five thematic sections (each with its own lucid and incisive introduction), contributors provide sometimes very detailed summaries of such topics as lower and higher auditory perception, song acquisition and memory formation, neurogenesis, song recognition, central sensory processing, sensorimotor control, the role of hormones in modulating song behavior, neuronal plasticity, and genomic mechanisms. A final section, "On a Personal Note," offers insights into ground-breaking discoveries and fast-paced development of birdsong research through a biographical sketch of the field's founder, William H. Thorpe, and personal recollections by his student Peter Marler and Marler's students Masakazu Konishi and Fernando Nottebohm.

Neuroscience of Birdsong

H. Philip Zeigler and Peter Marler, Eds.

Cambridge University Press, Cambridge, 2008. 568 pp. $160, £80. ISBN 9780521869157.

As to be expected and as the editors readily admit in the preface, some topics had to be cut short. Nonetheless, the limited attention given to evolutionary aspects of birdsong research is particularly unfortunate, in view of the often-expressed interest in songbirds as models for human speech. Findings in the past decade have shown that avian cognitive abilities are surprisingly sophisticated, in some cases arguably on a par with those of primates. Birds solve complex problems using similar brain regions to mammals, although the bird brain lacks the laminar structure that characterizes the mammalian brain. An international consortium of neuroscientists has revised the nomenclature of the avian brain in order to correct outdated views and misunderstandings intrinsic in the classic terminology. Their results (*2*) highlight the analogies between the avian and mammalian (including human) brain. Including more recent descriptions of the striking similarities between the organization of the song-system circuit and the human vocal control system would have enhanced the relevance of avian studies for neuroscience as a whole.

Also absent is recent work that may be of particular interest for comparisons with humans. For example, noninvasive, in vivo studies of songbird brains by high-resolution magnetic resonance imaging now enable us to disentangle relations among brain circuitry, neural activity, and behavior. The behavior of an individual can be related to changes in its neuronal substrate after having experienced (sometimes repeatedly) different environments or social contexts.

In spite of these absences, *Neuroscience of Birdsong* offers senior students an excellent in-depth introduction to, and provides specialists an insightful summary of current knowledge about, song learning and production in birds. The volume will reward anyone interested in the neurobiology and mechanisms underlying vocal communication.

Well-studied singer: white-crowned sparrow, *Zonotrichia leucophrys*.

We are still only beginning to address the neuroscience of communication. For example, birds use other acoustic signals besides vocalizations. Songs or calls may be accompanied or even replaced by instrumental sounds (e.g., the wing-flapping of the flappet lark, *Mirafra rufocinnamomea*). The neural substrates of such sounds remain unexplored. In addition, singing in many birds (like speaking in humans) is only the vocal part of a composite audiovisual communication, and these complementary bodily gestures have yet to be tackled by neuroscientists.

Similarly, we are only beginning to explore birdsong as the nearest known nonhuman analogy to our language. Women are better than men at spelling and grammar as well as verbal recall tasks, although both sexes have about the same vocabulary. Sex differences in human brain anatomy have been long known; now data on sex differences in the neurocognition of language are slowly accumulating as a result of neuroimaging studies. Dueting male and female dark-backed weavers, *Ploceus bicolor*, produce similar song repertoires of similar complexity, yet they have different-sized brain areas and different complex network properties (*3*). Through comparison with human data, models of dueting avian species may help unravel the mystery of how different brains can achieve the same vocal performances.

Contrasting studies of birdsong and monkey calls, Marler acknowledges, "The neurobiology of the song system, so productive for the study of learning, has told us virtually nothing about the meaning of vocal signals." Yet with great faith and undiminished enthusiasm, he notes, "There are endless new frontiers in the neuroscience of vocal behavior still awaiting our attention."

The reviewer is in the Sub-Department of Animal Behaviour and the Department of Experimental Psychology, University of Cambridge, Downing Street, Cambridge CB2 3EB, UK. E-mail: lss30@cam.ac.uk

References

1. I. Kant, *Über Pädagogik* (Friedrich Nicolovius, Königsberg, 1803).
2. A. Reiner *et al.*, *J. Comp. Neurol.* **473**, 377 (2004).
3. M. Gahr, R. Metzdorf, D. Schmidl, W. Wickler, *PLoS One* **3**, e3073 (2008).

10.1126/science.1168935

EARTH SCIENCES

Bounties from Biomarkers

Katherine H. Freeman

> I keep six honest serving-men
> (They taught me all I knew);
> Their names are What and Why and When
> And How and Where and Who.
> I send them over land and sea,
> I send them east and west;
> But after they have worked for me,
> I give them all a rest.
>
> –Rudyard Kipling (*1*)

In the High and Far off times, O Best Beloved, there were young scientists who were full of 'satiable curiosities about the molecules of bygone life. These scholars of fossils molecules, much like the Elephant's Child, followed a seeming esoteric curiosity against the wisdom and recommendations of those on a more established track. Their enthusiasm and thirst for discovery were paired with intellectual discipline and timed perfectly to harness rapidly emerging analytical tools in chromatography and mass spectrometry. This happy confluence opened the field of Organic Geochemistry and fostered discovery of how (and when and why and where and by whom) the molecular remains of life are transformed and preserved in the mud of oceans, lakes, and, even, perhaps, in the sediments of the great gray-green, greasy Limpopo River itself.

In *Echoes of Life: What Fossil Molecules Reveal About Earth History*, the delightful writing of lead author Susan Gaines is infused with the enthusiasm and extensive knowledge of her collaborators, Geoffrey Eglinton and Jürgen Rullkötter. This unlikely trio—a writer, a founding father of molecular biogeochemistry, and a leading petroleum organic geochemist—have created a work that is as difficult to characterize as it is fun to read.

The reviewer is at the Department of Geosciences, 235 Deike Building, Pennsylvania State University, University Park, PA 16802, USA. E-mail: kate@geosc.psu.edu

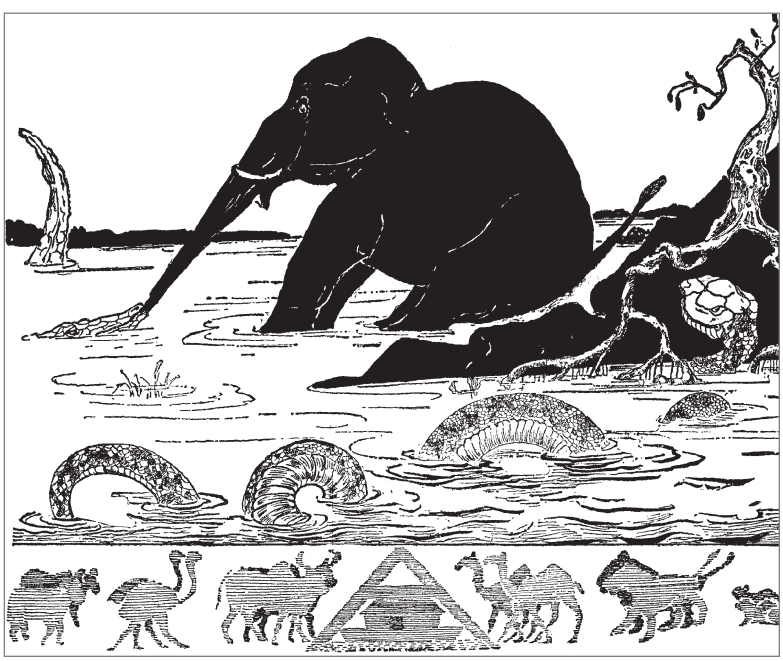
Echoes of Life
What Fossil Molecules Reveal About Earth History

by Susan M. Gaines, Geoffrey Eglinton, and Jürgen Rullkötter

Oxford University Press, New York, 2009. 375 pp. $35, £18.99. ISBN 9780195176193.

It is most emphatically not a standard textbook, although it is riddled with details and contains many helpful and sometimes humorous illustrations, a scholarly bibliography, a glossary, and a useful table of fossil molecules. It is also not a book aimed for a broad audience with limited tolerance for scientific terms and discourse, even though the prose is accessible and technical concepts are whimsically translated into clear analogies. To illustrate, here is the authors' description of how mass spectrometry can be used to discern molecular structure:

> It's like dropping a tray full of wineglasses: a few will escape unscathed, some will shatter completely, but most will break into three pieces—the round cup, the stem, and the thick-glassed base.… By examining all these fragments one can, in principle, figure out what an intact wineglass looked like….



CREDIT: RUDYARD KIPLING/FROM *JUST SO STORIES*

The prose is relentlessly precise and clear, and the authors carefully define concepts and terms as they go. For someone new to the discipline, the book is best read from start to finish (perhaps with an unbroken wineglass at your side) as the terms and details accumulate with each new definition or idea.

This book will be enjoyed by anyone who is curious about the molecular remnants of life and the tales they tell about ancient Earth. The birth, growth, and maturation of biogeochemistry are revealed through vignettes of discovery that capture the thrill of the hunt for insight and snips of the human story behind the science. The engaging anecdotes are filled with intrigue, chance encounters, and even a touch of romance (scientists, including organic geochemists, are human after all). Both the human and scientific tales are told with warmth and excitement.

Organic geochemistry is the study of the origin, transformation, and fate of life-derived matter on Earth. Processes that affect organic molecules stretch from biochemistry to geochemistry and transport us from star dust to mountains to the humble mud of ancient oceans. As illustrated throughout *Echoes of Life*, the study of ancient molecules weaves together sedimentary and physical processes, environmental biology, and the analytical prowess of geochemistry. The molecules' structures and isotopic signatures record their biological origin and history. Thus the molecules provide clues about, and serve as proxies for, conditions in ancient environments and within geologic basins. The chemical, biological, and geologic processes affecting carbon on Earth (and recorded by biomarkers) are fundamental to understanding both climate and fossil energy, timely topics indeed.

Echoes of Life will delight those who know the work but not the individuals, and it will charm those who know both the science and the researchers. But these groups represent a relatively narrow reach. Far more important, the book will be invaluable for readers approaching biogeochemistry from an allied science or as a student. It has an even broader appeal as a story or, more accurately, a collection of stories about discovery. The work offers a festive celebration of why science is fun and of the "rampant human curiosity" that fuels science, scientists, and young elephants alike.

References and Notes

1. This is the first stanza of the poem that appears at the end of "The Elephant's Child" in Kipling's *Just So Stories for Little Children* (Macmillan, London, 1902).

10.1126/science.1167929

ECOLOGY

Should Whales Be Culled to Increase Fishery Yield?

We examine the scientific evidence for the assertion that commercial fisheries are negatively impacted by whales in tropical breeding areas.

Leah R. Gerber,[1]* Lyne Morissette,[1,2] Kristin Kaschner,[3,1] Daniel Pauly[4]

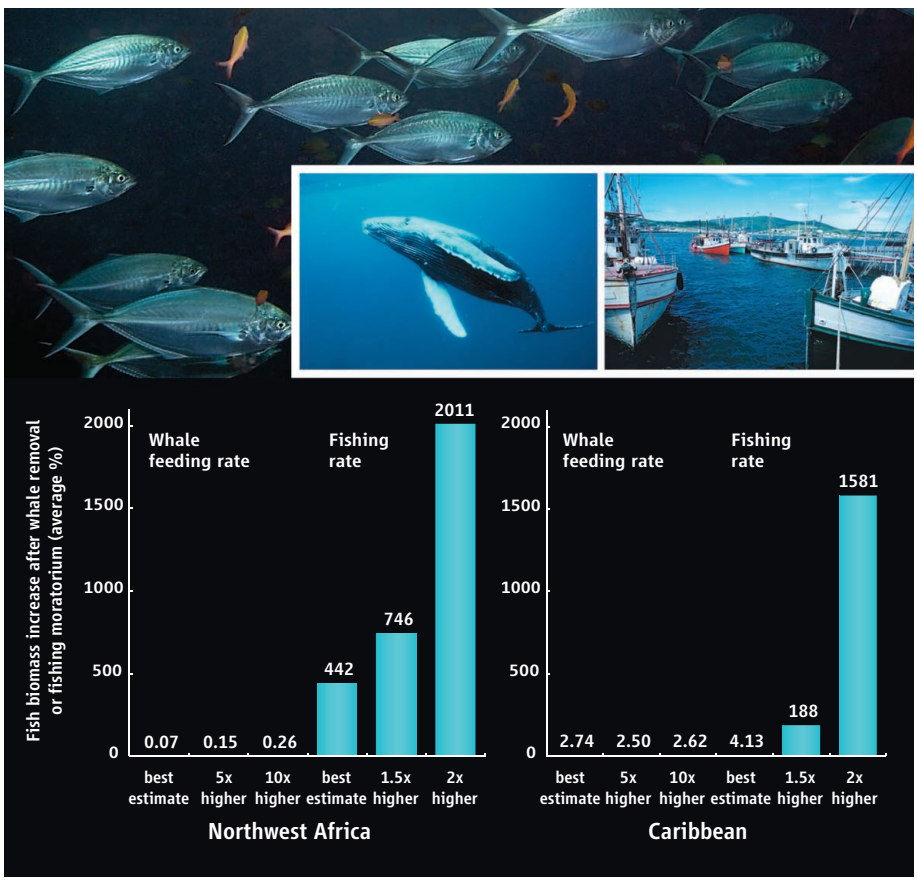
Science and international politics play complicated roles in the global arena of whale conservation and the management of the resources of the world's oceans. The International Whaling Commission (IWC), charged with the global conservation of whales and the management of whaling, introduced a moratorium on commercial whaling in 1986 because of the widespread depletion of whale species and stocks. Despite a lack of scientific data to indicate that many whale stocks have recovered, every year a heated debate takes place at the IWC meeting about the future of commercial whaling. Recently, whaling countries have introduced a new argument for resuming whaling by blaming whale populations for the decline in commercial fish stocks.

Couched in terms of "ecosystem management," whaling countries, including Japan, advocate the culling of whales as a solution to recover overexploited fish stocks and to increase fishery yield (*1*, *2*). Some developing countries, which may benefit economically and politically by supporting pro-whaling nations at IWC (*3*–*7*), have also supported the "whales-eat-fish" assertion. The Caribbean-driven St. Kitts Declaration at the 58th Annual Meeting of the IWC stated: "scientific research has shown that whales consume huge quantities of fish making the issue a matter of food security for coastal nations" (*6*). This issue was also claimed to be one of global concern at a 2008 symposium of IWC members in the Northwest Africa region (*8*).

When scientific information about the role of whales in marine ecosystems and for the economies of developing nations are considered, it becomes clear that delegates from developing countries who support the pro-whaling nations at the IWC may in fact be acting against the best interest of their countries. Whaling does not provide direct benefit to the fisheries that these countries closely depend on (*9*), but rather leads to the loss of species that are important for the structural integrity of their ecosystem (*10*–*12*). Living whales, on the other hand, may actually represent an alternative source of income through whale watching (*13*, *14*).

The rationale for whaling as the solution to depleted fisheries has been questioned by many in the scientific community in light of documented overfishing in oceans globally (*15*), a lack of spatially explicit overlap of resource exploitation between fisheries and whales (*2*), and the unpredictable consequences of culling (*16*, *17*). Based on stomach content analyses of whales caught during the Japanese scientific whaling program and available data on whale abundance, Japanese scientists estimate that whales consume several times as much food as the combined global fisheries catch in recent years (*18*). However, the methodology used by Japanese researchers to support their claim that whales' consumption of fish is an important component of fish declines has been repeatedly criticized (*19*–*22*). Although these discussions have been insightful, they have not stimulated movement within the IWC to break the current deadlock.

One of the obstacles in scientific studies of whales is that there are few data and models available to inform policy discussions. This is particularly true in the tropical waters bordering many of the developing countries that support the resumption of commercial



Negligible effects of removing whales on commercial fish biomass relative to the effect of a fishing moratorium. Estimated increases in fish biomass for best estimates of whale feeding and fishing rates, 5- and 10-fold underestimates of whale feeding, and 1.5 and 2-fold underestimates of fishing. For details, see (*9*).

PHOTO CREDITS (LEFT TO RIGHT): ERIKA ANTONIAZZO/ALAMY; LEAH GERBER; JUPITER IMAGES

[1]Ecology, Evolution, and Environmental Science, School of Life Sciences, Arizona State University, Tempe, AZ 85287–4501, USA. [2]Institut des Sciences de la Mer de Rimouski, 310 Allée des Ursulines, Rimouski, QC, G5L 3A1, Canada. [3]Institute of Biology I (Zoology), Evolutionary Biology and Ecology Laboratory, Albert-Ludwigs-University, Freiburg, Germany. [4]Sea Around Us Project, Fisheries Centre, University of British Columbia, Vancouver, BC, V6T 1Z4, Canada.

*Author for correspondence. E-mail: leah.gerber@asu.edu

Published by AAAS

whaling, although these areas are known to be primarily breeding (not feeding) grounds for baleen whales (*23*–*27*). We conducted an extensive literature search to compile and make use of all available sources of local data to provide a scientific starting point to the discussion (*9*). We also sought to actively involve scientific advisers of delegates who support Japan's position at the IWC meetings and to foster regional collaboration and active dissemination of our findings to inform discussions in local communities among scientists, managers, and other local experts (e.g., 2008 "Whales-Eat-Fish" regional workshops held in Senegal and Barbados, http://lenfestocean.org/whales_fisheries.html).

Using data available from the literature, and, e.g., the Sea Around Us Project (www.seaaroundus.org) and obtained during our regional stakeholder workshops, we developed ecosystem models to examine the potential increase in the biomass of commercially important fish stocks that would result from a reduction in whale abundance in the Northwest African and Caribbean ecosystems (*9*). Any discussion about the interactions between whales and fisheries must be considered in an ecosystem context, which allows investigation of the complex indirect effects of trophic relationships that would otherwise be very difficult to study. Although the IWC Scientific Committee maintains that "Ecosystem modelling cannot be used to predict interactions between marine mammals and fisheries" (*28*–*30*), other studies provide evidence to the contrary that mammals and fisheries can be studied with ecosystem models (*31*–*32*).

Our approach to addressing concerns about scientific uncertainty was to conduct extensive sensitivity analyses to explore the results emerging from a range of assumptions about ecosystem structure and the quality of our input data (table S2). For a wide range of assumptions about whale abundance, feeding rates, and fish biomass, even a complete eradication of baleen whales in these tropical areas does not lead to any appreciable increase in the biomass of commercially exploited fish. In contrast, just small changes in fishing rates lead to considerable increases in fish biomass (see figure, p. 880). We found little overlap between fisheries and whale consumption in terms of prey types, and we also found that fisheries remove far more fish biomass than whales consume (*9*). Moreover, because some whale prey species compete with commercially targeted fish for plankton and prey occupying a lower trophic level in the food web, it is possible that removing whales from marine ecosystems could result in fewer fish available to the fisheries (*9*).

Today, the majority of fish stocks (*33*) and many whale populations (*34*) are seriously depleted, but most available evidence points toward human overexploitation as the root of the problem. When developing tropical countries are encouraged to focus on the notion that "whales eat fish," they risk being diverted from addressing the real problems that their own fisheries face, primarily, overexploitation of their marine resources by distant-water fleets (*35*).

Here, we offer a set of recommendations for rational decision-making by effectively applying ecosystem management concepts to managing whales.

First, the question of "who is eating our fish" should be considered in a larger context (with respect to foreign fleets, ecosystem collapses, and climate change). Indirect social and economic benefits of whales in tropical ecosystems [e.g., tourism (*36*, *37*)] should also be taken into account.

Second, despite complicated politics, science should be an integral component of the discussions about managing whale and fishery interactions. An effort must be made to actively engage scientists and managers from countries that support Japan's claims (*3*–*5*) to help them investigate this issue within an ecosystem context in their own regions. In many cases, fisheries officers in tropical areas, such as the Caribbean, do not necessarily believe the whales-eat-fish arguments. Rather, the arguments are endorsed for reasons related to their aid relationship with Japan, especially in the fisheries sector.

Third, ecosystem modeling tools should be developed in order to bring the best available science to decision-making about the conservation of whales. Research aimed at filling the gaps on key scientific parameters (e.g., abundance, consumption rates, and diet information for key marine organisms) should be supported.

Finally, it is important to recognize that the goal of ecosystem-based management is to manage the whole system for long-term sustainability rather than modifying particular trophic levels in an attempt to maximize fishery yield (*38*). Broad-based, ecosystem management can and should increase an ecosystem's value so that it can provide benefits for future generations.

References and Notes

1. M. Komatsu, S. Misaki, *The Truth Behind the Whaling Dispute* (Institute of Cetacean Research, Tokyo, 2001).
2. K. Kaschner, D. Pauly, in *The State of Animals III: 2005*, D. J. Salem and A. N. Rowan, Eds. (The Humane Society of the United States Press, Gaithersburg, MD, 2005). pp. 95–117.
3. L. Busby, in *Global Corruption Report*, R. Hodess, T. Inowlocki, D. Rodriguez, T. Wolfe, Eds. (Transparency International and Pluto Press, London, 2004), pp. 76–88.
4. A. R. Miller, N. Dolšak, *Glob. Environ. Polit.* **7**, 69 (2007).
5. Third Millennium Foundation, *Japan's "Vote Consolidation Operation" at the International Whaling Commission* (Third Millennium Foundation, Paciano, Italy, 2007).
6. IWC, *Chairman's Report from the International Whaling Commission 58th Annual Meeting*, St. Kitts and Nevis, 16 to 20 June 2006 (IWC, St. Kitts, 2006), 15 pp.
7. K. Stringer, *Dipl. Statecraft* **17**, 547 (2006).
8. IWC, in *Symposium sur l'Utilisation Durable des Ressources Marines Vivantes de la Région Africaine*, Rabat, 11 to 12 February 2008; originally made available at the request of the Republic of Guinea via Circular Communication IWC.CCG.672 of 26 February 2008.
9. Materials and methods are available as supporting materials on *Science* Online.
10. J. E. Duffy, *Ecol. Lett.* **6**, 680 (2003).
11. A. M. Springer *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 12223 (2003).
12. M. R. Heithaus, A. Frid, A. J. Wirsing, B. Worm, *Trends Ecol. Evol.* **23**, 202 (2008).
13. E. Hoyt, G. T. Hvenegaard, *Coast. Manage.* **30**, 381 (2002).
14. J. E. S. Higham, D. Lusseau, *Conserv. Biol.* **21**, 554 (2007).
15. R. A. Myers, B. Worm, *Nature* **423**, 280 (2003).
16. R. T. Paine, M. J. Tegner, E. A. Johnson, *Ecosystems* **1**, 535 (1998).
17. M. Scheffer, S. Carpenter, J. A. Foley, C. Folke, B. Walker, *Nature* **413**, 591 (2001).
18. T. Tamura, in *Responsible Fisheries in Marine Ecosystems*, M. Sinclair and G. Valdimarsson, Eds. (Food and Agricultural Organization of the United Nations & CABI Publishing, Wallingford, UK, 2003), pp. 143–170.
19. N. J. Gales, T. Kasuya, P. J. Clapham, R. L. Brownell, *Nature* **435**, 883 (2005).
20. P. J. Clapham *et al.*, *Mar. Policy* **31**, 314 (2007).
21. S. J. Holt, *Mar. Pollut. Bull.* **54**, 1081 (2007).
22. P. Corkeron, *Mar. Ecol. Prog. Ser.* **375**, 305 (2009).
23. S. K. Katona, J. A. Beard, *Rep. Int. Whaling Comm.* **12**, 295 (1990).
24. P. J. Corkeron, R. C. Connor, *Mar. Mamm. Sci.* **15**, 1228 (1999).
25. W. F. Perrin, B. Würsig, J. G. M. Thewissen, Eds., *Encyclopedia of Marine Mammals* (Academic Press, San Diego, CA, 2002).
26. B. Jann *et al.*, *J. Cetacean Res. Manag.* **5**, 125 (2003).
27. J. Acevedo *et al.*, *Mar. Mamm. Sci.* **23**, 453 (2007).
28. IWC, *ICES J. Mar. Sci.* **4**, 325 (2002).
29. IWC, *ICES J. Mar. Sci.* **6**(suppl.), 413 (2004).
30. IWC, Annex K1: Ecosystem Modelling, in *Scientific Committee Report*, 60th annual meeting, Santiago, Chile, 1 to 13 June 2008 (IWC, St. Kitts, 2008); www.iwcoffice.org/_documents/sci_com/SCRepfiles2008/SCReportFINAL.pdf.
31. B. Bogstad, K. H. Hauge, Ø. Ulltang, *J. Northwest Atl. Fish. Sci.* **22**, 317 (1997).
32. G. Stefansson, J. Sigurjónsson, G. A. Víkingsson, *J. Northwest Atl. Fish. Sci.* **22**, 357 (1997).
33. B. Worm *et al.*, *Science* **309**, 1365 (2005).
34. J. Schipper, *Science* **322**, 225 (2008).
35. D. Pauly, Worrying about whales instead of managing fisheries: A personal account of a meeting in Senegal. *Sea Around Us Project Newsl.* (47), 1 (May-June 2008).
36. P. J. Corkeron, *Conserv. Biol.* **18**, 847 (2004).
37. G. E. Herrera, P. Hoagland, *Mar. Policy* **30**, 261 (2006).
38. E. K. Pikitch *et al.*, *Science* **305**, 346 (2004).

39. Supported by the Lenfest Ocean Program of the Pew Charitable Trusts. We thank the participants of the Dakar and Barbados workshops for contributing to our ecosystem models and J. Melgo for assisting with data management. We also thank M. Bowman and C. Hudson, past and present Lenfest Ocean Program Directors and L. Busby, Pew Whale Conservation Project Coordinator.

Supporting Online Material

www.sciencemag.org/cgi/content/full/323/5916/880/DC1

10.1129/science.1169981

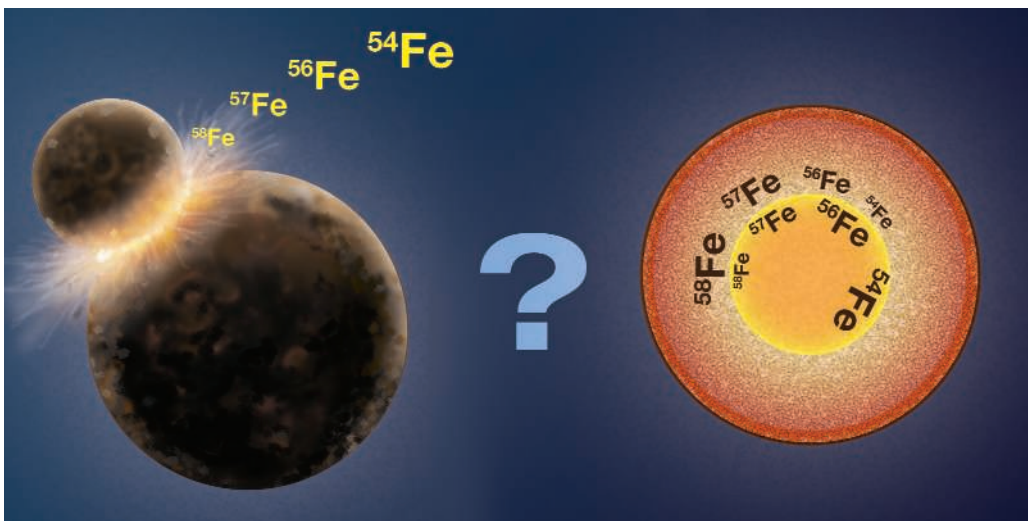
GEOCHEMISTRY

Probes of the Ancient and the Inaccessible

Iron isotopes may be witnesses of the interplanetary impact that formed the Moon or probes of processes occurring at Earth's core-mantle interface.

Franck Poitrasson

The Earth-Moon system was found to have an iron isotope composition enriched with heavier isotopes relative to meteorites, including those thought to originate from Mars (*1*). This isotopic difference was interpreted as fingerprints of a giant impact between the early Earth and a smaller planet, which resulted in the formation of the Moon. Such a high-energy event likely involved partial planet vaporization that would have led to the loss of the lighter Fe isotopes to space, resulting in Earth and the Moon being enriched in the remaining heavy Fe isotopes (see the figure, left panel). On page 912 of this issue, Polyakov (*2*) proposes an alternative interpretation for the difference in isotopic composition. Studying minerals under conditions occurring near Earth's core-mantle boundary, he considers the possible effects of pressures above 100 GPa on the Fe isotope fractionation process between minerals.



Isotope fractionation. Currently proposed explanations of the heavy Fe isotope composition of the Earth-Moon system relative to meteorites. (**Left**) Giant interplanetary impact that led to the formation of the Moon and was accompanied by the preferential loss to space of light iron isotopes. (**Right**) High-pressure core-mantle fractionation that results in the enrichment of light iron isotopes in the metallic core and heavy iron isotopes in Earth's mantle. Symbol size is proportional to isotopic relative enrichment.

Studying the early Earth remains a challenge because of the very few remaining witnesses after 4.5 billion years of geological history. Equally challenging is the study of processes occurring at depths of 2900 km, near the core-mantle boundary. Iron isotope geochemistry may provide an approach to learn about these difficult problems.

Taking current estimates of the deep mantle mineralogical composition, Polyakov (*2*) uses computed Fe isotope fractionation factors at 130 GPa and 4000 K of postperovskite [$(Fe,Mg)SiO_3$], ferropericlase [(Fe,Mg)O], and metallic iron to propose that the Earth-Moon heavier Fe isotope composition relative to other planetary bodies may simply be explained by Earth's core-mantle equilibration at the very high pressures found at depths of 2900 km (see the figure, right panel). Although this idea has already been raised (*3*, *4*), Polyakov takes into account the effect of the very high pressures occurring in the deep mantle on the Fe isotope fractionation between minerals.

If confirmed, this finding has important implications for our understanding of Earth's early differentiation and bulk composition estimates. The current paradigm envisions a final metal-silicate equilibration in a magma ocean several hundred kilometers deep, prior to metallic core segregation from the silicate mantle (*5*). Recent experimental evidence suggests that such a process should not affect Fe isotope signatures of Earth materials (*6*), in agreement with previous conclusions reached from meteorite studies (*7*). However, if we follow Polyakov's interpretation that the heavier Fe isotope composition of the Earth-Moon system results from the equilibration of a metallic core and a silicate mantle, it implies that the early terrestrial magma ocean model cannot lead to a correct understanding of the bulk silicate Earth composition. This has ramifications for our knowledge of deep Earth interior properties and for models of Earth's origins.

However, the original finding was that whereas Earth is isotopically heavier than meteorites and other planets, the Moon bulk Fe isotope composition estimate is actually twice as heavy as that of Earth (*1*). This topic was subsequently hotly debated (*8*–*10*) and remains an open question. Part of the discussion revolves around whether certain types of lunar basalts, characterized by either their low or high content of Ti, were produced through an as yet unknown magmatic process that may affect Fe isotope signatures. If the heavier Fe isotope composition of the Moon relative to Earth is confirmed, it follows that the high-pressure metal-silicate Fe isotope fractionation process proposed by Polyakov cannot alone explain the Fe isotope systematics between planets; the Moon is only 1% of the mass of Earth, so the required high pressures in the interior could not be achieved. Depending on the level of homogenization of the whole mantle, another possible outcome of Polyakov's prediction is that basalts from volcanoes, sampling zones from the deeper mantle known as "hot spots," should be isotopically heavier than the vast majority of the basalts originating from the shallower mantle. However, the current database does not support this (*1*, *3*, *11*). Although Fe isotope variations were recently found in an evolving Hawaiian mafic lava lake, possibly linked to kinetic processes specific to this geological

Laboratoire d'étude des Mécanismes de Transfert en Géologie, Centre National de la Recherche Scientifique, Université de Toulouse, Institut de Recherche pour le Développement, 14 Avenue Edouard Belin, 31400 Toulouse, France. E-mail: franck.poitrasson@lmtg.obs-mip.fr

setting (*12*), this database shows that it is an uncommon feature in mafic igneous rocks. Contrary to Weyer's assertions (*13*), terrestrial bulk mafic rocks generally do not show Fe isotope variations through differentiation: They are characterized by a very homogeneous Fe isotope composition (*1*, *3*, *11*).

Further work is therefore required to understand this apparent contradiction between Polyakov's predictions and our current knowledge of the Fe isotope signatures of igneous rocks. For instance, new ab initio techniques have correctly reproduced the iron isotope fractionation observed experimentally between aqueous species (*14*). Polyakov's Fe isotope fractionation factors between minerals, based on spectroscopies (*15*), should therefore be subjected to such computations. Also, estimating the isotopic composition of the silicate portion of a planet on the basis of a few samples taken at the surface remains a challenge, and this implies that a correct understanding of iron isotope fractionation laws between minerals in magmatic systems has yet to be properly established. The work on bulk planetary stable iron isotope estimates and systematics is promising, as it may eventually lead us to see the Moon's igneous history or Earth's deep mantle processes in a new light.

References

1. F. Poitrasson *et al.*, *Earth Planet. Sci. Lett.* **223**, 253 (2004).
2. V. B. Polyakov, *Science* **323**,912 (2009).
3. R. Schoenberg, F. von Blanckenburg, *Earth Planet. Sci. Lett.* **252**, 342 (2006).
4. H. M. Williams *et al.*, *Earth Planet. Sci. Lett.* **250**, 486 (2006).
5. J. Li, C. B. Agee, *Geochim. Cosmochim. Acta* **65**, 1821 (2001).
6. F. Poitrasson, M. Roskosz, A. Corgne, *Earth Planet. Sci. Lett.*, 10.1016/j.epsl.2008.12.025 (2009).
7. F. Poitrasson, S. Levasseur, N. Teutsch, *Earth Planet. Sci. Lett.* **234**, 151 (2005).
8. S. Weyer *et al.*, *Earth Planet. Sci. Lett.* **256**, 638 (2007).
9. B. L. Beard, C. M. Johnson, *Earth Planet. Sci. Lett.* **256**, 633 (2007).
10. F. Poitrasson, *Earth Planet. Sci. Lett.* **256**, 484 (2007).
11. B. L. Beard *et al.*, *Chem. Geol.* **195**, 87 (2003).
12. F.-Z. Teng, N. Dauphas, R. T. Helz, *Science* **320**, 1620 (2008).
13. S. Weyer, *Science* **320**, 1600 (2008).
14. A. D. Anbar, A. A. Jarzecki, T. G. Spiro, *Geochim. Cosmochim. Acta* **69**, 825 (2005).
15. V. B. Polyakov, S. D. Mineev, *Geochim. Cosmochim. Acta* **64**, 849 (2000).

10.1126/science.1169281

CELL BIOLOGY

The ABCs of Lipophile Transport

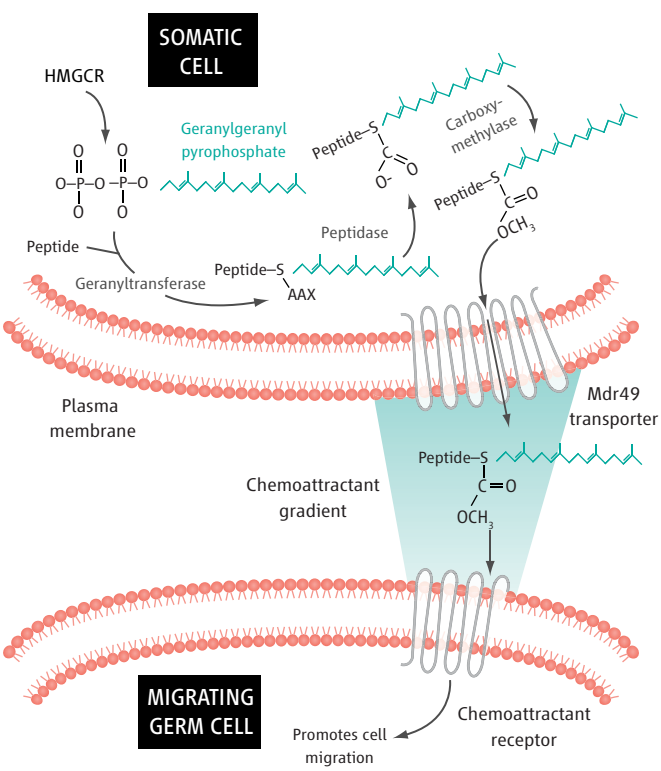
Proteins pump lipophilic molecules out of cells, establishing gradients that guide cell movement during development.

Timothy Hla[1] and Dong-Soon Im[1,2]

During animal development, certain embryonic cells export lipophilic molecules into their surroundings to guide tissue and organ formation. Two proteins that function as cellular transporters for such molecules have now been identified. On page 943 of this issue, Ricardo and Lehmann characterize a transporter in the fruit fly *Drosophila melanogaster* that exports a lipophilic molecule essential for germ cell migration (*1*). In another study, Kawahara *et al.* identify a transporter in the zebrafish *Danio rerio* that exports a lipophilic molecule to direct muscle cell movement during heart development (*2*). The findings underscore the generality and importance of cellular export mechanisms for molecules that presumably establish precise gradients in space and time to control cell migration.

In *Drosophila*, germ cells migrate toward somatic gonadal precursor cells to form gonads. The process is controlled by the activity of an enzyme in the somatic gonadal precursor cells, 3-hydroxy-3-methylglutaryl coenzyme A reductase (HMGCR), which catalyzes the formation of a lipophilic geranylgeranyl moiety that is added posttranslationally to the carboxyl termini of peptides. Loss of functional HMGCR, or its misexpression in the wrong cell type, results in aberrant migration of germ cells and defective gonad formation (*3*). Ricardo and Lehmann show that an ABC-type transporter, Mdr49, acts downstream of HMGCR. It turns out that Ste6p, an ABC-type transporter in yeast cells (*Saccharomyces cerevisiae*), exports a mating factor that is also geranylgeranylated. The authors determined that Ste6p could substitute functionally in *Drosophila* cells lacking Mdr49. Remarkably, orthologs of other enzymes involved in the yeast mating factor export pathway—prenylated peptide protease and carboxymethylase—are also required for germ cell migration toward somatic gonad precursor cells (see the figure). The findings not only suggest that the lipophile export machinery is conserved in evolution, but also that the germ cell chemoattractant may be structurally similar to a yeast mating factor (a geranylgeranylated peptide). A challenge will be to elucidate the structure of the germ cell chemoattractant.

The germ cell migratory system in *Drosophila* also relies on the lipid phosphate phosphatase enzymes Wunen-1 and -2, which dephos-



Go that way. Somatic cells in the developing fly gonad secrete a lipophilic molecule (geranylgeranylated peptide) using the transporter Mdr49. A molecular gradient is established that attracts germ cells by activating receptors for the chemoattractant. AAX (A, Ala; X, any amino acid).

CREDIT: N. KEVITIYAGALA/*SCIENCE*

[1]Center for Vascular Biology, University of Connecticut Health Center, Farmington, CT 06030, USA. [2]College of Pharmacy, Pusan National University, Busan 609-735, South Korea. E-mail: hla@nso2.uchc.edu

phorylate extracellular lysophospholipids. This modification, which occurs in both somatic and germ cells, is thought to create a lysophospholipid gradient essential for germ cell migration (*4*).

The work of Kawahara *et al.* (*2*) describes the cellular export of a bioactive lipid called sphingosine-1-phosphate (S1P) in zebrafish. It was previously found that S1P activates a G protein–coupled receptor called Miles Apart (Mil) in zebrafish, and that loss of functional Mil results in the formation of split hearts (cardia bifida) because of the arrested migration of cardiomyocyte precursors. Mil expression in mesodermal cells somehow guides the migration of cardiomyocyte precursor cells toward the midline of the developing embryo, an essential step in the formation of the primitive heart tube (*5*).

Kawahara *et al.* conducted a mutagenesis screen in zebrafish and identified a transporter-like protein, Spns2, that acts upstream of Mil. The authors showed that expression of the *spns2* gene in yolk syncytial layer cells is essential for the function of Mil in early heart development. The same gene was identified in another study with a similar mutagenesis strategy (*6*). The authors further showed that Spns2 protein (or its human or fish orthologs) exports S1P, which then activates Mil in the mesoderm, allowing the proper migration of cardiomyocyte precursors. Moreover, loss of functional Spns2 results in the formation of split hearts, suggesting the essential nature of S1P export in heart development (fig. S1).

In contrast to the *Drosophila* Mdr49 system, substrate recognition by Spns2 may be more specific, because the related protein Spns1 could not transport S1P (*2*, *6*). Previous work suggests that ABCC1 and ABCA1 transporters (which export sterols, eicosanoids, and other lipophilic molecules) may transport S1P as well (*7*, *8*). However, genetic evidence that these transporters move S1P is lacking, and given the new findings on Spns2, it will be important to assess the role of various transporters in exporting S1P from cells.

Nonetheless, the expression and function of Spns2 may be tightly regulated to establish and/or maintain precise S1P gradients spatially and temporally during embryogenesis. S1P activation of the Mil receptor in the mesodermal cells would then create a permissive environment for the migration of cardiomyocyte precursor cells toward the midline, thereby allowing the formation of the primitive heart tube.

The precise spatial and temporal establishment of lipophilic molecule gradients as guidance cues underlies many diverse biological processes. S1P gradients guide lymphocyte egress from thymus and peripheral lymphoid organs; indeed, modulation of this process is the basis for therapeutic development in autoimmune diseases (*9*). Further determining how the expression of lipophilic molecules and their transporters is regulated under physiological conditions may reveal whether and how they contribute to pathological processes. Given that transport inhibitors achieve great therapeutic efficacy (as in the case of selective serotonin reuptake inhibitors), better knowledge in this area may contribute to identifying potential drug targets. In addition, geranylgeranylated secreted factors would be inhibited by HMGCR inhibitors (such as statins), which are widely used to control heart disease and stroke. Whether such a mechanism is a part of statins' efficacy remains to be determined.

References

1. S. Ricardo, R. Lehmann, *Science* **323**, 943 (2009).
2. A. Kawahara *et al.*, *Science* **323**, 524 (2009); published online 11 December 2008 (10.1126/science.1167449).
3. A. C. Santos, R. Lehmann, *Dev. Cell* **6**, 283 (2004).
4. A. D. Renault, Y. J. Sigal, A. J. Morris, R. Lehmann, *Science* **305**, 1963 (2004); published online 26 August 2004 (10.1126/science.1102421).
5. E. Kupperman, S. An, N. Osborne, S. Waldron, D. Y. Stainier, *Nature* **406**, 192 (2000).
6. N. Osborne *et al.*, *Curr. Biol.* **18**, 1882 (2008).
7. P. Mitra *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **103**, 16394 (2006).
8. K. Sato *et al.*, *J. Neurochem.* **103**, 2610 (2007).
9. S. R. Schwab *et al.*, *Science* **309**, 1735 (2005).

Supporting Online Material

www.sciencemag.org/cgi/content/full/323/5916/883/DC1
Fig. S1

10.1126/science.1170009

VIROLOGY

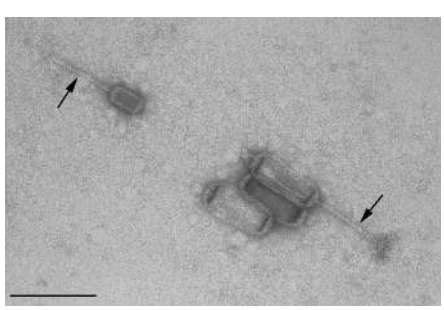
Making Nice with Viruses

Genomics reveals the origin of a polydnavirus lineage and a new way for viruses and their hosts to live together.

Donald B. Stoltz[1] and James B. Whitfield[2]

The polydnaviruses have long been regarded as an anomaly in virology. For example, virus particles, or virions, will readily infect a variety of cell lines, but never replicate in them. The reason is that unlike all other known viruses, that portion of the polydnavirus genome necessary for making progeny virions is in fact not packaged into them (*1*). On page 926 of this issue, Bézier *et al.* reveal where the missing genes are: embedded within the chromosomal DNA of the virus's wasp host (*2*). More importantly, Bézier *et al*. have uncovered a viral origin for these essential genes—at least in theory, they could have been of wasp origin. In so doing, the authors solve a long-standing mystery, and at the same time establish a new paradigm in virology: Whereas viruses have been typically seen as either parasites or commensals, we must now recognize a potential for obligatory mutualism.

Beginning in 1967 (*3*), viruslike particles were being observed in the ovaries of numerous species of parasitic wasps (parasitoids). These entities were present in all females of all affected species, suggesting vertical transmission. It was also evident that the particles were designed for export, initially into oviducts, and subsequently (along with parasitoid eggs) into the body cavity of the host—usually a caterpillar—being parasitized by the wasp. Viruslike particles from two different parasitic wasp species were eventually isolated and shown to package circular double-stranded DNAs of varying sizes (*4*, *5*). This led to the establishment of a new virus family, *Polydnaviridae* (from polydisperse DNA virus) (*6*). At least in one case, purified particles



A tale of two tails. An electron micrograph shows similar tail-like appendages (arrows) associated with a bracovirus nucleocapsid (upper left) and a nudivirus capsid (lower right), suggesting a possible phylogenetic relationship between the two types of virus. Unlike nudiviruses, which are conventional viruses, the bracoviruses do not package genes that are necessary for viral morphogenesis. Until now, convincing support for a viral origin of the bracoviruses has remained elusive. Scale bar, 200 nm.

CREDIT: D. STOLTZ

[1]Department of Microbiology and Immunology, Dalhousie University, Halifax, Nova Scotia, B3H 1X5, Canada. E-mail: dstoltz@ dal.ca [2]Department of Entomology, Urbana, IL 61801, USA. E-mail: jwhitfie@life.uiuc.edu

were immunosuppressive, protecting parasitoid eggs from the lethal defensive response that would otherwise be mounted in parasitized caterpillars (*7*). Because tens of thousands of parasitoid wasp species were probably carrying this type of virus, it was widely assumed that we were looking at mutualism on a grand scale. But, were the polydnaviruses really viruses?

The answer is a matter of semantics, and of how we define a virus. Perhaps the more important issue has to do with origins; that is, are the polydnaviruses a cellular "invention" [a nuclear secretion sensu (*8*)] or did they originate from conventional viruses? Given that genes encoding viral structural and replicative functions are not packaged into polydnavirions, then they must be sought in wasp genomic DNA. If such genes could be found and shown to have authentic viral relatives, then whatever we decide to call these entities, they must have descended from typical viruses.

As has now been convincingly demonstrated, homologs of genes derived from conventional viruses do indeed reside within wasp chromosomal DNA. Specifically, Bézier *et al.* have discovered 22 putative nudivirus gene homologs in the wasp genome, including 12 core genes that are also present in a close and well-known relative, the baculoviruses. Interestingly, a putative relationship between bracoviruses (one of the two major recognized polydnavirus lineages) and *Oryctes rhinoceros* virus was suggested about 30 years ago, on the basis of a remarkable structural resemblance (*9*) (see the figure). At that time, the *Oryctes* virus was thought to represent a nonoccluded baculovirus, but recently, it has been assigned to the baculovirus-related genus, *Nudivirus* (*10*).

So, why is this important for virology? Again, there's the issue of how to define a virus. Should we exclude the polydnaviruses? Consider a typical definition for "virus": a transmissible agent that in extracellular form consists of a DNA or RNA genome minimally packaged within a protein coat, replicating by the coordinated assembly of subviral components, rather than by growth and division. "Extracellular" implies that all viruses are transmitted horizontally, from one host to another, even if some of them (e.g., temperate bacteriophages) can also on occasion be transmitted vertically within host genomes. In typical definitions of a virus, there is no mention of what the genome must encode, on the assumption that for every virus there is a host that is permissive for productive infection, from which progeny virions will result. This is impossible for polydnaviruses. Much like the recombinant viral vectors used in gene therapy experiments, polydnaviruses are crippled, and so can rightly be viewed as naturally occurring gene-delivery vehicles, or even wasp organelles (*11*). If we do not wish to consider polydnaviruses as bona fide viruses, then the standard definition of a virus must specify that the packaged genome encodes all functions required for viral replication. Finally, it is ironic that in all 1259 pages of the most recent compendium on virus taxonomy (*12*), "virus" is left undefined.

We would suggest that the more interesting lesson here for virologists and for evolutionary biologists may be that there is now reason to start thinking about virus-host relationships in much broader terms, so as to include not only mutualism, which may be a lot more common than previously contemplated (*13*, *14*), but also obligatory mutualism, as exemplified by the wasp-nudivirus story. How did this kind of relationship arise? Parasitoid larvae feeding within their hosts are exposed to a variety of viruses and will likely become infected by some. Moreover, the parasitoid ovary may represent a permissive environment for maintaining persistent virus infections (*15*). It may be that bilateral gene transfer has preserved some elements of these that were of mutual benefit.

References

1. E. Espagne *et al.*, *Science* **306**, 286 (2004).
2. A. Bézier *et al.*, *Science* **323**, 926 (2009).
3. S. Rotheram, *Nature* **214**, 700 (1967).
4. P. J. Krell, D. B. Stoltz, *J. Virol.* **29**, 1118 (1979).
5. P. J. Krell, D. B. Stoltz, *Virology* **101**, 408 (1980).
6. D. B. Stoltz *et al.*, *Intervirology* **21**, 1 (1984).
7. K. M. Edson *et al.*, *Science* **211**, 582 (1981).
8. W. N. Norton *et al.*, *Cell Tissue Res.* **162**, 195 (1975).
9. D. B. Stoltz, S. B. Vinson, *Adv. Virus Res.* **24**, 125 (1979).
10. Y. Wang *et al.*, *Arch. Virol.* **152**, 519 (2007).
11. B. A. Federici, Y. Bigot, *J. Insect Physiol.* **49**, 419 (2003).
12. C. M. Fauquet *et al.*, Eds., *Virus Taxonomy*: VIIIth Report of the *International Committee on Taxonomy of Viruses* (Elsevier, London, 2005).
13. J. J. Barondess, J. Beckwith, *Nature* **346**, 871 (1990).
14. F. Arnaud *et al.*, *PLoS Pathog.* **3**, 1716 (2007).
15. D. B. Stoltz, J. B. Whitfield, *J. Hym. Res.* **1**, 125 (1992).

10.1126/science.1169808

PLANETARY SCIENCE

Seeing the Missing Half

Gregory A. Neumann[1] and Erwan Mazarico[1,2]

Results from the Japanese SELENE mission shed light on differences between the far and nearsides of the Moon.

The Moon is our closest planetary neighbor and the only extraterrestrial body to which humans have traveled, yet many questions about its origin and early history remain unanswered. Four papers published in this issue by scientific teams of the Japanese SELENE (Kaguya) mission (*1*–*4*) offer a new global view of the Moon that helps to elucidate how the Moon evolved to its present state.

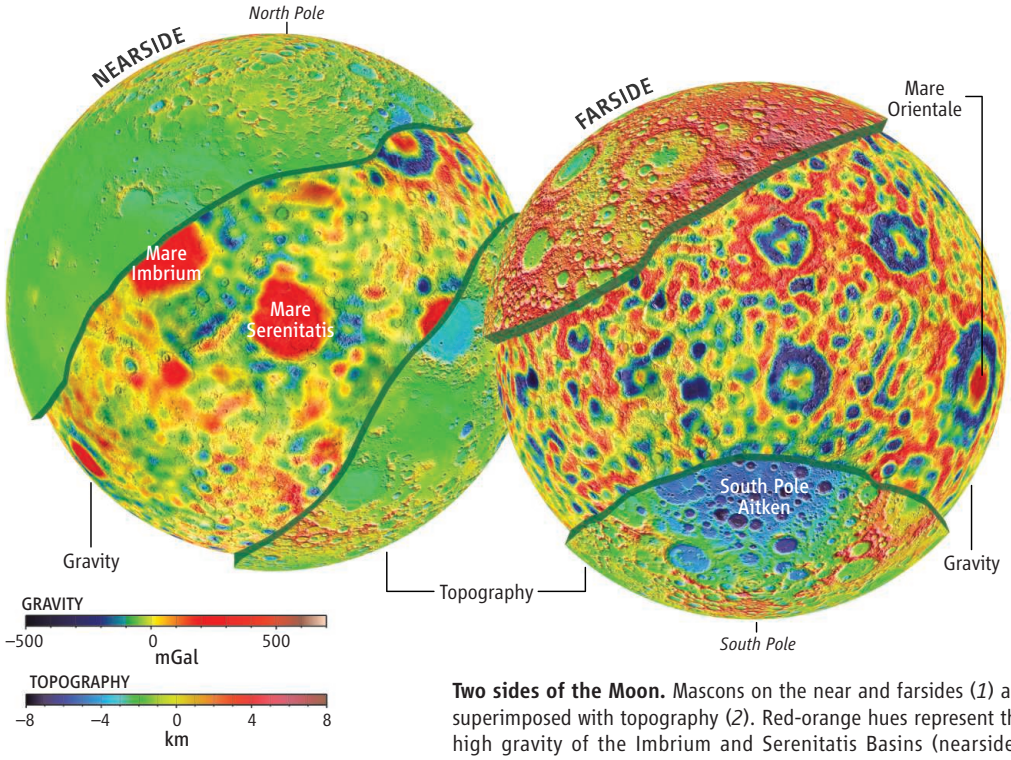
The Moon is lopsided: Its visible nearside (tidally locked to face the Earth) is covered with smooth, dark volcanic mare, whereas the far-side mainly consists of more heavily cratered, bright highland material. The differences in crustal thickness and density, apparent surface age, composition, and volcanic activity between the two sides are variously ascribed to external causes (such as a giant impact) or to internal causes (such as core formation, mantle convection, and crustal differentiation). The key to resolving these questions will be better data.

[1]NASA Goddard Space Flight Center, Code 698, Greenbelt, MD 20771, USA. [2] NASA Postdoctoral Program Fellow, Goddard Space Flight Center, Greenbelt, MD 20771 USA. E-mail: gregory.a.neumann@nasa.gov; erwan.m.mazarico@nasa.gov

The Apollo missions that ended in 1972 led to the current paradigm for the Moon's formation following a collision between the early Earth and a Mars-sized body (*5*). Analysis of lunar samples led to the hypothesis that the Moon was initially engulfed in a deep magma ocean and then differentiated to form a crust different from that of Earth. This crust subsequently hardened but was battered by meteorites during the late heavy bombardment that ended around 3.8 billion years ago, resulting in a surface covered by basins.

Some basins are as large as 2500 km across and 13 km deep (*6*), unlike anything on Earth. Their preservation indicates that the Moon's lithosphere formed rapidly and has since remained intact, but gives few clues to the present structure and thermal state of the lunar interior. The Apollo seismometers resolved the shallow crustal structure in a few locations, and detected deeper moonquakes whose origin remains elusive. A core-mantle boundary has not been detected, but data from more than 30 years of laser ranging to retroreflectors left on the lunar surface are consistent with a small liquid core (*7*).

The Moon's gravity field can be used to infer heterogeneities in the density of its interior. Large positive gravity anomalies associated with nearside mare basins were recognized very early from spacecraft radio tracking (*8*). Are these "mascons" (for mass concentrations) a result of loading by the volcanic mare deposits that filled the basins after their formation (*9*), or were they produced by dynamic mantle uplift during impact (*10*)? Given the intense heat produced by large impacts, preservation of the dynamic uplift seemed unlikely. But tracking data from the Clementine and Lunar Prospector missions in the 1990s showed the same mascon features with better spatial resolution and revealed new mascons not covered with visible mare (*11*, *12*).



Two sides of the Moon. Mascons on the near and farsides (*1*) are superimposed with topography (*2*). Red-orange hues represent the high gravity of the Imbrium and Serenitatis Basins (nearside). Green-blue hues represent gravity lows and topographic depressions such as the large South Pole–Aitken Basin (bottom right). Note the different character between the nearside mascons and the farside mascons with their large negative anomaly rings.

Unlike most nearside mascons, the mascon beneath Mare Orientale on the Western limb (see the figure, lower right) is surrounded by a ring of negative gravity anomaly; this feature could not be readily explained by loading of volcanic material (*13*). Speculation arose as to whether similar mascons, created by dynamic uplift, might exist beneath the nonmare-filled basins on the farside. However, no valid comparison of gravity anomalies on the near and farsides was possible, because lunar orbiters cannot be tracked by direct link from Earth when the orbiter traverses the farside.

The SELENE orbiter overcomes this problem by using a small companion satellite in a higher elliptical orbit to relay the radio signals of the main orbiter to and from Earth. On page 900, Namiki *et al.* (*1*) present a dramatically improved gravity model that includes the first farside tracking data. The results reveal farside mascons with small central gravity highs and wider rings of negative gravity anomaly that do not resemble their nearside counterparts (see the figure). The authors interpret the new mascons as indicating cooler, more rigid early conditions on the farside than previously thought.

As a further means to understand the structure of the basins, the expected contribution of topography to the gravity signal can be removed to reveal variations in crustal thickness and mantle topography under the basins. Altimetry data from the 1994 Clementine mission revealed a global dichotomy between the crustal thicknesses of the nearside basins and those of the farside highlands. However, the resolution was too low to elucidate geophysical processes such as the formation of large basins (like Mare Orientale) and the subsequent volcanic activity leading to present-day mare.

On page 897, Araki *et al.* report global topographic mapping of the Moon with the laser altimeter onboard SELENE (*2*). The data offer 100-fold higher density and better vertical resolution than previously available. Spectral analysis of the topography shows more power at short wavelengths than the Earth. This could indicate that, similarly to Venus and Mars, the Moon's mantle is depleted in volatiles such as water. As shown by the use made of Martian laser altimeter data (*14*), new topography models of the Moon will be invaluable for geophysical lunar research.

Mascon basins commonly display associated tectonic features, such as extensional linear channels and compressional mare ridges (*15*). On page 909, Ono *et al.* use data from the SELENE Lunar Radar Sounder (*3*) to map echoes from subsurface stratigraphy beneath nearside mare regions. They did not identify the 1.6-km-deep reflective interface in Serenitatis observed in early Apollo experiments (*16*), but discovered shallower reflectors. The authors infer from the stratigraphic thicknesses that the ridges seen in the most recent volcanic flows were deformed not only by mascon stresses arising during their emplacement, as previously thought, but by a regional compressive stress that occurred sometime later, probably during an episode of global cooling and contraction.

The Terrain Camera investigation described by Haruyama *et al.* on page 905 (*4*) has imaged volcanic flows with optimal illumination at high resolution, allowing farside mare basins to be dated based on cratering statistics (*17*). The results suggest that episodic farside volcanic activity continued until at least 2.5 billion years ago. Such late activity in the thicker farside crust is an important constraint for a comprehensive model of the Moon's history.

The new results (*1*–*4*) do not offer a complete picture of the Moon's evolution, but it is now clear that the farside lithosphere was stronger than the nearside at the time of the late heavy bombardment when the mascons were formed. More data should soon become available not only from SELENE, but also from the Chinese Chang'e-1, the Indian Chandrayaan-1, and, this spring, the NASA Lunar Reconnaissance Orbiter. The 2011 GRAIL (Gravity Recovery and Interior Laboratory) mission will map the lunar gravity with an accuracy at least three orders of magnitude greater than the Kaguya gravity mission, and the International Lunar Network stations planned for 2013 and afterward will carry a suite of geophysics experiments. It will then

be the turn of lunar modelers to determine which theory for the asymmetrical evolution best fits the data.

References and Notes

1. N. Namiki *et al.*, *Science* **323**, 900 (2009).
2. H. Araki *et al.*, *Science* **323**, 897 (2009).
3. T. Ono *et al.*, *Science* **323**, 909 (2009).
4. J. Haruyama *et al.*, *Science* **323**, 905 (2009).
5. A. G. W. Cameron, W. R. Ward, *Lunar Sci.* **7**, 120 (1976).
6. M. T. Zuber *et al.*, *Science* **266**, 1839 (1994).
7. J. G. Williams *et al.*, *J. Geophys. Res.* **106**, 27,933 (2001).
8. P. M. Muller, W. L. Sjogren, *Science* **161**, 680 (1968).
9. J. E. Conel, G. B. Holstrom, *Science* **162**, 1403 (1968).
10. D. U. Wise, M. Y. Yates, *J. Geophys. Res.* **75**, 261 (1970).
11. F. G. Lemoine *et al.*, *J. Geophys. Res.* **102**, 16,339 (1997).
12. A. S. Konopliv *et al.*, *Science* **281**, 1476 (1998).
13. G. A. Neumann *et al.*, *J. Geophys. Res.* **101**, 16,841 (1996).
14. D. E. Smith *et al.*, *J. Geophys. Res.* **106**, 23,689 (2001).
15. S. C. Solomon, J. W. Head, *Rev. Geophys. Space Phys.* **18**, 107 (1980).
16. R. J. Phillips *et al.*, in *Proceedings of the Fourth Lunar Science Conference*, R. Brett *et al.*, Eds. (Pergamon, New York, 1973), pp. 2821–2831.
17. J. W. Head, *Rev. Geophys. Space Phys.* **14**, 265 (1976).

18. E. M. was supported by an appointment to the NASA Postdoctoral Program at the Goddard Space Flight Center, administered by Oak Ridge Associated Universities through a contract with NASA.

10.1126/science.1170655

CLIMATE CHANGE

Sentinels of Change

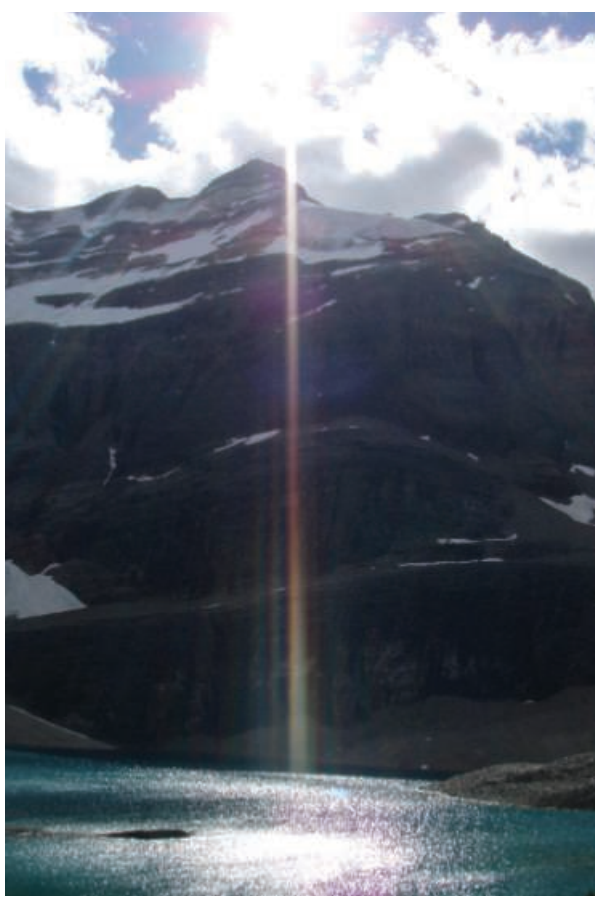
Lakes and reservoirs provide key insights into the effects and mechanisms of climate change.

Craig E. Williamson,[1] Jasmine E. Saros,[2] David W. Schindler[3]

Given the vast and complex landscape of Earth's surface, where do we look for signals of how climate change influences ecosystems? Lakes and reservoirs are an important part of the answer. Although they make up a small percentage of Earth's surface, lakes and reservoirs act as sentinels by providing signals that reflect the influence of climate change in their much broader catchments (*1*). The sediments of inland waters integrate these signals over time, and the deposition of terrestrially derived carbon and outgassing of greenhouse gases make them hot spots of carbon cycling and, thus, climate change regulation. Furthermore, given that freshwater is one of Earth's resources most jeopardized by changing climate (*2*, *3*), being able to detect changes that are detrimental to water quantity or quality is critically important.

Lake levels are highly visible signals of change in water quantity, with records that can span many decades (*4*). For instance, declining levels in closed-basin prairie lakes in North America indicate that at some time in this century, climate change may render these areas much drier than they have been since the late 19th century (*5*). Many of these changes are caused by long-term, cyclic climatic changes driven by ocean circulation patterns (*4*). For example, levels in smaller lakes in Wisconsin rise and fall with those of the nearby Laurentian Great Lakes, suggesting climate oscillations as a common driver (*6*). Lake sediments also store signals that are proxies for prehistoric climate change. The height and position of beach ridges that can be dated are similarly useful in deducing the state of water supplies in past millennia.

Some of these indicators are worrisome. For example, the Laurentian Great Lakes are often viewed as vast reservoirs for solving future water shortages in other parts of North America. But recent studies show that their water balance is precarious, with water renewal rates of less than 1% per year (*7*). In much of North America, the 20th century now appears to have been the wettest in the past millennium or more, with droughts up to several decades long commonplace in earlier centuries (*8*). Water levels in the Great Lakes have been fairly stable in the past century, varying by less than 2.1 m, but proxy data suggest that climate-driven hydrologic imbalance led the connecting rivers and Niagara Falls to dry up between 8770 and 8290 years ago, leaving the Great Lakes as separate basins (*7*). Pollen and seeds in dated lake sediments have shown that between 3000 and 8000 years ago, grasslands around Lake Winnipeg extended well north of the present tree line; the climate in the catchment seems to have been very similar to 20th-century Medicine Hat, Alberta, one of the driest parts of the Canadian Prairies (*9*).



CREDIT: CRAIG E. WILLIAMSON/MIAMI UNIVERSITY

Lakes as sentinels. Alpine lakes, such as Lake Oesa in the Canadian Rockies pictured here, produce and store signals of climate change. Scientists are using these signals to determine how climate influences both terrestrial and aquatic ecosystems and to elucidate the role of inland waters in regulating climate change.

Other recent studies have elucidated the role of lakes as regulators in the global carbon cycle. Over 90% of the estimated 304 million lakes worldwide are small (less than 0.01 km^2) (*10*) and shallow, with ample nutrients, light, and water to make them optimal environments for high levels of biological productivity. Much of the carbon reaching lakes originates in the vegetation and soils of their catchments. About twice

[1]Department of Zoology, Miami University , Oxford, OH 45056, USA. E-mail: craig.williamson@muohio.edu [2]Climate Change Institute and School of Biology and Ecology, University of Maine, Orono, ME 04469, USA. E-mail: jasmine.saros@maine.edu [3]Department of Biological Sciences, University of Alberta Edmonton, Alberta, Canada T6G 2E9. E-mail: d.schindler@ualberta.ca

as much terrestrial carbon is received by inland waters as reaches the world's oceans (*11*). Calcareous catchments also release considerable carbon to lakes as bicarbonate and carbonate. In total, inland waters may bury about four times as much carbon as do the oceans (*10*).

Much of the organic carbon received by lakes is mineralized and emitted to the atmosphere as carbon dioxide or methane. According to recent studies, inland waters—including eutrophic reservoirs and saline lakes as well as running waters—emit roughly as much carbon as is absorbed by the world's oceans (*10*, *12*). These contributions are given little consideration in global climate models. Including lakes and reservoirs in models may shift estimates for many landscapes to greater sources of carbon dioxide.

The dissolved organic carbon originating from terrestrial systems is usually colored and serves important functions in lakes. For example, it attenuates both damaging ultraviolet radiation and the longer wavelength solar radiation that regulates the thermal structure and physical habitat of the lake, thus changing nutrient and contaminant cycling. Dissolved organic carbon can be flocculated, contributing to carbon sedimentation, or mineralized through photolytic and biological processes (*13*). Widespread regional changes in dissolved organic carbon concentrations in inland waters remain poorly understood (*14*).

The biodiversity of many freshwater habitats is gravely threatened. During the 20th century in North America alone, 123 species of freshwater animals were recorded as extinct, with 49% of mussels, 23% of gastropods, 33% of crayfishes, 26% of amphibians, and 21% of freshwater fishes imperiled by the end of the century (*15*). The Laurentian Great Lakes are in the process of an "invasional meltdown" (*16*), due largely to species imported in the ballast water of Eurasian ships. Reservoirs and other human-made impoundments, including constructed ponds and wetlands, are 2.4 to 300 times as likely to harbor invasive species as natural lakes (*17*). They accelerate the spread of invasive species by decreasing the distance to the nearest "stepping stone" of water (*17*). The Three Gorges Dam has allowed the spread of 55 invasive species throughout its 58,000 km^2 catchment, including the water hyacinth, regarded by many as the world's worst invasive species (*18*).

Lakes also serve a sentinel function for many of the changes to forests and wetlands caused by a warming climate. For example, a 74 to 118% increase in the area burned by forest fires in Canada is predicted to result from increasingly severe dry weather by the end of the century (*19*). After a fire in the catchment of Moab Lake, Jasper National Park, Canada, increased nutrients and mercury caused changes in food web structure that resulted in greatly increased mercury concentrations in fish (*20*).

The outlook for lakes and reservoirs and the ecosystem services that they provide is bleak. Yet records from these inland waters may provide the insights necessary to address the dual challenges of climate change and increased human domination and their effects on lakes and the larger landscape. Global lake observatory networks that monitor and integrate these signals are needed in combination with experimental studies if we are to decipher all the information contained in the waters and sediments of lakes.

References and Notes

1. C. E. Williamson, W. Dodds, T. K. Kratz, M. A. Palmer, *Frontiers Ecol. Environ.* **6**, 247 (2008).
2. Millennium Ecosystem Assessment, 2005; www.millenniumassessment.org/en/Global.aspx
3. Intergovernmental Panel on Climate Change (IPCC) Fourth Assessment Report; see www.ipcc.ch/ipccreports/assessments-reports.htm
4. D. A. Wilcox, T. A. Thompson, R. K. Booth, J. R. Nicholas, *Lake Level Variability and Water Availability in the Great Lakes* (U.S. Geological Survey Circular 1311, 2007).
5. G. Van der Kamp, D. Keir, M. S. Evans, *Can. Water Resour. J.* **33**, 23 (2008).
6. C. A. Stow, E. C. Lamon, T. K. Kratz, C. E. Sellinger, *EOS* **89**, 389 (2008).
7. C. F. M. Lewis *et al.*, *Eos* **89**, 541 (2008).
8. C. A. Woodhouse, J. T. Overpeck, *Bull. Am. Meteorol. Soc.* **79**, 2693 (1998).
9. C. F. M. Lewis *et al.*, *Geology* **29**, 743 (2001).
10. J. A. Downing *et al.*, *Global Biogeochem. Cycles* **22**, GB1018 (2008).
11. J. J. Cole *et al.*, *Ecosystems* **10**, 171 (2007).
12. C. Duarte *et al.*, *J. Geophys. Res.* **113**, GO4141 (2008).
13. E. von Wachenfeldt, S. Sobek, D. Bastviken, L. Tranvik, *Limnol. Oceanogr.* **53**, 2416 (2008).
14. D. T. Monteith *et al.*, *Nature* **450**, 537 (2007).
15. A. Ricciardi, J. B. Rasmussen, *Conserv. Biol.* **13**, 1220 (1999).
16. A. Ricciardi, *Can. J. Fish. Aquat. Sci.* **58**, 2513 (2001).
17. P. T. J. Johnson, J. D. Olden, M. J. V. Zanden, *Frontiers Ecol. Environ.* **6**, 359 (2008).
18. J. Q. Ding, R. N. Mack, P. Lu, M. X. Ren, H. W. Huang, *Bioscience* **58**, 317 (2008).
19. M. D. Flannigan *et al.*, *Clim. Change* **72**, 1 (2005).
20. E. N. Kelly *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **103**, 19380 (2006).

21. Support from the NSF (NSF DEB-IRCEB-0552283) is acknowledged.

10.1126/science.1169443

PHYSICS

Fast Electrons Tie Quantum Knots

Special relativity and quantum physics combine to generate unusual arrangements of electron spins in two different solids.

Jan Zaanen

Objects traveling near the speed of light bring space travel to mind, but in familiar solids containing heavy elements (such as in a gold ring), the velocities of electrons are high enough that effects arising through special relativity come into play. The main effect for electrons is that the orientation of their spins depends on their trajectories. This effect, called spin-orbit coupling, seemed to be understood thoroughly by the 1960s, but new theoretical and experimental studies are driving its revival. Two reports in this issue discuss how spin-orbit coupling causes novel forms of electronic organization that can be captured by topology, the branch of mathematics that classes objects on the basis of properties that do not change upon deformation (and so lumps coffee cup handles together with doughnuts).

On page 915, Mühlbauer *et al.* (*1*) report the most complex form of magnetic order yet observed: In the presence of a magnetic field, the electron spins in manganese silicide (MnSi) form a lattice of topological "particles" called skyrmions (see the figure, panels A and B). On page 919, Hsieh *et al.* (*2*), using spin-resolved photoemission, find a superficially similar picture in the orientations of the electron spins in the metallic surface of bismuth antimonide ($Bi_{1-x}Sb_x$) (see the figure, panel C). However, these patterns now occur in the space of wave vectors of the electron wave functions associated with this metallic surface. Their shapes and topology are dictated by a bulk insulating state. Unlike ordinary insulators, this bulk is a macroscopic object that carries a net quantum entanglement, referring to the eerie property

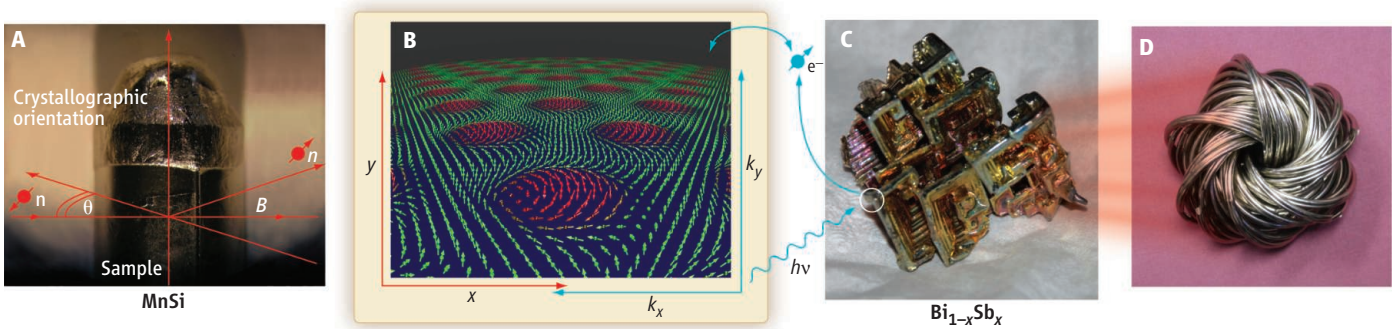
Lorentz Institute for Theoretical Physics, Leiden University, 2300 RA Leiden, Netherlands. E-mail: jan@lorentz.leidenuniv.nl

of the quantum world where every state is influenced by every other state in a way that has no counterpart in the realm of everyday experience. This bulk entanglement in turn dictates how the surface quantum states are linked together.

Relativistic effects in quantum mechanics were first treated with the Dirac equation, and electron spin was discovered through its solution. The aspect of relativity that matters for spin-orbit coupling is the unification of electricity and magnetism by the principle of relative motion. When a magnetic dipole, such as the electron spin, moves relative to an electrical field at rest, it experiences the latter as a magnetic field. As a result the spins will precess, that also maximizes magnetic ordering. Such helical magnetism forms spontaneously from a paramagnetic phase in MnSi at temperatures just below 30 K. However, when a magnetic field is applied (between 0.1 and 0.2 T), a mysterious "A phase" is found (*3*). Using neutron scattering, Mühlbauer *et al*. have now resolved the spin structure of this phase: It is the "skyrmion lattice" (see the figure, panel B). The reasons for its stability are complicated and include help from thermal fluctuations, but the key feature is that the red regions in panel B of the figure (the skyrmions) crystallize as if they were atoms.

Skyrmions were introduced in 1962 by the high-energy physicist Skyrme (*4*). Topology quantum computing. Typically, only a few microscopic degrees of freedom can be entangled because under normal circumstances, the contact with the classical macroscopic world will destroy the entanglement long before the quantum system itself becomes macroscopically large. However, in $Bi_{1-x}Sb_x$, topological effects help protect its net entanglement against collapse.

To see how strange this state is, we need to compare it with conventional insulators. The electron waves in crystals diffract against the potentials formed by the lattice to create energy bands; in an insulator, there are no electronic states at the Fermi energy because an energy gap separates the valence and con-



Topological knots. In the metal MnSi (**A**), a form of magnetic order, probed by neutron scattering (neutrons *n* scattering by angle θ in a magnetic field *B*), occurs in real space (*x* and *y*)that can be viewed as a lattice of topological particles called skyrmions, which are the red lumps in (**B**). In the "topological" insulator $Bi_{1-x}Sb_x$ (**C**), the entanglement of the individual electrons, probed by photoemission, adds up to an overall macroscopic quantum state whose topology is like that of metal rings (**D**) (*7*), having the effect that the spins order in a pattern similar to the skyrmions (see text) with wave vectors k_x and k_y describing the surface electron waves (*2*).

and this eventually causes the phenomena reported by Mühlbauer *et al.* and Hsieh *et al.*

The magnetism exhibited by MnSi is unusual but can be described in a physically intuitive way. Magnetism in metals, such as iron, can occur spontaneously through the ordering of its unpaired electron spins. Electrons try to move as freely as possible, but they also repel each other and hinder each other's motions. Also, the Pauli exclusion principle forbids two electrons to occupy the same quantum state, so electrons automatically avoid each other when their spins point in the same direction, thereby forming a ferromagnet. However, unlike a metal such as iron, the crystal structure of MnSi lacks inversion symmetry. This symmetry breaking adds the effect of spin-orbit coupling, and the spins of the electrons now precess. However, this motion can be organized in such a way that a frozen-out magnetic structure forms.

The simplest structure of this kind is the helical magnet, in which the spins precess around a preferred axis. For particular values of the pitch of this spin helix, the precessional motions neatly synchronize in a static pattern classes objects together if they can be smoothly morphed into the same shape, so the hole in a doughnut is the "same" as the handle on a coffee cup. The skyrmion is such a "handle," but rather than forming in a ceramic, it forms in vectors that can point in arbitrary directions in space, such as the magnetization of MnSi. The red regions can be removed only by coordinating the rotations of a huge number of spins, which is highly improbable. Because of this integrity, the skyrmions are said to be "topological particles." Like atoms in a crystal, they can be stacked in a lattice, which is how the magnetic order revealed by Mühlbauer *et al.* should be understood.

Physical intuition is harder to come by in understanding the topological band insulator of Hsieh *et al*., because of the remarkable phenomenon that a macroscopic piece of bismuth antimonide carries a net entanglement. Entanglement refers to the "spooky" quantum phenomenon that correlates the information in quantum states together in ways that make it possible to compute exponentially faster than with classical states; this is the idea behind duction bands. These can be viewed as an electronic "nothingness" that is quite akin to the fundamental vacuum of the Dirac theory. Two years ago, however, two groups predicted theoretically (*5*, *6*) that this "insulating nothingness" can have topological structure. Resting on quite fanciful topology (see the figure, panel D) (*7*), the quantum entanglements of all occupied electron states can combine in one overall topological quantity, forming the "topological insulator" that carries a global entanglement (*5*, *6*).

Despite the spooky nature of this entanglement, it can still be probed experimentally. The surface of the crystal corresponds with the interface between the normal fundamental vacuum and the topological insulating bulk of the crystal. The entanglement topology of the bulk now dictates that, at the surface, the band gap of the insulator has to fill up with special states, which causes the surface to become a special "helical metal" (*5*, *6*). In this case, helical refers to the odd number of electron bands crossing the Fermi energy of the metal that forms at the surface; in a normal metal, this number is always even.

The special nature of the single-electron wave functions required for topological insulators arises naturally in insulators that have small band gaps and strong spin-orbit coupling because they contain heavy atoms. Experimental evidence for the strange surface states was found shortly after their prediction (*8*) in transport measurements in a device containing a thin layer of HgTe forming a "quantum spin Hall insulator," the two-dimensional version of the topological insulator (*9*). Shortly thereafter, it was realized that topological insulators could form in three dimensions (*6*), and the heavy semimetal bismuth, alloyed with antimony to turn it into a small-gap insulator, was identified as a candidate (*5*). Transport measurements on the surface of a crystal are very difficult, but angle-resolved photoemission can image the surface electron bands directly. Last year, Hsieh *et al*. (*10*) showed that there are an uneven number of surface bands crossing the Fermi energy.

Spin-orbit coupling lies at the heart of the topological insulator, but how does this relate to the effects of relativity discussed in the context of a MnSi skyrmion lattice? Electrical fields are present at the $Bi_{1-x}Sb_x$ crystal surface, but these will not give rise to magnetism. However, under the influence of the topological bulk, the surface spins do not order in physical position space, but rather in the space formed by the wave vectors of the quantum waves describing the electrons moving on the surface. This two-dimensional wave vector space repeats periodically, and because the surface is metallic, it contains a periodic array of Fermi "surfaces" enclosing the regions with occupied states.

When the bulk is a topological insulator, the remarkable coincidence is that the skyrmion lattice described by Mühlbauer *et al.* forms an acceptable cartoon of what this "magnetism in wave vector space" looks like. The skyrmions are now regions of occupied states, and their rims are the Fermi surfaces. The spins at the Fermi energy are precisely oriented as the whirls formed by the "golden" spins.

However, the cartoon is not a literal description as electron energies move away from the Fermi energy. The whirl-like arrangement of the Fermi surface spins should actually persist both for the occupied and unoccupied states, with the spins slowly vanishing upon moving away from the Fermi surface. Using spin-resolved photoemission, Hsieh *et al.* observe precisely this "wave vector space magnetism," which is direct evidence that $Bi_{1-x}Sb_x$ is a topological insulator.

The discovery by Mühlbauer *et al*. that spins can order in the form of a lattice of topological particles confirms that skyrmions indeed can behave like atoms and opens up new avenues of research related to electrical transport, especially in relation to the very strange metallic states found in MnSi when it is put under pressure. Hsieh *et al*. show that a simple alloy of bismuth and antimony allows us to hold something very nonintuitive—a macroscopic quantum entangled state—in the palms of our hands, and the theorists continue to suggest new ideas for experimental study. The electrodynamics of topological insulators is also quite strange: When an electrical charge is brought to the surface, it will bind automatically to a magnetic monopole formed in the bulk, and this "dyon" should behave like a particle with fractional quantum statistics (*11*). Alternatively, when a superconductor is brought into contact with a topological insulator, its magnetic vortices are predicted to turn into particles that can be used for topological quantum computing (*12*).

References

1. S. Mühlbauer *et al*., *Science* **323**, 915 (2009).
2. D. Hsieh *et al*., *Science* **323**, 919 (2009).
3. Y. Ishikawa, M. Arai, *J. Phys. Soc. Jpn*. **53**, 2726 (1984).
4. T. H. R. Skyrme, *Nucl. Phys*. **31**, 556 (1962).
5. L. Fu, C. L. Kane, E. J. Mele, *Phys. Rev. Lett*. **98**, 106803 (2007).
6. J. E. Moore, L. Balents, *Phys. Rev. B* **75**, 121306 (2007).
7. J. E. Moore, Y. Ran, X. G. Wen, *Phys. Rev. Lett*. **101**, 186805 (2008).
8. B. A. Bernevig, T. L. Hughes, S.-C. Zhang, *Science* **314**, 1757 (2006).
9. M. König *et al*., *Science* **318**, 766 (2007); published online 19 September 2007 (10.1126/science.1148047).
10. D. Hsieh *et al*., *Nature* **452**, 970 (2008).
11. X.-L. Qi, R. Li, J. Zang, S.-C. Zhang, *Science*, in press; published online 29 January 2009 (10.1126/science.1167747).
12. L. Fu, C. L. Kane, *Phys. Rev. Lett*. **100**, 096407 (2008).

10.1126/science.1169344

NEUROSCIENCE

Pains and Pleasures of Social Life

Analyses of brain activity reveal a link between social and physical pains and pleasures.

Matthew D. Lieberman and Naomi I. Eisenberger

Life is full of complex social events such as being accepted or rejected, treated fairly or unfairly, and esteemed or devalued by others. Our responses to these events depend primarily on our psychological interpretation of them, in contrast to events like spraining an ankle or eating chocolate, for which our responses seem more dependent on the physical acts themselves. Nevertheless, our emotional responses to these psychological events rely on much of the same neural circuitry that underlies the simplest physical pains and pleasures. On page 937 of this issue, Takahasi *et al*. (*1*) show that experiencing envy at another person's success activates pain-related neural circuitry, whereas experiencing schadenfreude—delight at someone else's misfortune—activates reward-related neural circuitry.

Neuroscientists have identified neural systems responsible for experiences of pain and pleasure. The cortical pain network consists primarily of the dorsal anterior cingulate cortex (dACC), insula, and somatosensory cortex, with subcortical contributions from the periaqueductal gray and thalamus (*2*) (see the figure). Whereas the somatosensory cortex is associated with sensory aspects of cutaneous physical pain (e.g., its location on the body), the dACC is associated with the distressing aspect of pain.

The brain's reward circuitry (see the figure) consists of neural structures receiving the neurotransmitter dopamine from the ventral tegmental area, and responds to physically rewarding stimuli such as food, drugs, and sexual activity. The nucleus accumbens in ventral striatum plays a critical role in reward learning and pleasurable states, while the ventromedial prefrontal cortex and amygdala are also major dopaminergic targets that have been implicated in reward processes (*3*).

Although it is expected that these networks produce robust responses to physical pains and pleasures, it is surprising that social pains and pleasures activate these same networks. For example, being socially excluded activates the dACC and insula, with the dACC showing greater activity to the extent that an individual feels greater social pain (*4*). Grieving over the death of a loved one and being treated unfairly also activate these regions (*5*, *6*). Alternatively, social rewards

Department of Psychology, 1285 Franz Hall, University of California at Los Angeles, Los Angeles, CA 90095–1563, USA. E-mail: lieber@ucla.edu

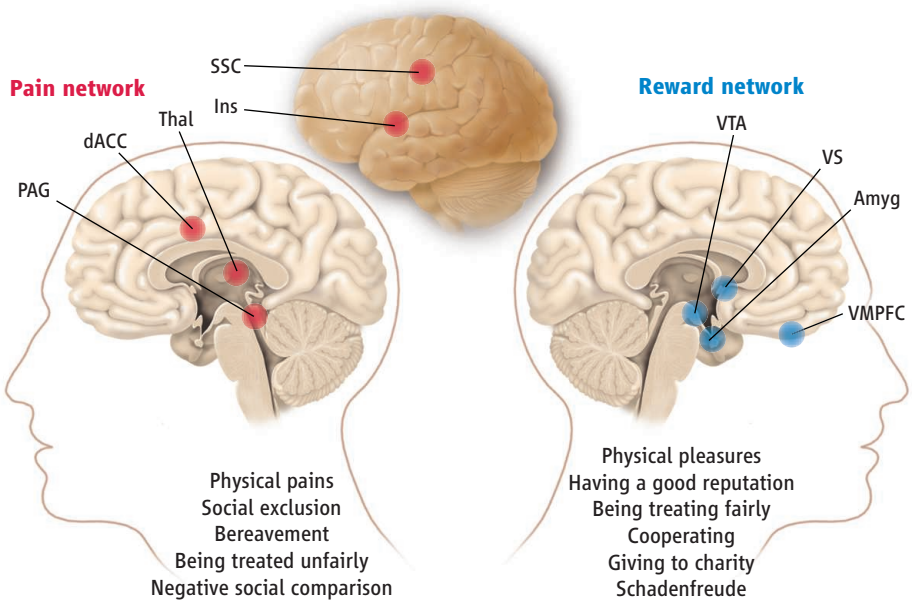
activate the same reward network as desirable foods and drinks. Having a good reputation, being treated fairly, and being cooperative all activate the ventral striatum (*7*–*9*). Strikingly, making charitable donations activates the reward network more than receiving the same sum of money for oneself (*10*).

Although most would describe being excluded as painful and giving to charity as pleasurable, the connotations of these descriptions change as we learn that these experiences activate the same brain regions that respond to physical pains and pleasures. Such findings suggest that the brain may treat abstract social experiences and concrete physical experiences as more similar than is generally assumed.

These overlaps suggest that certain social psychological concerns may have the same motivational importance as other physical survival needs. For every state of deprivation associated with a particular need, there is a pain. Lack of food begets hunger, lack of water begets thirst, and lack of shelter begets thermal discomfort. Each of these pains motivates us to seek out the salve that will take the pain away and satisfy the underlying need. The process of satisfying such needs is pleasurable and rewarding. All basic survival needs share these dynamics between need deprivation and pain and between need satiation and pleasure. Moreover, for physical survival needs, the greater the deprivation and attendant pain, the more pleasurable it is to satisfy the need (e.g., food tastes better on an empty stomach).

Takahashi *et al*. demonstrate, for the first time, this dynamic interplay between social pains and pleasures. If maintaining one's social value is a need like other physical needs, then the greater the pain caused by negative social comparisons, the greater the pleasure in response to seeing the comparison target socially devalued (schadenfreude). The authors found that greater envy and dACC activity in response to a negative social comparison was associated with greater schadenfreude and ventral striatum activity when learning of that comparison target's subsequent downfall.

Given that physical needs intuitively seem more critical to survival than social needs, why would the brain have evolved to treat them as motivationally similar? It is clear why food and water are needed and why their deprivation causes pain. But why use the neural system for physical pain to deal with social pains? One critical determinant may be the dependence of mammalian newborns on others for survival. Because newborn mammals are relatively immature—incapable of securing food, water, and shelter for themselves—they and the survival of their species depend on an ongoing bond between caregiver and infant (*11*). For both caregiver and infant to feel pain upon separation ensures social connection and thus offspring survival. In a sense, for mammalian infants, social needs take precedence over physical needs because meeting the social needs is what allows the physical needs to be met as well.



The pain and pleasure systems. The pain network consists of the dorsal anterior cingulate cortex (dACC), insula (Ins), somatosensory cortex (SSC), thalamus (Thal), and periaqueductal gray (PAG). This network is implicated in physical and social pain processes. The reward or pleasure network consists of the ventral tegmental area (VTA), ventral striatum (VS), ventromedial prefrontal cortex (VMPFC), and the amygdala (Amyg). This network is implicated in physical and social rewards.

In addition to the caregiver-infant bond, connections to a social group also promote survival. When responsibility for food acquisition, protection from predators, and care for offspring are distributed among group members (rather than being the sole responsibility of a single individual), individual group members are more likely to survive (*12*). Being fair, cooperative, or charitable may increase the survival of the group and thus one's own offspring. Moreover, group members who are not cooperative are more likely to be ostracized, which greatly lowers chances of survival (*13*). Thus, evolutionary pressures may have created internal mechanisms that register being socially cooperative as pleasurable and being ostracized as painful in order to promote the maintenance of group bonds and ensure survival.

The link between social and physical pains and pleasures adds to the growing chorus of neurocognitive findings that point to the critical importance of the social world for the surviving and thriving of humans. It seems noncoincidental that the size of the prefrontal cortex correlates with the size of social groups across primate species (*14*), that there is a dedicated neurocognitive network for social cognition that is preferentially activated when the mind is at rest (*15*), and that social and physical needs rely on shared neural substrates. Our attentiveness to the social world may sometimes seem like a diversion from more concrete concerns, but increasingly, neuroscience is revealing ways in which such attention is actually an adaptive response to some of our most vital concerns.

References

1. H. Takahashi *et al*., *Science* **323**, 937 (2009).
2. D. D. Price, *Science* **288**, 1769 (2000).
3. K. C. Berridge, M. L. Kringelbach, *Psychopharmacology* **199**, 457 (2008).
4. N. I. Eisenberger, M. D. Lieberman, K. D. Williams, *Science* **302**, 290 (2003).
5. M. F. O'Connor *et al*., *Neuroimage* **42**, 969 (2008).
6. A. G. Sanfey, J. K. Rilling, J. A. Aronson, L. E. Nystrom, J. D. Cohen, *Science* **300**, 1755 (2003).
7. K. Izuma, D. N. Saito, N. Sadato, *Neuron* **58**, 284 (2008).
8. G. Tabibnia, A. B. Satpute, M. D. Lieberman, *Psychol. Sci.* **19**, 339 (2008).
9. J. K. Rilling *et al*., *Neuron* **35**, 395 (2002).
10. J. Moll *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **103**, 15623 (2006).
11. J. Bowlby, *Attachment and Loss* (New York, Basic Books, 1969), vol. 1.
12. R. F. Baumeister, M. R. Leary, *Psychol. Bull.* **117**, 497 (1995).
13. K. D. Williams, *Annu. Rev. Psychol.* **58**, 425 (2007).
14. R. I. M. Dunbar, *Evol. Anthropol.* **6**, 178 (1998).
15. M. E. Raichle *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **98**, 676 (2001).

10.1126/science.1170008

Network Analysis in the Social Sciences

Stephen P. Borgatti, Ajay Mehra, Daniel J. Brass, Giuseppe Labianca

Over the past decade, there has been an explosion of interest in network research across the physical and social sciences. For social scientists, the theory of networks has been a gold mine, yielding explanations for social phenomena in a wide variety of disciplines from psychology to economics. Here, we review the kinds of things that social scientists have tried to explain using social network analysis and provide a nutshell description of the basic assumptions, goals, and explanatory mechanisms prevalent in the field. We hope to contribute to a dialogue among researchers from across the physical and social sciences who share a common interest in understanding the antecedents and consequences of network phenomena.

One of the most potent ideas in the social sciences is the notion that individuals are embedded in thick webs of social relations and interactions. Social network theory provides an answer to a question that has preoccupied social philosophy since the time of Plato, namely, the problem of social order: how autonomous individuals can combine to create enduring, functioning societies. Network theory also provides explanations for a myriad of social phenomena, from individual creativity to corporate profitability. Network research is "hot" today, with the number of articles in the Web of Science on the topic of "social networks" nearly tripling in the past decade. Readers of *Science* are already familiar with network research in physics and biology (*1*), but may be less familiar with what has been done in the social sciences (*2*).

History

In the fall of 1932, there was an epidemic of runaways at the Hudson School for Girls in upstate New York. In a period of just 2 weeks, 14 girls had run away— a rate 30 times higher than the norm. Jacob Moreno, a psychiatrist, suggested the reason for the spate of runaways had less to do with individual factors pertaining to the girls' personalities and motivations than with the positions of the runaways in an underlying social network (*3*). Moreno and his collaborator, Helen Jennings, had mapped the social network at Hudson using "sociometry," a technique for eliciting and graphically representing individuals' subjective feelings toward one another (Fig. 1). The links in this social network, Moreno argued, provided channels for the flow of social influence and ideas among the girls. In a way that even the girls themselves may not have been conscious of, it was their location in the social network that determined whether and when they ran away.

Moreno envisioned sociometry as a kind of physics, complete with its own "social atoms" and its laws of "social gravitation" (*3*). The idea of modeling the social sciences after the physical ones was not, of course, Moreno's invention. A hundred years before Moreno, the social philosopher Comte hoped to found a new field of "social physics." Fifty years after Comte, the French sociologist Durkheim had argued that human societies were like biological systems in that they were made up of interrelated components. As such, the reasons for social regularities were to be found not in the intentions of individuals but in the structure of the social environments in which they were embedded (*4*). Moreno's sociometry provided a way of making this abstract social structure tangible.

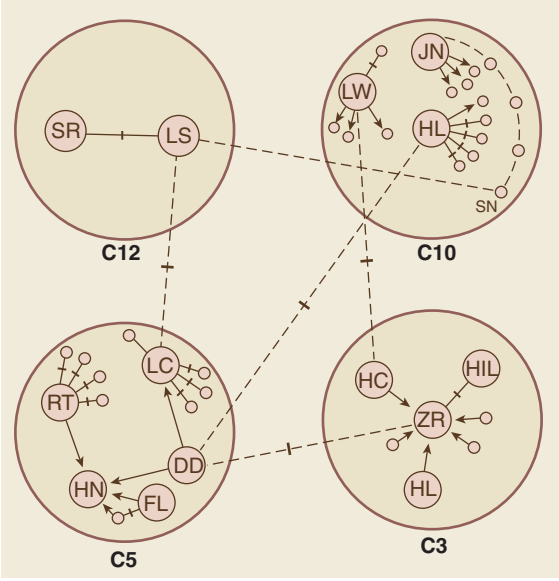


Fig. 1. Moreno's network of runaways. The four largest circles (C12, C10, C5, C3) represent cottages in which the girls lived. Each of the circles within the cottages represents an individual girl. The 14 runaways are identified by initials (e.g., SR). All nondirected lines between a pair of individuals represent feelings of mutual attraction. Directed lines represent one-way feelings of attraction.

In the 1940s and 1950s, work in social networks advanced along several fronts. One front was the use of matrix algebra and graph theory to formalize fundamental social-psychological concepts such as groups and social circles in network terms, making it possible to objectively discover emergent groups in network data (*5*). Another front was the development of a program of laboratory experimentation on networks. Researchers at the Group Networks Laboratory at the Massachusetts Institute of Technology (MIT) began studying the effects of different communication network structures on the speed and accuracy with which a group could solve problems (Fig. 2). The more centralized structures, such as the star structure, outperformed decentralized structures, such as the circle, even though it could be shown mathematically that the circle structure had, in principle, the shortest minimum solution time (*6*). Why the discrepancy? Achieving the mathematically optimal solution would have required the nodes to execute a fairly complex sequence of information trades in which no single node served as integrator of the information. But the tendency in human networks seemed to be for the more peripheral members of a network (i.e., the nodes colored blue in the "Star," "Y," and "Chain" networks in Fig. 2) to channel information to the most central node (i.e., the nodes colored red in Fig. 2), who then decided what the correct answer was and sent this answer back out to the other nodes. The fastest performing network structures were those in which the distance of all nodes from the obvious integrator was the shortest (*7*).

The work done by Bavelas and his colleagues at MIT captured the imagination of researchers in a number of fields, including psychology, political science, and economics. In the 1950s, Kochen, a mathematician, and de Sola Pool, a political scientist, wrote a highly circulated paper, eventually published in 1978 (*8*), which tackled what is known today as the "small world" problem: If two persons are selected at random from a population, what are the chances that they would know each other, and, more generally, how long a chain of acquaintanceship would be required to link them? On the basis of mathematical models, they speculated that in a population like the United States, at least 50% of pairs could be linked by chains with no more than two intermediaries. Twenty years later, Stanley Milgram tested their propositions empirically, leading to the now popular notion of "six degrees of separation" (*9*).

During this period, network analysis was also used by sociologists interested in studying the changing social fabric of cities. The common conviction at the time was that urbanization destroyed community, and that cities played a central role in this drama. These sociologists saw concrete relations between people—love, hate, support, and so

LINKS Center for Network Research in Business, Gatton College of Business and Economics, University of Kentucky, Lexington, KY 40506–0034, USA. E-mail: sborgatti@uky.edu (S.P.B.), ajay.mehra@uky.edu (A.M.), dbrass@uky.edu (D.J.B.), and joe.labianca@uky.edu (G.L.)

on—as the basic stuff of community, and they used network analysis to represent community structure. For example, researchers interviewed 1050 adults living in 50 northern Californian communities with varying degrees of urbanism about their social relations (*10*). The basic procedure for eliciting network data was to get respondents (egos) to identify people (alters) with whom they had various kinds of relationships and then to also ask ego about the relationships between some or all of the alters. They found that urbanism did in fact reduce network density, which, in turn, was negatively related to psychological measures of satisfaction and overall well-being. A similar study of 369 boys and 366 girls between the ages of 13 and 19 in a Midwestern town of about 10,000 residents found that the adolescents' behaviors were strongly influenced by the "cliques" to which they belonged (*11*). The representation and analysis of community network structure remains at the forefront of network research in the social sciences today, with growing interest in unraveling the structure of computer-supported virtual communities that have proliferated in recent years (*12*).

By the 1960s, the network perspective was thriving in anthropology. Influenced by the pioneering work of Radcliffe Brown (*13*), there were three main lines of inquiry. First, at the conceptual level, anthropologists like S. F. Nadel began to see societies not as monolithic entities but rather as a "pattern or network (or 'system') of relationships obtaining between actors in their capacity of playing roles relative to one another" (*14*). Second, building on the insights of the anthropologist Levi-Strauss, scholars began to represent kinship systems as relational algebras that consisted of a small set of generating relations (such as "parent of" and "married to") together with binary composition operations to construct derived relations such as "in-law" and "cousin." It was soon discovered that the kinship systems of such peoples as the Arunda of Australia formed elegant mathematical structures that gave hope to the idea that deep lawlike regularities might underlie the apparent chaos of human social systems (*15*, *16*).

Third, a number of social anthropologists began to use network-based explanations to account for a range of outcomes. For example, a classic ethnographic study by Bott (*17*) examined 20 urban British families and attempted to explain the considerable variation in the way husbands and wives performed their family roles. In some families, there was a strict division of labor: Husband and wife carried out distinct household tasks separately and independently. In other families, the husband and wife shared many of the same tasks and interacted as equals. Bott found that the degree of segregation in the role-relationship of husband and wife varies directly with the connectedness (or density) of the family's social network. The more connected the network, the more likely the couple would maintain a traditional segregation of husband and wife roles, showing that the structure of the larger network can affect relations and behaviors within the dyad.

In the 1970s, the center of gravity of network research shifted to sociology. Lorrain and White (*18*) sought ways of building reduced models of the complex algebras created when all possible compositions of a set of relations were constructed (e.g., the spouse of the parent of the parent of …). By collapsing together nodes that were structurally equivalent—i.e., those that had similar incoming and outgoing ties—they could form a new network (a reduced model) in which the nodes consisted of structural positions rather than individuals. This idea mapped well with the anthropologists' view of social structure as a network of roles rather than individuals, and was broadly applicable to the analysis of roles in other settings, such as the structure of the U.S. economy (*19*). It was also noted that structurally equivalent individuals faced similar social environments and therefore could be expected to develop similar responses, such as similar attitudes or behaviors (*20*).

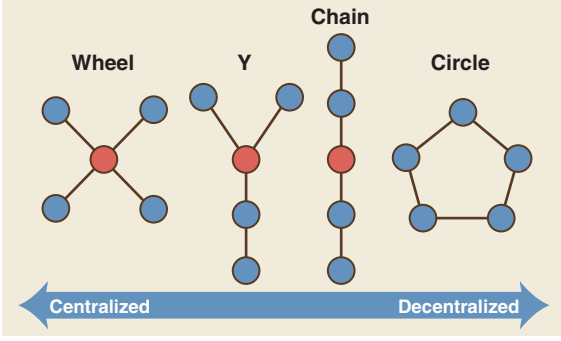


Fig. 2. Four network structures examined by Bavelas and colleagues at MIT. Each node represents a person; each line represents a potential channel for interpersonal communication. The most central node in each network is colored red.

Another key contribution was the influential strength of weak ties (SWT) theory developed by Granovetter (*21*). Granovetter argued that strong ties tend to be "clumpy" in the sense that one's close contacts tend to know each other. As a result, some of the information they pass along is redundant—what a person hears from contact A is the same as what the person heard from B. In contrast, weak ties (e.g., mere acquaintances) can easily be unconnected to the rest of one's network, and therefore more likely to be sources of novel information. Twenty years later, this work has developed into a general theory of social capital—the idea that whom a person is connected to, and how these contacts are connected to each other, enable people to access resources that ultimately lead them to such things as better jobs and faster promotions (*22*).

By the 1980s, social network analysis had become an established field within the social sciences, with a professional organization (INSNA, International Network for Social Network Analysis), an annual conference (Sunbelt), specialized software (e.g., UCINET), and its own journal (*Social Networks*). In the 1990s, network analysis radiated into a great number of fields, including physics and biology. It also made its way into several applied fields such as management consulting (*23*), public health (*24*), and crime/war fighting (*25*). In management consulting, network analysis is often applied in the context of knowledge management, where the objective is to help organizations better exploit the knowledge and capabilities distributed across its members. In public health, network approaches have been important both in stopping the spread of infectious diseases and in providing better health care and social support.

Of all the applied fields, national security is probably the area that has most embraced social network analysis. Crime-fighters, particularly those fighting organized crime, have used a network perspective for many years, covering walls with huge maps showing links between "persons of interest." This network approach is often credited with contributing to the capture of Saddam Hussein. In addition, terrorist groups are widely seen as networks rather than organizations, fueling research on how to disrupt functioning networks (*26*). At the same time, it is often asserted that it takes a network to fight a network, sparking military experiments with decentralized units.

Social Network Theory

Perhaps the oldest criticism of social network research is that the field lacks a (native) theoretical understanding—it is "merely descriptive" or "just methodology." On the contrary, there is so much of it that one of the main purposes of this article is to organize and simplify this burgeoning body of theory. We will give brief summaries of the salient points, using comparisons with the network approach used in the physical sciences (including biology).

Types of ties. In the physical sciences, it is not unusual to regard any dyadic phenomena as a network. In this usage, a network and a mathematical graph are synonymous, and a common set of techniques is used to analyze all instances, from protein interactions to coauthorship to international trade. In contrast, social scientists typically distinguish among different kinds of dyadic links both analytically and theoretically. For example, the typology shown in Fig. 3 divides dyadic relations into four basic types—similarities, social relations, interactions, and flows. Much of social network research can be seen as working out how these different kinds of ties affect each other.

The importance of structure. As in the study of isomers in chemistry, a fundamental axiom of social network analysis is the concept that structure matters. For example, teams with the same composition of member skills can perform very differently depending on the patterns of relationships among the members. Similarly, at the level of the individual node, a node's outcomes and

Similarities			Social Relations				Interactions	Flows
Location	Membership	Attribute	Kinship	Other role	Affective	Cognitive	e.g.,	e.g.,
e.g.,	e.g.,	e.g.,	e.g.,	e.g.,	e.g.,	e.g.,	Sex with	Information
Same spatial and temporal space	Same clubs	Same gender	Mother of	Friend of	Likes	Knows	Talked to	Beliefs
	Same events	Same attitude	Sibling of	Boss of	Hates	Knows about	Advice to	Personnel
	etc.	etc.		Student of	etc.	Sees as happy	Helped	Resources
				Competitor of		etc.	Harmed	etc.
							etc.	

Fig. 3. A typology of ties studied in social network analysis.

future characteristics depend in part on its position in the network structure. Whereas traditional social research explained an individual's outcomes or characteristics as a function of other characteristics of the same individual (e.g., income as a function of education and gender), social network researchers look to the individual's social environment for explanations, whether through influence processes (e.g., individuals adopting their friends' occupational choices) or leveraging processes (e.g., an individual can get certain things done because of the connections she has to powerful others). A key task of social network analysis has been to invent graph-theoretic properties that characterize structures, positions, and dyadic properties (such as the cohesion or connectedness of the structure) and the overall "shape" (i.e., distribution) of ties.

At the node level of analysis, the most widely studied concept is centrality—a family of node-level properties relating to the structural importance or prominence of a node in the network. For example, one type of centrality is Freeman's betweenness, which captures the property of frequently lying along the shortest paths between pairs of nodes (*27*). This is often interpreted in terms of the potential power that an actor might wield due to the ability to slow down flows or to distort what is passed along in such a way as to serve the actor's interests. For example, Padgett and Ansell (*28*) analyzed historical data on marriages and financial transactions of the powerful Medici family in 15th-century Florence. The study suggested that the Medici's rise to power was a function of their position of high betweenness within the network, which allowed them to broker business deals and serve as a crucial hub for communication and political decision-making.

Research questions. In the physical sciences, a key research goal has been formulating universal characteristics of nonrandom networks, such as the property of having a scale-free degree distribution. In the social sciences, however, researchers have tended to emphasize variation in structure across different groups or contexts, using these variations to explain differences in outcomes. For example, Granovetter argued that when the city of Boston sought to absorb two neighboring towns, the reason that one of the towns was able to successfully resist was that its more diffuse network structure was more conducive to collective action (*21*).

A research goal that the social and physical sciences have shared has been to explain the formation of network ties and, more generally, to predict a host of network properties, such as the clusteredness of networks or the distributions of node centrality. In the social sciences, most work of this type has been conducted at the dyadic level to examine such questions as: What is the basis of friendship ties? How do firms pick alliance partners? A host of explanations have been proposed in different settings, but we find they can usefully be grouped into two basic categories: opportunity-based antecedents (the likelihood that two nodes will come into contact) and benefit-based antecedents (some kind of utility maximization or discomfort minimization that leads to tie formation).

Although there are many studies of network antecedents, the primary focus of network research in the social sciences has been on the consequences of networks. Perhaps the most fundamental axiom in social network research is that a node's position in a network determines in part the opportunities and constraints that it encounters, and in this way plays an important role in a node's outcomes. This is the network thinking behind the popular concept of social capital, which in one formulation posits that the rate of return on an actor's investment in their human capital (i.e., their knowledge, skills, and abilities) is determined by their social capital (i.e., their network location) (*29*).

Unlike the physical sciences, a multitude of node outcomes have been studied as consequences of social network variables. Broadly speaking, these outcomes fall into two main categories: homogeneity and performance. Node homogeneity refers to the similarity of actors with respect to behaviors or internal structures. For example, if the actors are firms, one area of research tries to predict which firms adopt the same organizational governance structures (*30*); similarly, where the nodes are individuals, a key research area has been the prediction of similarity in time-to-adoption of an innovation for pairs of actors (*31*). Performance refers to a node's outcomes with respect to some good. For example, researchers have found that firm centrality predicts the firm's ability to innovate, as measured by number of patents secured (*32*), as well as to perform well financially (*33*). Other research has linked individual centrality with power and influence (*34*).

Theoretical mechanisms. Perhaps the most common mechanism for explaining consequences of social network variables is some form of direct transmission from node to node. Whether this is a physical transfer, as in the case of material resources such as money (*35*), or a mimetic (imitative) process, such as the contagion of ideas, the underlying idea is that something flows along a network path from one node to the other.

The adaptation mechanism states that nodes become homogeneous as a result of experiencing and adapting to similar social environments. Much like explanations of convergent forms in biology, if two nodes have ties to the same (or equivalent) others, they face the same environmental forces and are likely to adapt by becoming increasingly similar. For example, two highly central nodes in an advice network could develop similar distaste for the telephone and e-mail, because both receive so many requests for help through these media. Unlike the case of transmission, the state of "distaste for communication media" is not transmitted from one node to another, but rather is similarly created in each node because of their similar relations to others.

The binding mechanism is similar to the old concept of covalent bonding in chemistry. The idea is that social ties can bind nodes together in such a way as to construct a new entity whose properties can be different from those of its constituent elements. Binding is one of the mechanisms behind the popular notion of the performance benefits of "structural holes" (Fig. 4). Given an ego-network (the set of nodes with direct ties to a focal node, called "ego," together with the set of ties among members of the ego network), a structural hole is the absence of a tie among a pair of nodes in the ego network (*22*). A well-established proposition in social network analysis is that egos with lots of structural holes are better performers in certain competitive settings (*19*). The lack of structural holes around a node means that the node's contacts are "bound" together—they can communicate and coordinate so as to act as one, creating a formidable "other" to negotiate with. This is the basic principle behind the benefits of worker's unions and political alliances. In

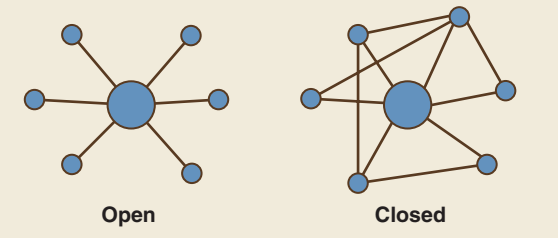


Fig. 4. Two illustrative ego networks. The one on the left contains many structural holes; the one on the right contains few.

contrast, a node with many structural holes can play unconnected nodes against each other, dividing and conquering.

The exclusion mechanism refers to competitive situations in which one node, by forming a relation with another, excludes a third node. To illustrate, consider a "chain" network (Fig. 5) in which nodes are allowed to make pairwise "deals" with those they are directly connected to. Node d can make a deal with either node c or node e, but not both nodes. Thus, node d can exclude node c by making a deal with node e. A set of experiments (*36*) showed that nodes b and d have high bargaining power, whereas nodes a, c, and e have low power. Of special interest is the situation of node c, which is more central than, and has as many trading partners as, nodes b and d. However, nodes b and d are stronger because each have partners (nodes a and e) that are in weak positions (no alternative bargaining partners).

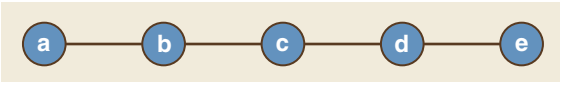


Fig. 5. A five-person exchange network. Nodes represent persons; lines represent exchange relations.

Having only strong nodes to bargain with makes node c weak. In this way, a node's power becomes a function of the powers of all other nodes in the network, and results in a situation in which a node's power can be affected by changes in the network far away from the node. An example of the exclusion mechanism occurs in business-to-business supply chains. When a firm intentionally locks up a supplier to an exclusive contract, competitor firms are excluded from accessing that supplier, leaving them vulnerable in the marketplace.

In quantum physics, the Heisenberg uncertainty principle describes the effects of an observer on the system being measured. A foreseeable challenge for network research in the social sciences is that its theories can diffuse through a population, influencing the way people see themselves and how they act, a phenomenon that Giddens described as the double-hermeneutic (*37*). For example, there has been an explosion in the popularity of social networking sites, such as Facebook and Linkedin, which make one's connections highly visible and salient. Many of these sites offer users detailed information about the structure and content of their social networks, as well as suggestions for how to enhance their social networks. Will this enhanced awareness of social network theories alter the way in which people create, maintain, and leverage their social networks?

Final Observations

A curious thing about relations among physical and social scientists who study networks is that each camp tends to see the other as merely descriptive. To a physical scientist, network research in the social sciences is descriptive because measures of network properties are often taken at face value and not compared to expected values generated by a theoretical model such as Erdos-Renyi random graphs. For their part, social scientists have reacted to this practice with considerable bemusement. To them, baseline models like simple random graphs seem naïve in the extreme—like comparing the structure of a skyscraper to a random distribution of the same quantities of materials.

More importantly, however, social and physical scientists tend to have different goals. In the physical sciences, it has not been unusual for a research paper to have as its goal to demonstrate that a series of networks have a certain property (and that this property would be rare in random networks). For social scientists, the default expectation has been that different networks (and nodes within them) will have varying network properties and that these variations account for differences in outcomes for the networks (or nodes). Indeed, it is the relating of network differences to outcomes that they see as constituting theoretical versus descriptive work.

Social scientists have also been more concerned than the physical scientists with the individual node, whether an individual or a collective such as a company, than with the network as a whole. This focus on node-level outcomes is probably driven to at least some extent by the fact that traditional social science theories have focused largely on the individual. To compete against more established social science theories, network researchers have had to show that network theory can better explain the same kinds of outcomes that have been the traditional focus of the social sciences.

Some physicists argue that direct observation of who interacts with whom would be preferable to asking respondents about their social contacts, on the grounds that survey data are prone to error. Social scientists agree that survey data contain error, but do not regard an error-free measurement of who interacts with whom to be a substitute for, say, who trusts whom, as these are qualitatively different ties that can have different outcomes. In addition, social scientists would note that even when objective measures are available, it is often more useful for predicting behavior to measure a person's perception of their world than to measure their actual world. Furthermore, the varying ability of social actors to correctly perceive the network around them is an interesting variable in itself, with strong consequences for such outcomes as workplace performance (*38*).

It is apparent that the physical and social sciences are most comfortable at different points along the (related) continua of universalism to particularism, and simplicity to complexity. From a social scientist's point of view, network research in the physical sciences can seem alarmingly simplistic and coarse-grained. And, no doubt, from a physical scientist's point of view, network research in the social sciences must appear oddly mired in the minute and the particular, using tiny data sets and treating every context as different. This is one of many areas where we can each take lessons from the other.

References and Notes

1. M. Newman, A. Barabasi, D. J. Watts, Eds., *The Structure and Dynamics of Networks* (Princeton Univ. Press, Princeton, NJ, 2006).
2. For a thorough history of the field, see the definitive work by Freeman (*39*).
3. J. L. Moreno, *Who Shall Survive?* Nervous and Mental Disease Publishing Company, Washington, DC, 1934).
4. E. Durkheim, *Suicide: A Study in Sociology* (Free Press, New York, 1951).
5. R. D. Luce, A. Perry, *Psychometrika* **14**, 95 (1949).
6. H. Leavitt, *J. Abnorm. Soc. Psychol.* **46**, 38 (1951).
7. Later experiments suggested that this result was contingent on other factors. For example, several experiments showed that, as the complexity of puzzles increased, decentralized networks performed better (*40*).
8. I. de S. Pool, M. Kochen, *Soc. Networks* **1**, 1 (1978).
9. S. Milgram, *Psychol. Today* **1**, 60 (1967).
10. C. S. Fischer, *To Dwell Among Friends* (Univ. of Chicago Press, Chicago, IL, 1948)
11. C. E. Hollingshead, *Elmtown's Youth* (Wiley, London, 1949).
12. B. Wellman *et al.*, *Annu. Rev. Sociol.* **22**, 213 (1996).
13. R. Brown, *Structure and Function in Primitive Society* (Free Press, Glencoe, IL, 1952).
14. S. F. Nadel, *The Theory of Social Structure* (Free Press, Glencoe, IL, 1957).
15. H. White, *An Anatomy of Kinship: Mathematical Models for Structures of Cumulated Roles* (Prentice Hall, Engelwood, NJ, 1963).
16. J. P. Boyd, *J. Math. Psychol.* **6**, 139 (1969).
17. E. Bott, *Family and Social Network* (Tavistock, London, 1957).
18. F. P. Lorrain, H. C. White, *J. Math. Sociol.* **1**, 49 (1971).
19. R. S. Burt, *Corporate Profits and Cooptation* (Academic Press, NY, 1983).
20. R. S. Burt, *Am. J. Sociol.* **92**, 1287 (1987).
21. M. S. Granovetter, *Am. J. Sociol.* **78**, 1360 (1973).
22. R. S. Burt, *Structural Holes: The Social Structure of Competition* (Harvard Univ. Press, Cambridge, MA. 1992).
23. R. Cross, A. Parker, *The Hidden Power of Social Networks* (Harvard Business School Press, Boston, MA, 2004).
24. J. A. Levy, B. A. Pescosolido, *Social Networks and Health* (Elsevier, London, 2002).
25. M. Sageman, *Understanding Terror Networks* (Univ. of Pennsylvania Press, Philadelphia, 2004).
26. S. P. Borgatti, in *Dynamic Social Network Modeling and Analysis: Workshop Summary and Papers*, R. Breiger, K. Carley, P. Pattison, Eds. (National Academy of Sciences Press, Washington, DC, 2003), p. 241.
27. L. C. Freeman, *Sociometry* **40**, 35 (1977).
28. J. F. Padgett, C. K. Ansell, *Am. J. Sociol.* **98**, 1259 (1993).
29. R. S. Burt, *Brokerage and Closure* (Oxford Univ. Press, New York, 2005).
30. G. F. Davis, H. R. Greve, *Am. J. Sociol.* **103**, 1 (1997).
31. T. W. Valente, *Soc. Networks* **18**, 69 (1996).
32. W. Powell, K. Koput, L. Smith-Doerr, *Adm. Sci. Q.* **41**, 116 (1996).
33. A. V. Shipilov, S. X. Li, *Adm. Sci. Q.* **53**, 73 (2008).
34. D. J. Brass, *Adm. Sci. Q.* **29**, 518 (1984).
35. N. Lin, *Annu. Rev. Sociol.* **25**, 467 (1999).
36. T. Yamagishi, M. R. Gilmore, K. S. Cook, *Am. J. Sociol.* **93**, 833 (1988).
37. A. Giddens, *The Constitution of Society* (Univ. of California Press, Berkeley and Los Angeles, 1984).
38. D. Krackhardt, M. Kilduff, *J. Pers. Soc. Psychol.* **76**, 770 (1999).
39. L. C. Freeman, *The Development of Social Network Analysis: A Study in the Sociology of Science* (Empirical Press, Vancouver, 2004).
40. M. E. Shaw, in *Advances in Experimental Social Psychology*, L. Berkowitz, Ed. (Academic Press, New York, 1964), vol. 1, p. 111.

41. We thank A. Caster, R. Chase, L. Freeman, and B. Wellman for their help in improving this manuscript. This work was funded in part by grant HDTRA1-08-1-0002-P00002 from the Defense Threat Reduction Agency and by the Gatton College of Business and Economics at the University of Kentucky.

10.1126/science.1165821

Tracking Long-Distance Songbird Migration by Using Geolocators

Bridget J. M. Stutchbury,[1]* Scott A. Tarof,[1] Tyler Done,[1] Elizabeth Gow,[1] Patrick M. Kramer,[1] John Tautin,[2] James W. Fox,[3] Vsevolod Afanasyev[3]

Until now it has been impossible to track migratory songbirds to their tropical wintering grounds. Songbirds are far too small for satellite tracking, and our current understanding of individual movements comes from brief snapshots of the journey via radar images, opportunistic recaptures of banded individuals, studies of migrants on the ground refueling, and an exceptional study that followed radio-tagged songbirds by airplane (*1*–*4*). We tracked songbird migration by mounting light-level geolocators (*5*, *6*) on 14 wood thrushes (*Hylocichla mustelina*) and 20 purple martins (*Progne subis*) breeding in northern Pennsylvania during 2007. The next summer we retrieved geolocators from five wood thrushes and two purple martins and analyzed sunrise and sunset times to reconstruct migration routes and estimate wintering locations (±300 km).

Rapid long-distance movement occurred in both species, and prolonged stopovers were common during fall migration (Fig. 1). Both purple martins flew south 2500 km to the Yucatan Peninsula in 5 days (500 km/day) and, on the basis of longitude estimates, then had a stopover of 3 to 4 weeks in the region (fig. S1). Four wood thrushes spent 1 to 2 weeks in the southeastern United States in late October before crossing the Gulf of Mexico (Fig. 1C), and two individuals stopped on the Yucatan Peninsula for 2 to 4 weeks before continuing migration (Fig. 1D).

Wood thrushes overwintered in a narrow band from 83.7° to 85.0°W (±1.4°) in Honduras or Nicaragua (Fig. 1), suggesting a level of connectivity not previously documented for migratory songbirds. Stable isotope analysis of black-throated blue warbler (*Dendroica carulescens*) feathers, for instance, showed that individuals wintering on western Caribbean islands originate from the northern portion of the breeding range, whereas those on easterly islands are from southern breeding areas (*7*).

Overall migration rate was 2 to 6 times more rapid in spring than in fall (table S1). One female martin (Fig. 1A) left the Amazon basin after the night of 12 April and flew about 7500 km in 13 days (577 km/day). Nine days involved migration flights, and 4 days were spent on stopover. Most wood thrushes returned to their breeding territory in only 13 to 15 days (233 to 271 km/day). One wood thrush did not cross the Gulf of Mexico on spring migration and took 29 days to complete the 4600-km overland route (Fig. 1D). Previous studies appear to greatly underestimate the true flight performance of migrating songbirds (*4*) because spring migration speed has typically been estimated at under 150 km/day.

Alarming long-term declines of migratory songbird species in North America and elsewhere heighten the urgency of mapping migration routes and wintering locations with far greater accuracy than is currently possible with stable isotope analysis (*8*). Tracking individuals to their wintering areas is essential for predicting the impact of tropical habitat loss and climate change (*7*, *9*). Survival estimates can now be obtained from regions where individuals from a specific breeding population overwinter, improving our understanding of how wintering versus breeding threats drive population fluctuations of migratory songbirds.

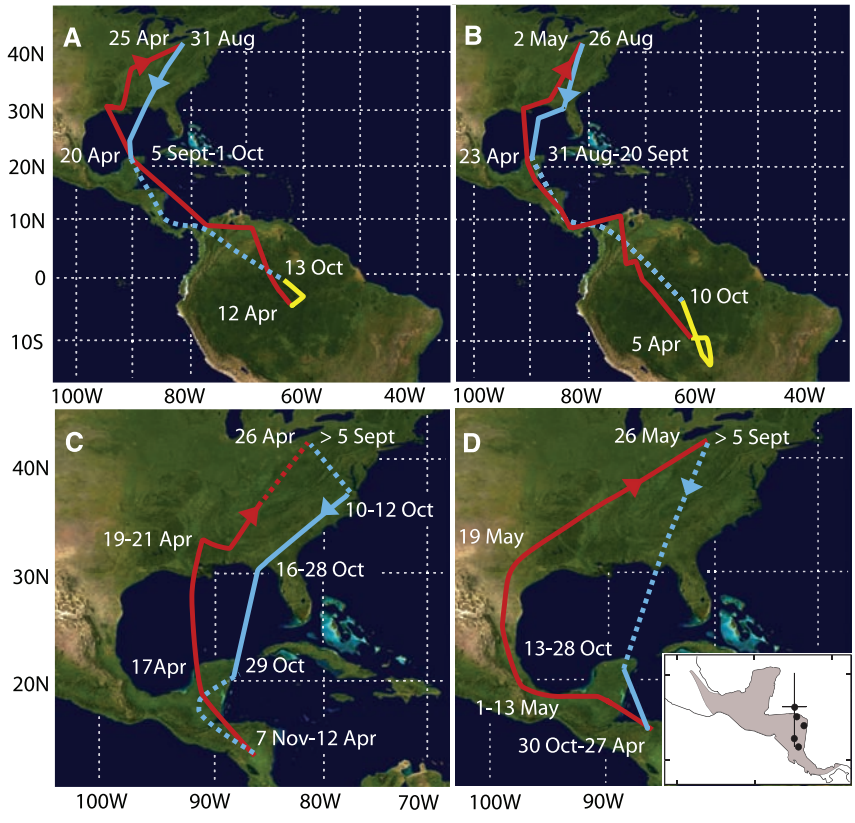


Fig. 1. Interpolated geolocation tracks of individual purple martins (**A** and **B**) and wood thrushes (**C** and **D**) that bred in northern Pennsylvania, USA (42°N, 80°W). Blue, fall migration; yellow, winter range movements; and red, spring migration. Dotted lines link locations when latitude could not be determined. Inset shows winter territory locations of wood thrush and species winter range (shaded); the standard deviation for one individual is shown.

References and Notes

1. S. Gauthreaux Jr., C. G. Belser, *Auk* **120**, 266 (2003).
2. K. C. Ruegg, T. B. Smith, *Proc. R. Soc. London Ser. B* **269**, 1375 (2002).
3. W. Yong, F. R. Moore, *Auk* **114**, 263 (1997).
4. M. Wikelski *et al.*, *Nature* **423**, 704 (2003).
5. J. P. Croxall, J. R. D. Silk, R. A. Phillips, V. Afanasyev, D. R. Briggs, *Science* **307**, 249 (2005).
6. S. A. Shaffer *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **103**, 12799 (2006).
7. D. R. Rubenstein *et al.*, *Science* **295**, 1062 (2002).
8. K. A. Hobson, *Auk* **122**, 1037 (2005).
9. T. S. Sillett, R. T. Holmes, T. W. Sherry, *Science* **288**, 2040 (2000).
10. B. J. M. Stutchbury, *Silence of the Songbirds* (Walker, New York, 2007).
11. We thank C. Graziano, R. Kresnik, T. MacIntosh, M. MacPherson, E. Pifer, T. Piraino, S. Puppi, C. Silverio, and C. Stanley for field assistance. For funding, we thank the Natural Sciences and Engineering Research Council of Canada, the National Geographic Society, the Purple Martin Conservation Association, and proceeds from (*10*).

Supporting Online Material

www.sciencemag.org/cgi/content/full/323/5916/896/DC1
Materials and Methods
Fig. S1
Table S1
References

1 October 2008; accepted 4 December 2008
10.1126/science.1166664

[1]Department of Biology, York University, Toronto, Ontario M3J 1P3, Canada. [2]Purple Martin Conservation Association, Tom Ridge Environmental Center, 301 Peninsula Drive, Erie, PA 16505, USA. [3]British Antarctic Survey, Natural Environment Research Council, High Cross, Madingley Road, Cambridge CB3 0ET, UK.

*To whom correspondence should be addressed. E-mail: bstutch@yorku.ca

Lunar Global Shape and Polar Topography Derived from Kaguya-LALT Laser Altimetry

H. Araki,[1]* S. Tazawa,[2] H. Noda,[1] Y. Ishihara,[2] S. Goossens,[2] S. Sasaki,[2] N. Kawano,[2] I. Kamiya,[3] H. Otake,[4] J. Oberst,[5] C. Shum[6]

A global lunar topographic map with a spatial resolution of finer than 0.5 degree has been derived using data from the laser altimeter (LALT) on board the Japanese lunar explorer Selenological and Engineering Explorer (SELENE or Kaguya). In comparison with the previous Unified Lunar Control Network (ULCN 2005) model, the new map reveals unbiased lunar topography for scales finer than a few hundred kilometers. Spherical harmonic analysis of global topographic data for the Moon, Earth, Mars, and Venus suggests that isostatic compensation is the prevailing lithospheric support mechanism at large scales. However, simple rigid support is suggested to dominate for the Moon, Venus, and Mars for smaller scales, which may indicate a drier lithosphere than on Earth, especially for the Moon and Venus.

Knowledge of the lunar shape and topography is fundamental for our understanding of the internal structure and surface evolution of the Moon. However, laser topographic mapping of the Moon had been limited. Early experiments were carried out by the Apollo 15 to 17 spacecraft for limited areas below their near-equatorial orbits (*1*). More recently, laser altimeter on the Clementine obtained topographic data covering almost the entire South Pole–Aitken Basin (*2*). Its coverage did not reach beyond 80° latitude owing to the elliptic orbit of the spacecraft, and the spatial resolution is 20 to 60 km (*3*). A full topographic map, including the polar regions, ULCN 2005, was produced using photogrammetric analysis of Clementine images combined with data from historical control points on the Moon (*4*). However, the map is known to suffer from large systematic errors. Radar topography from Earth includes gaps due to limitations of viewing and illumination conditions, whose effects are especially evident in the polar regions (*5*, *6*).

The LALT on board the Japanese lunar explorer Kaguya (SELENE) is designed to measure distances to the lunar surface from the altitude of 100 km (*7*). LALT measures the distance from the spacecraft to the surface of the Moon by transmitting Nd and Cr-doped yttrium-aluminum-garnet (Cr-doped Nd:YAG) laser pulses every second. The beam divergence is 0.4 milliradian, resulting in the laser spot size on the lunar surface of typically 40 m from the orbiter altitude of 100 km. To exploit the instrument's nominal range resolution of 1 m, the range data are calibrated for thermal variations of the internal clock frequency and the instrument delay. The errors related to data quantization, thermal variation of the clock and electronics, and instrument delay measurement are estimated to be 0.55 m (1 SD) (*8*, *9*). The first ranging tests were carried out successfully on 25 November 2007, and the nominal mapping phase began on 30 December 2007. Topography data were produced by incorporating precise orbits for the Kaguya main orbiter. These orbits are calculated from two-way Doppler data by the GEODYN-II software using the latest lunar gravity model SGM90em (SELENE Gravity Model) that is an adapted version of the model SGM90d for the purpose of orbit determination (*10*–*12*). Orbit precision is determined from orbit differences during overlapping parts, showing that the radial orbit error is generally within 1 m (*13*) and the total positioning error (computed using the root sum square over the radial, along-track, and cross-track directions) is found to be ~50 m. Thus, the radial topographic error originated from the orbit repeatability is 1 m (1 SD), the instrumental error is 0.55 m (1 SD), and the instrument range shift is between +2.5 m and +12 m (*8*, *9*), which are summarized ±4.1 m (1 SD) as the final budget where the range shift is incorporated as 4 m (1 SD). In the same way, the horizontal topographic error originated from the orbit repeatability is 50 m (1 SD), the pointing error is 175 m (maximum), and the time-tag error is 1.5 m (maximum), which are summarized as ±77 m (1 SD) as the final budget (*14*). Attitude and time-tag data are provided from the tracking and operation center of the Japan Aerospace Exploration Agency (JAXA). The number of geolocated points over the entire lunar surface is about 6.77×10^6 as of 31 March 2008 (*15*).

Figure 1 shows the topographic map obtained from LALT data. It clearly delineates the prom-

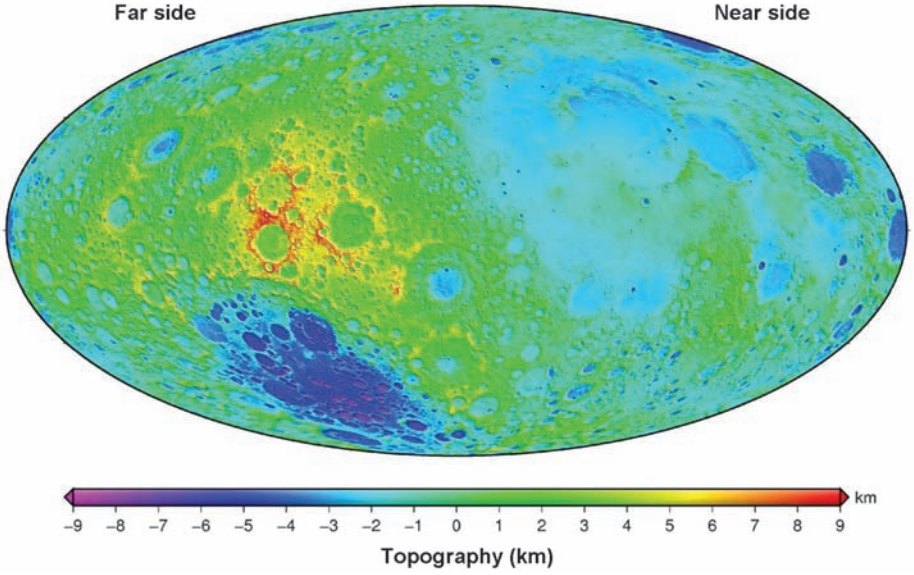


Fig. 1. Lunar global topographic map obtained from LALT altimetry data shown in Hammer equal-area projection. Lunar coordinates are based on the mean Earth/polar axis system. Reference of the height is a sphere whose radius is 1737.4 km and whose origin is set to the center of mass (*19*). The map center is 270°E, with the nearside on the right and the farside on the left. Full range of the topography is about 19.81 km. The highest point is on the southern rim of the Dirichlet-Jackson basin (–158.64°E, 5.44°N, +10.75 km), and the lowest point is inside Antoniadi crater (–172.58°E, 70.43°S, –9.06 km) in the South Pole–Aitken Basin.

[1]National Astronomical Observatory of Japan, 2-21-1 Osawa, Mitaka, Tokyo 181-8588, Japan. [2]National Astronomical Observatory of Japan, 2-12 Hoshigaoka, Mizusawa, Oshu, Iwate 023-0861, Japan. [3]Geographical Survey Institute, 1 Kitasato, Tsukuba, Ibaraki 305-0811, Japan. [4]Japan Aerospace Exploration Agency, 2-1-1 Sengen, Tsukuba, Ibaraki 305-8505, Japan. [5]German Aerospace Center (DLR), Rutherfordstraße 2, 12489 Berlin, Germany. [6]Ohio State University, School of Earth Sciences, 121 South Oval Mall, 275 Mendenhall Laboratory, Columbus, OH 43210, USA.

*To whom correspondence should be addressed. E-mail: arakih@miz.nao.ac.jp

inent lunar geologic landforms, especially the Feldspathic Highland Terrane (FHT), mare basins, and the South Pole–Aitken Basin Terrane (SPAT). Mare surfaces, which are covered with basaltic lava, are remarkably flat. Ridge systems have elevations of a few hundred meters. In Mare Serenitatis and Imbrium, for example, ridges form concentric circular patterns parallel to the outermost mare cliff. Complex sinusoidal ridges are also seen in Oceanus Procellarum and Mare Imbrium. These ridge systems may be formed by subsidence of the mare basalt modified by preexisting multiring or smaller topographic features (fig. S1).

Because Kaguya is in a polar orbit, surface trajectories appear approximately parallel at low latitudes, with cross-track spacing less than 0.5° (about 15 km at the equator) (fig. S2). In the polar regions, the nearest neighbor distance between ranged positions is no more than 2 km inside the region of 1° from the poles (fig. S2). A comparison between LALT and ULCN 2005 maps shows that the LALT topographic map far exceeds the ULCN 2005 in resolution and accuracy for features that are less than a few hundred kilometers across (fig. S3).

On the farside of the Moon, the mean difference of topography is 10 km between FHT and the SPAT (Fig. 1, and fig. S4). The highest point on the Moon is on the southern rim of the Dirichlet-Jackson basin, and the lowest one is in the Antoniadi crater in the SPAT. The full-range topography spans about 19.81 km, which is greater than the ULCN 2005 result of 17.53 km for the next highest and lowest points, the positions of which are generally identical to our highest and lowest points, with differences of less than a few degrees (Fig. 1) (*16*).

Observations by LALT provide the first polar maps with complete coverage (Fig. 2). Small depressions or craters are observed for the first time in the polar image. For example, our map of the South Pole region disclosed a crater, with a diameter of ~15 km, inside of de Gerlache crater and a topographic depression on the farside of Shackleton and de Gerlache craters. The new polar map will be a valuable asset for planning future lunar exploration, especially for the sun-lighting conditions.

We have developed a spherical harmonic model complete to degree and order 359 [STM359_grid-02 (SELENE Topography Model)] using the LALT topographic data as of 31 March 2008. Topographic data were interpolated and assembled to 0.0625° by 0.0625° gridded data, boxcar filtered and resampled to a 0.25° by 0.25° grid (*17*), and then converted to spherical harmonic coefficients (*18*). The N degree and order spherical harmonic model of the global topography H is given as follows:

$$H(\lambda,\varphi) = \sum_{l=0}^{N} \sum_{m=0}^{l} \overline{P}_{lm}(\sin\varphi)(\overline{C}_{lm} \cos m\lambda + \overline{S}_{lm} \sin m\lambda) \quad (1)$$

where φ and λ are the lunar-centric latitude and longitude, respectively, based on the mean Earth/polar axis system (*19*). $\overline{P}_{lm}$ is the normalized associated Legendre function of degree l and order m. $\overline{C}_{lm}$ and $\overline{S}_{lm}$ are normalized spherical harmonic coefficients with units in meters. Low degree and order coefficients (0, 1, and 2) are listed in Table 1. Considering our known ranging errors, C_{00}, the mean radius of the Moon is estimated to be 1737.15 ± 0.01 km. It is considered that the error of C_{00} resulting from the orbit and attitude errors is averaged to be small in total. The mean radius derived from ULCN 2005 is 1736.93 ± 2.1 km, which is in good agreement with our value (*16*). The COM-COF (center of mass–center of figure) offset is derived from the C_{1m} (m = 0, 1) as (–1.772, –0.731, 0.239) km along three axes in the mean Earth/polar axis coordinates or 1.93 km in total. The direction of the COM-COF vector is displaced 22.41°E from the prime meridian and 7.10°N from the equator on the farside. This is slightly north of the highest elevation point of the Moon, which suggests that irregular mass distribution associated with shifted COM-COF offset around the axis of lunar minimum moments of inertia is virtually compensated due to the variations in crustal thickness. Our results on the COM-COF offset generally agree with those obtained from Clementine LIDAR (light detection and ranging) altimetry (*3*). The polar and mean equatorial radii are derived to be 1735.66 km and 1738.64 km, respectively, from the lunar mean radius (1737.15 km) and C_{20} coefficient. The polar flattening is thus estimated to be 1/581.9 from the two radii.

We compared STM359_grid-02 with the ULCN 2005 spherical harmonic model developed using the same procedure as for STM359_grid-02. The amplitude spectrum, which is derived as the square root of the total sum of squares of the spherical harmonic coefficients for each degree (Fig. 3), are nearly identical for degree and order less than 30 (half wavelength scale is more than 180 km on the Moon). However, the STM359_grid-02 spectral amplitudes are larger by a factor of two or more for degree and order 100 or at smaller scales

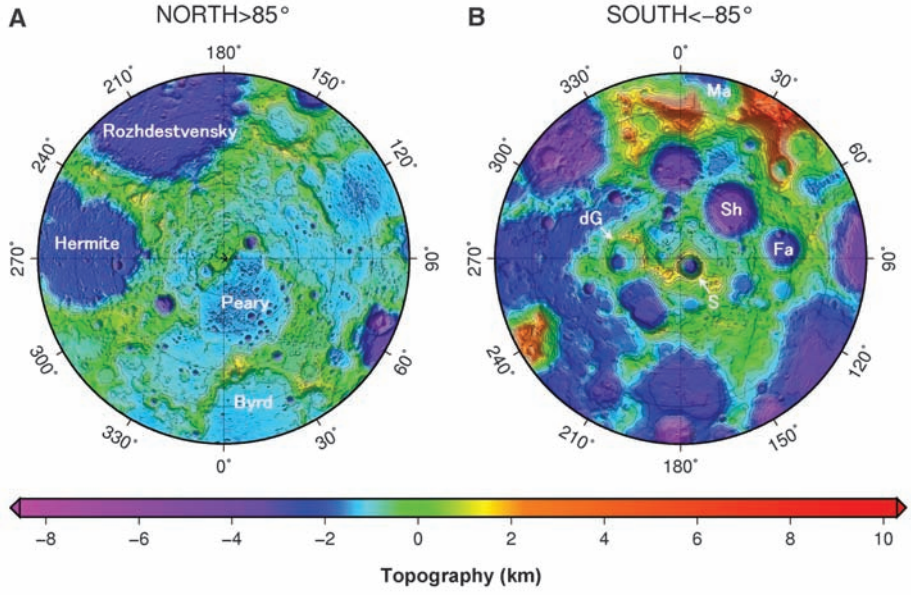


Fig. 2. Polar topographic maps of the Moon produced by LALT. The projections are stereographic for the North Pole (**A**) and the South Pole (**B**). Lunar coordinate system and the height reference are the same as Fig. 1. Some craters are labeled and some are abbreviated: Ma, Malapert; Sh, Shoemaker; Fa, Faustini; S, Shackleton; and dG, de Gerlache. The grid resolution is 0.015625° (1/64°) in latitude and 0.179° in longitude (0.473 km at 85°).

Table 1. Normalized coefficients of degree and order 0, 1, and 2, obtained from the lunar shape model (STM359_ grid-02) from the spherical harmonic analysis. Normalized associated Legendre function $\overline{P}_{lm}$ is $\overline{P}_{lm}(\sin\varphi) = \sqrt{(2l+1)(2-\delta_{l0})(l-m)!/(l+m)!}P_{lm}(\sin\varphi)$, where P_{lm} is an associated Legendre function. $\overline{P}_{lm}$ is used for our spherical harmonic analysis.

Degree: l	Order: m	$\overline{C}$ (m)	$\overline{S}$ (m)
0	0	1,737,156.3	—
1	0	137.9	—
1	1	–1023.2	–421.9
2	0	–668.1	—
2	1	–769.5	–16.9
2	2	109.2	383.4

(less than 55 km), indicating that the Moon is rougher at scales of less than 180 km than is indicated by ULCN 2005 model.

Topographic amplitude spectra T_p of the Moon and three terrestrial planets are compared in Fig. 4, where p is for the Moon (M), Mars (Ma), Venus (V), and Earth (E). T_M is calculated from STM359_grid-02, and T_{Ma}, T_V, and T_E are from results of spherical harmonic analysis of the three bodies' shape data (*16*). To plot and compare T_p on the same horizontal scale, a half-wavelength scale for the *N*-degree spherical harmonic function is introduced as $\lambda/2 = (2\pi R_p/2N)$ in place of *N*, where R_p is the body's radius. The relationship $T_p \propto (\lambda/2)$ for $\lambda/2 \geq 600$ km is generally confirmed for each body.

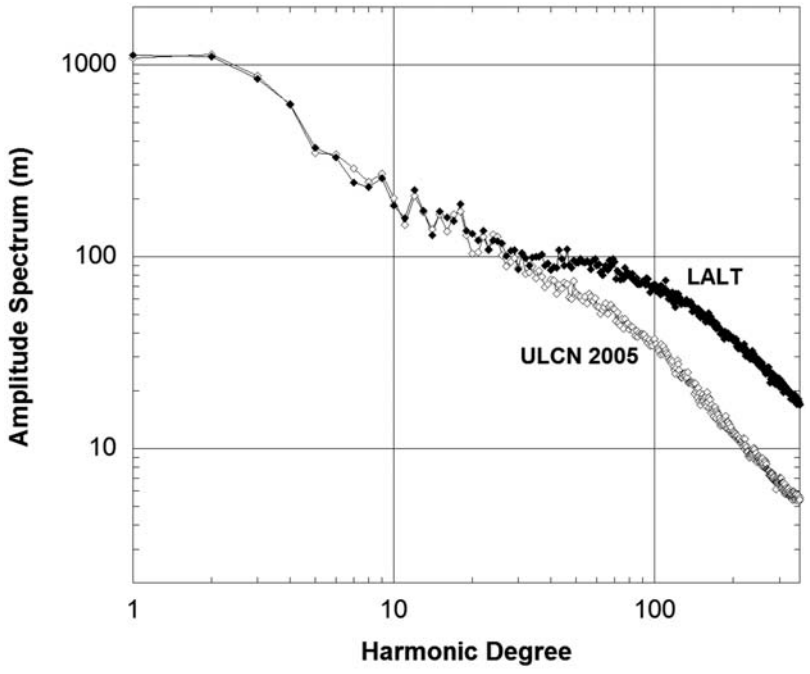


Fig. 3. Lunar topographic amplitude spectra from STM359_grid-02 (filled diamonds) and the spherical harmonic model from ULCN 2005 (open diamonds) (*4*). The two plots have similar characteristics for low-order coefficients. However, for scales smaller than 180 km, the Moon is clearly rougher than was previously known.

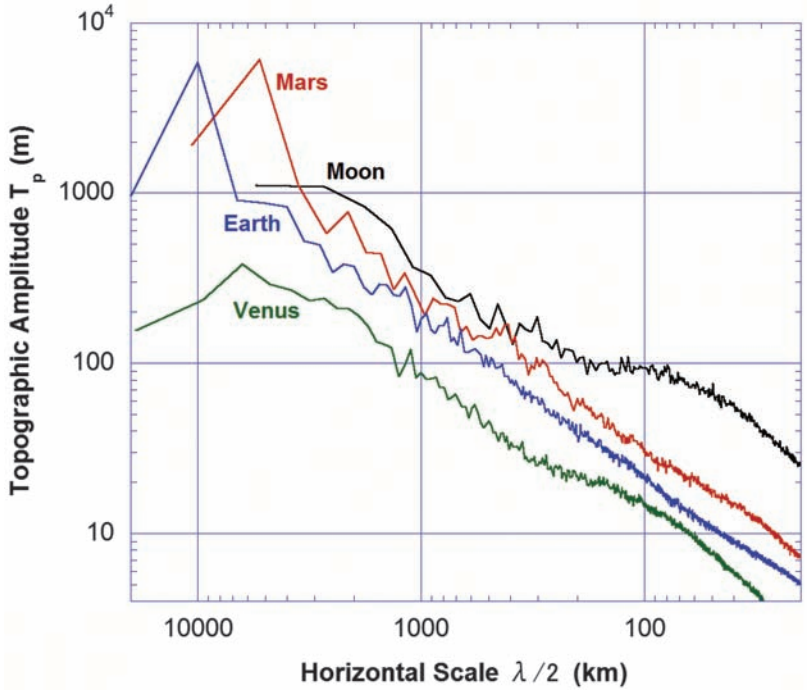


Fig. 4. Topographic amplitude spectra T_p for the Moon and three terrestrial planets. T_p is compared on the same horizontal scale $\lambda/2 = (2\pi R_p/2N)$, where R_p is the radius of each body.

The compensation state of Earth's interior is well explained by an Airy isostatic model. For the other three bodies, some degree of isostatic compensation is known to be achieved for the long-wavelength topography by comparing spectra of the observed and calculated gravity field, assuming that their short-wavelength topography is not compensated (*20*). Lunar free-air anomaly is moderate on the FHT or SPAT, where $\lambda/2$ is from 2000 to 4000 km, which indicates the isostatic equilibrium by the upwelling of the lower crust and/or upper mantle (*2*). However, T_M for $\lambda/2 \leq$ 400 km exhibits a somewhat larger value than the prediction from the simple linear relationship [$T_M \propto (\lambda/2)$] and shows its virtually flat spectrum for 90 km $\leq \lambda/2 \leq$ 180 km. These two characteristics contribute the excess of topographic spectrum for $\lambda/2 \leq$ 400 km. A similar profile is observed for Venus (T_V), where the excess and flat spectrum is found for $\lambda/2 \leq$ 300 km and 150 km $\leq \lambda/2 \leq$ 180 km, respectively. For Mars (T_{Ma}), the excess spectrum is observed clearly for 400 km $\leq \lambda/2 \leq$ 600 km and $\lambda/2 \cong$ 300 km. The excess and/or flat spectrum common to the Moon, Venus, and Mars may indicate that the compensation mechanism by the buoyancy and elasticity of their lithosphere is changed to the simple rigid support. In the same way, the lack of such characteristics for the topographic spectra of Earth may suggest that the compensation mechanism is valid for Earth at least down to the topography whose horizontal scale is several tens of kilometers.

Why do the lithospheres of the Moon, Venus, and Mars seem to be more rigid for small-scale topography than Earth? The reason may be that volatile materials such as water, which weaken the mechanical property of the lithosphere, are relatively poor in the crust and/or upper mantle of these three bodies compared with Earth. It is confirmed that lunar surface rock does not have any signs of water or other volatile alteration, and the lunar lithosphere is considered to be extremely dry, which is probably the case for Venus, too, owing to its high-temperature surface environment (*21*). The interior of Mars may contain more water and other volatile materials than the Moon and Venus if we consider the hypothesis that a vast amount of liquid water had existed on its surface. However, based on the fact that no plate tectonics is observed on Mars, the water inside Mars would have very limited effects to weaken the rigidity of the lithosphere. In contrast, the water inside Earth plays an important role in the formation of the continents and in plate tectonics. Thus, the amount and activity of volatile materials within the terrestrial planets is considered to be a key parameter for understanding their global isostatic state as well as the mechanical properties of the lithosphere itself.

References and Notes

1. G. A. Neumann, *Int. Arch. Photogramm. Remote Sens.* **XXXIV-3** (W4), 73 (2001).
2. M. T. Zuber *et al.*, *Science* **266**, 1839 (1994).
3. D. E. Smith *et al.*, *J. Geophys. Res.* **102**, 1591 (1997).
4. B. A. Archinal *et al.*, "The Unified Lunar Control Network 2005" (Open-File Report 2006-1367 ver. 1.0, U.S. Geological Survey, 2006).
5. B. A. Campbell, D. B. Campbell, *Icarus* **180**, 1 (2006).
6. B. Bussey, P. Spudis, *The Clementine Atlas of the Moon* (Cambridge Univ. Press, Cambridge, 2004).
7. T. Tsubokawa *et al.*, *Proc. of the 23rd International Symposium on Space Technology and Science* (*ISTS*), Matsue Shimane, Japan, 26 May to 2 June 2002 (2002), pp. 1986–1991.
8. H. Araki *et al.*, *Adv. Space Res.* **42**, 317 (2008).
9. One remaining concern is about systematic errors in measurements of pulse arrivals due to distortion of the return pulse caused by the sloped and/or rough target terrain in combination with unknown albedo effects. This error (range shift) may result in underestimates of ranges by 12 m in the very worst case of 30° slope and 30% surface reflectance before the pulse spreading correction. For moderately flat surfaces, systematic range errors are expected to be 2.5 m (*8*).
10. D. E. Pavlis *et al.*, "GEODYN II system description, vols. 1-5," Contractor Report, SGT Inc. (2006).
11. J. J. McCarthy, "SOLVE program user's guide," Contractor Report between NASA/GSFC and Hughes/STX (2007).
12. N. Namiki *et al.*, *Science* **323**, 900 (2009).
13. Orbits are determined by full-scale precision force and measurement modeling. For each data arc, estimated parameters include the state vector at epoch, a solar radiation pressure coefficient, empirical accelerations with a once per orbital revolution signature in the along-track and cross-track direction, and measurement biases to absorb systematic effects and mismodeling. Orbit precision has been evaluated by computing orbit overlap differences. Overlap analysis showed a radial consistency of 1 m in general, with outliers (that were excluded from topography data processing) up to 4 m.
14. The pointing error is considered to be <0.1° (175 m for 100-km altitude), based on the thermal and other deformation analysis of the main orbiter. Time-tag error is <1 msec (1.5 m along-track for 100-km altitude) through the correction that takes into account the propagation time from the main orbiter to each station and the processing delay on each tracking station. These values (175 m and 1.5 m) are incorporated into the final budget as 3 SD errors.
15. The laser power has decreased since the middle of April 2008 for yet unexplained reasons, disrupting our data analysis schedule.
16. B. A. Archinal *et al.*, *Lunar Planet. Sci.* **XXXVIII**, 1904 (2007).
17. P. Wessel, W. H. F. Smith, *EOS* **72**, 441 (1991).
18. M. A. Wieczorek, in *Treatise on Geophysics*, G. Schubert, Ed. in Chief (Elsevier, Amsterdam, 2007), vol. 10, pp. 165–206.
19. P. K. Seidelmann *et al.*, *Celestial Mech. Dyn. Astron.* **98**, 155 (2007).
20. A. B. Watts, *Isostacy and Flexure of the Lithosphere* (Cambridge Univ. Press, Cambridge, 2001).
21. S. J. Mackwell *et al.*, *J. Geophys. Res.* **103**, 975 (1998).

22. We would like to express our great appreciation for the contributions of all staff engineers of NEC Co. Ltd. and the entire staff of the Kaguya (SELENE) project at JAXA in their development and operational support of LALT. We would also express our special thanks to M. Wieczorek for SHTOOLS on his Web site (www.ipgp.jussieu.fr/~wieczor) used for the spherical harmonic analysis and F. Lemoine and his colleagues at NASA Goddard Space Flight Center for invaluable discussions on the orbit accuracy of the Kaguya main orbiter. Finally, we are grateful to the reviewers for their helpful review and comments and to editors for their careful checking of our manuscript. All of the science data obtained by the Kaguya mission will be made available to the public in a PDS-like format (not PDS itself) via the Internet and/or some other method on 1 November 2009. LALT data sets are summarized in table S1.

Supporting Online Material

www.sciencemag.org/cgi/content/full/323/5916/897/DC1
Figs. S1 to S4
Table S1

4 August 2008; accepted 23 December 2008
10.1126/science.1164146

Farside Gravity Field of the Moon from Four-Way Doppler Measurements of SELENE (Kaguya)

Noriyuki Namiki,[1]* Takahiro Iwata,[2] Koji Matsumoto,[3] Hideo Hanada,[3] Hirotomo Noda,[3] Sander Goossens,[3] Mina Ogawa,[4] Nobuyuki Kawano,[3] Kazuyoshi Asari,[3] Sei-itsu Tsuruta,[3] Yoshiaki Ishihara,[3] Qinghui Liu,[3] Fuyuhiko Kikuchi,[3] Toshiaki Ishikawa,[3] Sho Sasaki,[3] Chiaki Aoshima,[5] Kosuke Kurosawa,[6] Seiji Sugita,[6] Tadashi Takano[7]

The farside gravity field of the Moon is improved from the tracking data of the Selenological and Engineering Explorer (SELENE) via a relay subsatellite. The new gravity field model reveals that the farside has negative anomaly rings unlike positive anomalies on the nearside. Several basins have large central gravity highs, likely due to super-isostatic, dynamic uplift of the mantle. Other basins with highs are associated with mare fill, implying basalt eruption facilitated by developed faults. Basin topography and mantle uplift on the farside are supported by a rigid lithosphere, whereas basins on the nearside deformed substantially with eruption. Variable styles of compensation on the near- and farsides suggest that reheating and weakening of the lithosphere on the nearside was more extensive than previously considered.

In the beginning of the space age, Apollo missions and their precursors discovered that the topography and crustal thickness of the Moon differed on the near and farsides (*1*). Post-Apollo global mapping missions such as Clementine, Lunar Prospector (LP) and Small Missions for Advanced Research in Technology–1 (SMART-1) further revealed that the near and farsides also differ in their chemical composition (*2*–*4*), which might be a consequence of asymmetric crystallization of a primordial magma ocean (*5*). Here, we present gravity data of the farside from SELENE (Kaguya) (*6*), which helps resolve the origin of this dichotomy.

The gravity field is a fundamental physical quantity for the study of the internal structure and the evolution of planetary bodies. The Moon was the first target of planetary gravimetry. In 1966, the Luna 10 mission began the study of the gravity field from observed orbital motion of the spacecraft. It was followed by Lunar Orbiter missions (LO) from I to V and Apollo 15 and 16 subsatellites (A15/16ss), and more recently Clementine and LP. Muller and Sjogren (*7*) discovered large positive gravity anomalies called "mascons" within maria basins on the nearside. The elliptic orbit of the Clementine spacecraft improved the lower degrees and sectorial terms of the gravity field (*8*). The low circular, polar orbit of LP increased the spatial resolution of the nearside gravity field (*9*), and we will use the degree and order 100 model (LP100K) (*9*) for comparisons with our model.

The most important problem of the previous lunar gravity models is the lack of direct observa-

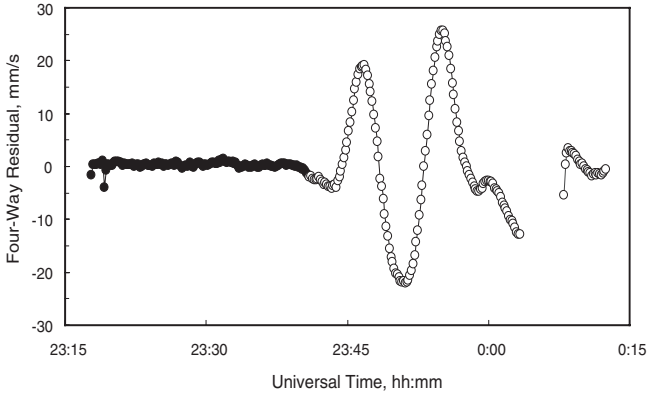


Fig. 1. Four-way Doppler residuals in pass from 23:17 UT on 5 November to 0:12 UT on 6 November 2007. Solid and open symbols indicate the visibility and invisibility of Main from UDSC, respectively. Four-way data are separated into two arcs by unloading of momentum wheels between 00:03 and 00:08.

[1]Kyushu University, 6-10-1 Hakozaki, Higashi-ku, Fukuoka 812-8581, Japan. [2]Institute of Space and Astronautical Science, Japan Aerospace Exploration Agency, 3-1-1 Yoshinodai, Sagamihara, Kanagawa 229-8510, Japan. [3]National Astronomical Observatory of Japan, 2-12 Hoshigaoka, Mizusawa, Oshu, Iwate 023-0861, Japan. [4]Japan Aerospace Exploration Agency, Tsukuba Space Center, 2-1-1 Sengen, Tsukuba, Ibaraki 305-8505, Japan. [5]Fujitsu Ltd., 1-9-3 Nakase, Mihama-ku, Chiba 261-8588, Japan. [6]University of Tokyo, 5-1-5 Kashiwanoha, Kasiwa, Chiba 277-8561, Japan. [7]Nihon University, 7-24-1 Narashinodai, Funabashi, Chiba 274-8501, Japan.

*To whom correspondence should be addressed. E-mail: nori@geo.kyushu-u.ac.jp

tions of the farside gravity signals (*9*). Synchronous rotation of the Moon with its orbit inhibits a direct link between an Earth-based ground tracking station and a lunar-orbiting spacecraft over the farside. Even though a 100-km altitude allows tracking of spacecraft within about 20° over the limb, about one-third of the lunar surface remains uncovered by direct tracking data. Before SELENE, the information regarding the farside gravity depended on the long-wavelength farside gravitational effect on the spacecraft orbit, particularly from tracking of the A15/16ss and Clementine, whose integrated effect somewhat reduced the uncertainties of the farside gravity (*10*). Nonetheless, the lack of farside tracking degrades the quality of spherical harmonic expansions (*9*) that require global distribution of observations. Consequently, the coefficients of the previous gravity models are strongly correlated with each other (*9*). To compensate for the lack of tracking data on the farside, an a priori constraint (*11*) has been used in processing tracking data to produce a global lunar gravity field model (*7*–*9*). This constraint successfully prevents unreasonably large oscillations in farside gravity, resulting from the lack of observations, while simultaneously maintaining the signal as observed on the nearside (*9*). However, the farside gravity field will remain unresolved until global coverage of gravimetric observations is achieved.

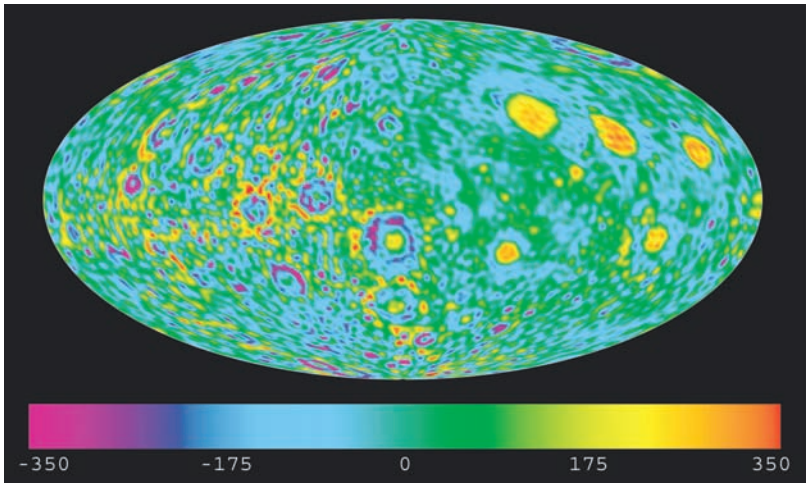


Fig. 2. Free-air gravity anomaly map from SGM90d. In this and Figs. 4 and 5, the lunar nearside is on the right side of the figure, and the farside is on the left. The color bar at the bottom indicates the gravity anomaly in milli-Galileo (1 mGal = 10^{-5} m s^{-2}). Positive gravity anomalies greater than 400 mGal or negative anomalies smaller than −400 mGal are truncated in Fig. 2. The maximum and minimum values of the map are 640 and −720 mGal. High-low-high patterns at high latitudes in the north for SGM90d results from the data gap of the four-way Doppler data between orbital tracks (fig. S2), and possibly from correlated coefficients.

To efficiently track a spacecraft over the lunar farside, we developed a satellite-to-satellite Doppler tracking subsystem (RSAT) for SELENE (*12*) (fig. S1). SELENE consists of three satellites: the main orbiter (Main), the relay subsatellite (Rstar), and the very-long-baseline interferometry (VLBI) subsatellite (Vstar) (*6*, *13*, *14*). When Main is orbiting over the farside of the Moon, a tracking signal in the S-band frequency, transmitted from Usuda Deep Space Center (UDSC) of the Japan Aerospace Exploration Agency (JAXA), is relayed by Rstar to Main keeping the phase coherence. Then Main returns the coherent tracking signal to Rstar, and Rstar converts the S-band (2.2 GHz) signal into X-band (8.5 GHz) to downlink a coherent Doppler signal to UDSC, thus establishing tracking data of Main over the farside (four-way Doppler measurement) (*12*) (fig. S1). At the same time, conventional range and range rate measurements are carried out between Rstar and UDSC (two-way Doppler and range measurements) (fig. S1).

Tracking data have been independently processed at both the National Astronomical Observatory of Japan (NAOJ) and the Flight Dynamics Division of JAXA, with different software systems to independently produce estimates of new lunar gravity field models. Our lunar gravity field model was developed at NAOJ with SELENE tracking data from 31 October 2007 to 1 April 2008, and with historical tracking data. The processed tracking data of SELENE include

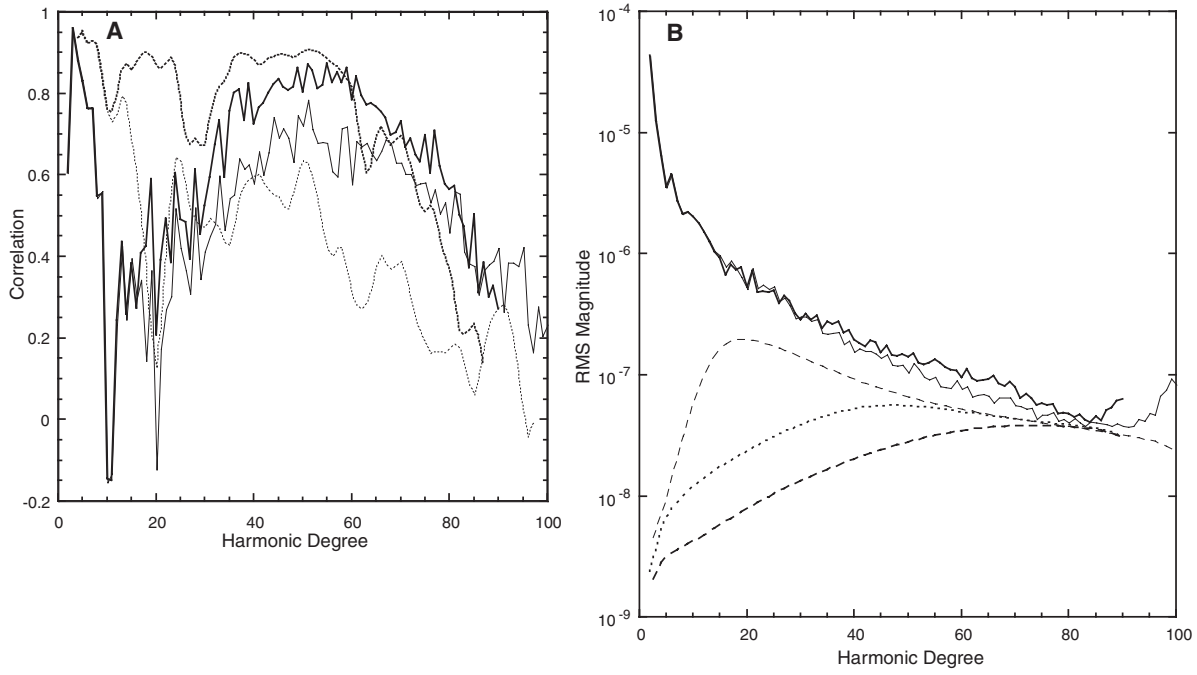


Fig. 3. (**A**) Correlation of gravity and topography models as a function of spherical harmonic degree summed over orders and (**B**) amplitude spectrum of the lunar gravity field as a function of spherical harmonic degree. (A) Thick solid line, SGM90d; thin solid line, LP100K (*9*); thick dashed line, SGM90d localized on the farside; thin dashed line, LP100K localized on the farside. (B) Thick solid line, from SGM90d; thin solid line, from LP100K (*9*); thick dashed line, the errors in the spectrum for SGM90d; thick dotted line, the errors in the spectrum for SGM90d recalculated for four-way Doppler data with a weight of 5 mm/s; and thin dashed line, for LP100K.

four-way Doppler data, two-way tracking data of Main using the 11-m Ground Network (GN) antennas of JAXA, and two-way range and Doppler data for both Rstar and Vstar using UDSC (*14*). All data are integrated to 10 s. Data weights in the gravity solution are summarized in table S1. The data weights are chosen to emphasize these new data and to balance the discrepancy between the amount of data for the near and farsides (table S1). We simultaneously solve for both orbits of Rstar and Main as well as the gravity field model (*14*).

The four-way tracking started at 22:00 UT on 31 October 2007 during the initial check-out phase. In Fig. 1, residuals of four-way data (the observed data minus data computed from precision modeling), using LP100K as the a priori gravity model, are shown for a measurement pass starting at 23:17 on 5 November and ending at 0:12. During this pass, Main was initially over the nearside while the four-way link was established, and thus directly visible from UDSC, before becoming invisible from UDSC when it disappeared over the farside. Solid and open symbols in the figure indicate visibility and invisibility of Main from UDSC, respectively. Figure 1 shows the long-wavelength, large-amplitude features of the residuals (up to 30 mm/s) when Main goes over the farside, whereas the residuals when Main is over the nearside are much smaller [with a root mean square (RMS) of 0.6 mm/s and an amplitude of about 1 mm/s].

Although the farside coverage is not completed in these data (fig. S2), it is sufficient to resolve a preliminary global gravity field model with an a priori constraint (*11*) in the form of $3.6 \times 10^{-4}/l^2$, where l is the spherical harmonic degree. The a priori constraint was necessary, with the current four-way data coverage, to stabilize solutions when expanded up to degree and order beyond 60. The leading constant of the constraint is scaled under the equal stress assumption of Earth and Moon (*9*) in the same manner as previous gravity models (*8*, *9*). Because Rstar's apolune was initially located at 45°S, the southern hemisphere is better covered than the northern hemisphere and, consequently, the gravity field is better constrained in the south and equatorial region.

We estimated a new spherical harmonic model of the lunar gravity field, complete to degree and order 90, from the data sets, designated SGM90d (SELENE Gravity Model with the maximum degree and a version number) (*15*) (Fig. 2). We evaluated the global gravitational field of the Moon at the spheroidal surface with a flattening value of 0.000308935, based on the newly observed second-degree zonal harmonics.

Correlations of coefficients between SGM90d and the SELENE Topography Model (STM) are shown in Fig. 3A. We adopt STM-359_grid-02 from the laser altimeter experiment of SELENE (*16*). For degree greater than 17, correlations of SGM90d are higher than those of LP100K owing to four-way Doppler coverage over the farside and the limb (fig. S2). In particular, a large negative correlation of LP100K for degree 20 that was attributed to five principal nearside mascons (*9*) is not discernible in SGM90d.

The power and error spectrum of SGM90d (Fig. 3B) shows improvement for the degrees lower than 60 with respect to LP100K. Besides, the error spectrum illustrates the substantial contribution of the four-way data: The errors are much less dominated by the a priori power law. This dominance of the four-way data still holds even if we increase the weight of the four-way data from 1 mm/s to 5 mm/s: The error level increases but maintains its slowly increasing trend, in contrast to the convex shape for LP100K. Performance of orbit determination by SGM90d is examined by using LP tracking data during July 1998. SGM90d proves to give as good a data fit as LP100K on the nearside (table S2). Discrepancies are left among different gravity models depending on a priori constraint, data weight, and coverage.

On the nearside, a comparison of Fig. 2 with a previous lunar gravity model reveals general agreement (fig. S3A). The five principal gravity highs in red color on Imbrium, Serenitatis, Crisium, Nectaris, and Humorum are clearly visible, as in the previous models. The mechanism to support the positive gravity anomaly of the nearside mascons remains controversial. Combined gravitational attraction of lava fills in the mare basin (*17*), and uplifted mantle beneath the basins (*18*) are thought to be sources of the positive gravity anomalies; however, the relative contribution of each endmember is difficult to evaluate.

On the farside, in contrast, the gravity field model shows several circular signatures that correspond to topographic structures such as Moscoviense, Freundlich-Sharonov, Mendeleev, Hertzsprung, Korolev and Apollo basins (Fig. 2), unlike linear signatures in previous models (fig. S3B). Free-air gravity anomalies in the South Pole–Aitken Basin Terrane are subdued, with the exception of the Apollo and Planck basins. Many large craters found in topographic models of Clementine (*19*) and the laser altimeter experiment of SELENE appear as spots of negative gravity anomalies without rings of positive anomalies.

There are distinctive differences between the anomalies of the nearside principal mascons and the farside basins. These differences are particularly important for thermal evolution of the Moon because the basin structure possibly reflects the state of the lithosphere in the early development of mare volcanism (*9*, *20*, *21*). As shown previously (*8*, *9*), the nearside principal mascons have sharp shoulders, with a gravity plateau and a weakly negative gravity anomaly in the surroundings (fig. S5, A and B). In contrast, the farside basins are characterized by concentric rings of positive and negative anomalies. For example, gravity anomalies over Korolev (fig. S5C) have a central gravity high in a ring of negative anomalies within the basin, surrounded by a positive anomaly. These gravity signatures were suggested as negative anomalies at first (*8*, *22*); however, Konopliv and colleagues (*9*) proposed that the farside had six "mascons" (Hertzsprung, Moscoviense, Korolev, Freundlich-Sharonov, Coulomb-Sarton, and Dirichlet-Jackson) and that the positive circular anomalies do not match with topographic highs around the basins. On the contrary, the circular gravity highs in Fig. 2 agree well with the topographic rims of the basins revealed by STM-359_grid-02 (*16*). These characteristics of the farside basins are like those of basins on the limb that used to be recognized as mascon basins. In our gravity model, Orientale, Mendel-Rydberg, Lorentz, and Humboldtianum show more affinity with the farside basins than the nearside principal mascons.

As shown in Fig. 2, Korolev (fig. S5, C and D), Mendeleev, Planck, and Lorentz basins have sharp central peaks of which the magnitude in free-air anomalies is almost equivalent to the one in Bouguer anomalies. On the other hand, Orientale, Mendel-Rydberg, Humboldtianum, Moscoviense (fig. S5, E and F), and Freundlich-Sharonov basins have a broad peak of which the magnitude in free-air anomalies is 20 to 60% smaller than the one in Bouguer anomalies. We call the former basins Type I and the latter Type II (Fig. 4 and Table 1). Most Type II basins, except for the Hertzsprung basin, are associated

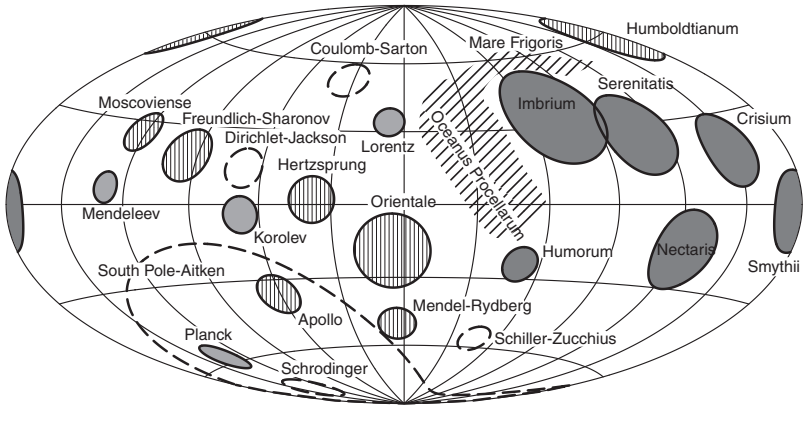


Fig. 4. Locations of major lunar basins. The lunar nearside is on the right side of the figure, and the farside is on the left. The western limb of the Moon viewed from Earth (270°E) is at the center front. Lightly shaded, Type I; hatched, Type II; shaded, the principal mascon; dashed lines, unclassified.

with mare fill, for example, in Orientale, Mendel-Rydberg, Humboldtianum, Apollo, and Moscoviense basins (*23*, *24*). The Freundlich-Sharonov basin has been regarded to have only a small amount of mare basalts inside (*23*, *25*); however, it is possible that more abundant mare basalt could be covered by ejecta from the surrounding crust because the Freundlich-Sharonov basin belongs to the oldest group in Table 1. On the other hand, the presence of mare basalts does not necessarily correspond to positive gravity anomalies. Some areas identified to be covered by mare basalts (*23*) do not show positive anomalies in Fig. 2.

To calculate the Bouguer gravity anomaly map (Fig. 5), we adopted the STM-359_grid-02 (*16*) and assume that the crustal density is 2800 kg m^{-3} (*19*). On the nearside, the general characteristics of the Bouguer gravity anomalies agree well with previous models. The farside is composed of two terranes; gravity is high in the southern hemisphere and low in the northern hemisphere. These divisions correspond to the South Pole–Aitken Basin Terrane (SPAT) and the Feldspathic Highland Terrane (FHT) proposed in (*3*). Higher Bouguer anomalies in SPAT indicate thinner crust beneath this terrane and vice versa in FHT. The subdued appearance of gravity anomalies inside both terranes suggests that the crust-mantle boundary has low topography. A third proposed terrane (*3*), Procellarum KREEP Terrane (PKT), is an area between Oceanus Procellarum and Mare Serenitatis. The Bouguer gravity anomalies in Oceanus Procellarum and Mare Frigoris (Figs. 2 and 4) are weakly positive in SGM90d (Fig. 5), perhaps reflecting gravitational attraction from mare basalts covering the entire Oceanus Procellarum and Mare Frigoris.

Concentric rings of positive and negative gravity anomalies of Type I basins in the free-air gravity anomaly map (Fig. 2 and fig. S5C) are not evident in the Bouguer anomaly map (Fig. 5 and fig. S5D). This change implies that the surface topography is a dominant source of these ringlike gravity anomalies. On the other hand, the gravity high at the center relative to the surrounding floor of the Type I basins remains in the Bouguer anomaly (fig. S5, C and D). Similarly the anomalies of Type II basins disappear in the

Table 1. Classification of major basins with distinctive gravity anomalies (*21*, *25*, *32*), listed in order of relative age. Locations of basins are shown in Fig. 4.

Basin	Type*	Age	Mare volcanism	Diameter (km)†	Central high in free-air gravity anomaly (mGal) relative to surrounding floor	Central high in Bouguer gravity anomaly (mGal) relative to surrounding floor
Orientale	II	Imbrian	Y	1900, 1300, 930, 620, 480, 320	500	700
Imbrium	PM	Imbrian	Y	1700, 1160, 790, 550	350	350
Hertzsprung	II	Nectarian		570, 380, 255, 150	300	400
Serenitatis	PM	Nectarian	Y	1800, 1300, 920, 620, 410	450	600
Crisium	PM	Nectarian	Y	1600, 1080,740, 540, 360	500	650
Humorum	PM	Nectarian	Y	1195, 800, 570, 425, 340, 210	400	600
Humboldtianum	II	Nectarian	Y	1350, 1050, 650, 460, 340, 250	400	500
Mendeleev	I	Nectarian		365, 140	200	200
Korolev	I	Nectarian		810, 590, 440, 220	300	300
Moscoviense	II	Nectarian	Y	630, 420, 300, 220, 140	400	900
Mendel-Rydberg	II	Nectarian	Y	630, 420, 300, 200	300	650
Nectaris	PM	Nectarian	Y	1320, 860, 620, 400, 240	250	600
Apollo	II	Pre-Nectarian	Y	720, 480, 240	300	400
Freundlich-Sharonov	II	Pre-Nectarian	Limited	600	250	600
Planck	I	Pre-Nectarian	Y	325, 160	300	300
Lorentz	I	Pre-Nectarian		365, 170	250	250
Smythii	PM	Pre-Nectarian	Y	1130, 740, 540, 370, 260	300	600

*I, II, and PM indicate Type I, Type II, and primary mascon basins, respectively. †Lunar multiring basins have more than one ring. The major topographic basin rim is underlined.

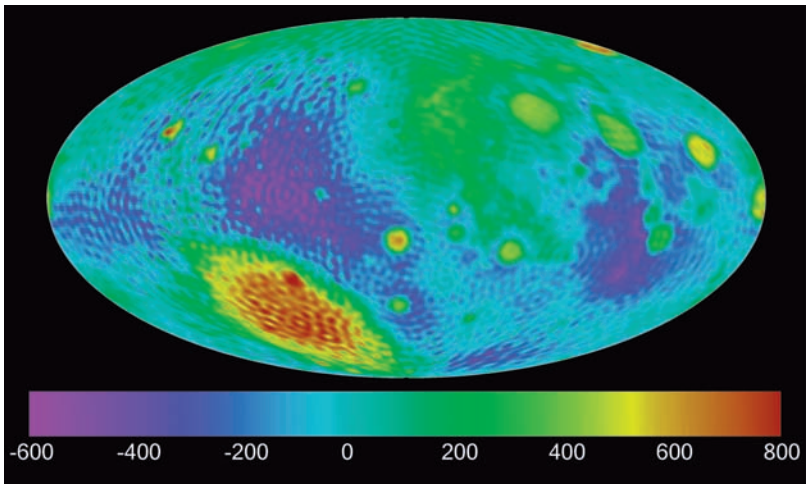


Fig. 5. Bouguer gravity anomaly map from SGM90d. Units are mGal.

Bouguer anomaly map (fig. S5, E and F), except for the prominent central gravity high relative to the surrounding floor, which is enhanced. The primary mascon basins on the nearside are distinguished from Type I and II basins by a weakly to nonnegative anomaly in the surroundings (fig. S5A). The diameter of the gravity plateau is larger than that of Type II basins. In addition to the five principal mascon basins, we include the Smythii basin in this type (Table 1). This classification is ambiguous, and transition from Type I to primary mascon basins could be gradual.

The central gravity high of Type I basins in Bouguer anomalies suggests the existence of excess mass below the center. Because mare fill is absent from Type I basins, the central gravity high is most likely a manifestation of mantle uplift beneath the basin. The magnitude of the central high of Type I basins of 200 to 300 mGal corresponds to an undulation of the crust-mantle boundary of 10 to 15 km. Estimates of the depth of Type I basins from their diameters (*26*) are 3 to 4 km. This depth is too shallow for the central gravity high to be attributed to topographic relaxation of the basins. We therefore consider that the crust-mantle surface was overcompensated, as proposed from a dynamic mantle rebound model (*20*).

The free-air gravity map (Fig. 2) illustrates that the topography and overcompensation beneath the center of Type I basins have been supported by either rigid crust or the lithosphere beneath the FHT. The topography and gravity anomalies of Type I therefore rule out both the viscous relaxation due to flow in the lower part of the thick crust and the isostatic compensation of the lithosphere. Such a stiff lower crust may be unexpected; however, recent studies of rock rheology under dry conditions show that plagioclase is even harder than olivine at temperatures lower than 1000 K (*27*, *28*). We estimate that viscosity of the 100-km-thick crust is ~10^{24} Pa·s for viscous relaxation to occur from viscous layer model calculations (*29*), for a flow time scale of 1 billion years (*14*). If the lower crust is either nearly equal or harder than the upper mantle, the effective viscosity of the lithosphere is required to be >10^{27} Pa s. The lower bound of the lithospheric viscosity implies that the temperature of the crust-mantle boundary was between about 700 and 800 K beneath FHT at the time of the basin formation (*27*, *28*, *30*).

Peak height of positive Bouguer anomalies of Type II ranges from 400 to 900 mGal in comparison to those in free-air anomalies, which range from 250 to 500 mGal. This difference can be attributed to local compensation at the center of the Type II basins. As for the Type I basins, the concentric rings of positive and negative free-air anomalies suggest that rims and topographic depression of Type II basins is not compensated and is supported by a rigid lithosphere. Therefore, neither the viscous relaxation nor the elastic compensation is plausible as a mechanism of the central compensation of the Type II basins. We propose a brittle deformation resulting from a load of uplifted mantle. Mare volcanism is inferred as a result of a fault system developed at the center of the Type II basins.

The inferences on overcompensation, mare volcanism, and topographic relaxation of Type I and II basins also constrain models of the formation of the principal mascon basins on the nearside. Table 1 shows little relation between the class and formation age. On the other hand, there are fewer large lunar basins on the farside. It is unlikely that large impacts concentrated on one side of the Moon and smaller impacts on the other side, or that a thinner crust on the nearside resulted in larger basins than on the farside, because crater diameter depends mostly on impacting energy and momentum, not on the properties of the target (*31*). A plausible hypothesis is that the primary mascon basins on the nearside have deformed more after their initial formation. Tectonic ridges and troughs on the nearside maria indicate that deformation continued after emplacement of mare basalt. The generally lower magnitude of the positive gravity plateau of primary mascon with respect to Type II basins further implies that not only the surface but also the crust-mantle boundary relaxed. Studies of rock rheology (*27*, *28*) suggest that a temperature at the crust-mantle boundary higher than 800 K allows isostatic compensation in ~1 billion years (*14*).

The SELENE lunar gravity field implies that, since the formation of pre-Nectarian basins, the lithosphere of the farside has been sufficiently thick to support the topography of Type I and II basins and that the temperature at the crust-mantle boundary has been less than 800 K. On the nearside, deformation of the primary mascon basins continued during and after the onset of basalt eruption, leading to tectonic deformation even after the cessation of basalt eruption, and temperatures in the crust and the mantle were higher than 1000 K, allowing viscous relaxation in the lower part of the crust.

References and Notes

1. S. R. Taylor, *Lunar Science: A Post-Apollo View* (Pergamon, New York, 1975).
2. J. J. Gillis, B. L. Jolliff, R. L. Korotev, *Geochim. Cosmochim. Acta* **68**, 3791 (2004).
3. B. L. Jolliff, J. J. Gillis, L. A. Haskin, R. L. Korotev, M. A. Wieczorek, *J. Geophys. Res.* **105**, 4197 (2000).
4. R. L. Korotev, B. L. Jolliff, R. A. Zeigler, J. J. Gillis, L. A. Haskin, *Geochim. Cosmochim. Acta* **67**, 4895 (2003).
5. T. Arai, H. Takeda, A. Yamaguchi, M. Ohtake, *Earth Planets Space* **60**, 433 (2008).
6. Kaguya is the Japanese nickname of SELENE. Rstar and Vstar also have nicknames, Okina and Ouna.
7. P. M. Muller, W. L. Sjogren, *Science* **161**, 680 (1968).
8. F. G. Lemoine, D. E. Smith, M. T. Zuber, G. A. Neumann, D. D. Rowlands, *J. Geophys. Res.* **102**, 16339 (1997).
9. A. S. Konopliv, S. W. Asmar, E. Carranza, W. L. Sjogren, D. N. Yuan, *Icarus* **150**, 1 (2001).
10. R. Floberghagen, *Lunar Gravimetry: Revealing the Far-Side* (Springer, New York, 2002).
11. W. M. Kaula, *Theory of Satellite Geodesy* (Blaisdell, Waltham, MA, 1966).
12. T. Iwata *et al.*, *J. Geod. Soc. Japan* **47**, 558 (2001).
13. M. Kato, S. Sasaki, K. Tanaka, Y. Iijima, Y. Takizawa, *Adv. Space Res.* **42**, 294 (2008).
14. Materials and methods are available as supporting material on *Science Online*.
15. SGM90d is the latest version of a series of preliminary gravity models developed by NAO]. There is another series of lunar gravity models [SELENE Gravity Model Tsukuba (SGMT)] developed by JAXA for the purpose of operation and orbital prediction of SELENE. While NAO] uses GEODYN II/SOLVE programs, JAXA developed programs called Lunar Gravity Estimation Function on the basis of their Unified Flight Dynamics System. In the discussions, we focus on common features among SGM90d and gravity models developed by JAXA.
16. H. Araki *et al.*, *Science* **323**, 897 (2009).
17. R. J. Phillips, J. E. Conel, E. A. Abbot, W. L. Sjogren, J. B. Morton, *J. Geophys. Res.* **77**, 7106 (1972).
18. D. U. Wise, M. T. Yates, *J. Geophys. Res.* **75**, 261 (1970).
19. M. T. Zuber, D. E. Smith, F. G. Lemoine, G. A. Neumann, *Science* **266**, 1839 (1994).
20. G. A. Neumann, M. T. Zuber, D. E. Smith, F. G. Lemoine, *J. Geophys. Res.* **101**, 16841 (1996).
21. P. D. Spudis, *The Geology of Multi-Ring Impact Basins: The Moon and Other Planets* (Cambridge Univ. Press, Cambridge, 1993).
22. M. P. Ananda, *J. Geophys. Res.* **103**, 3709 (1977).
23. J. L. Whitford-Stark, *Moon Planets* **26**, 323 (1982).
24. M. J. S. Belton *et al.*, *Science* **255**, 570 (1992).
25. D. E. Wilhelms, J. F. McCauley, N. J. Trask, *The Geologic History of the Moon, U.S. Geol. Surv. Prof. Pap. 1348* (U.S. Government Printing Office, Washington, DC, 1987).
26. $R = 1.83\ D^{0.10}$, where R and D are apparent depth and diameter in kilometers, respectively (*33*).
27. E. Rybacki, G. Dresen, *J. Geophys. Res.* **105**, 26017 (2000).

28. J. Korenaga, S. Karato, *J. Geophys. Res.* **113**, B02403 (2008).
29. S. C. Solomon, R. P. Comer, J. W. Head, *J. Geophys. Res.* **87**, 3975 (1982).
30. Dislocation creep of olivine is assumed to be the dominant deformation process in the mantle.
31. H. J. Melosh, *Impact Cratering: A Geologic Process* (Oxford Univ. Press, Oxford, 1989).
32. A. C. Cook, T. R. Watters, M. S. Robinson, P. D. Spudis, *J. Geophys. Res.* **105**, 12023 (2000).
33. R. J. Pike, *Proc. Lunar Planet. Sci. Conf.* **8**, 3427 (1977).
34. We thank the engineers of NEC/Toshiba Space Systems Ltd., Nippon Antenna Co. Ltd., and Japan Aircraft Mfg. Co. Ltd. and the entire staff of the SELENE mission, and especially F. Fuke, who passed away 2 months after the launch. We thank D. Rowlands for help with modeling the four-way Doppler observable; F. Lemoine for discussions; and T. Arai, H. Takeda, C. K. Shum, and two anonymous reviewers.

Supporting Online Material
www.sciencemag.org/cgi/content/full/323/5916/900/DC1
Materials and Methods
Figs. S1 to S7
Tables S1 and S2
References

3 November 2008; accepted 6 January 2009
10.1126/science.1168029

Long-Lived Volcanism on the Lunar Farside Revealed by SELENE Terrain Camera

Junichi Haruyama,[1]* Makiko Ohtake,[1] Tsuneo Matsunaga,[2] Tomokatsu Morota,[1] Chikatoshi Honda,[1] Yasuhiro Yokota,[1] Masanao Abe,[1] Yoshiko Ogawa,[2] Hideaki Miyamoto,[3] Akira Iwasaki,[4] Carle M. Pieters,[5] Noriaki Asada,[6] Hirohide Demura,[6] Naru Hirata,[6] Junya Terazono,[6] Sho Sasaki,[7] Kazuto Saiki,[8] Atsushi Yamaji,[9] Masaya Torii,[10] Jean-Luc Josset[11]

We determined model ages of mare deposits on the farside of the Moon on the basis of the crater frequency distributions in 10-meter-resolution images obtained by the Terrain Camera on SELENE (Selenological and Engineering Explorer) (Kaguya). Most mare volcanism that formed mare deposits on the lunar farside ceased at ~3.0 billion years ago, suggesting that mare volcanism on the Moon was markedly reduced globally during this period. However, several mare deposits at various locations on the lunar farside also show a much younger age, clustering at ~2.5 billion years ago. These young ages indicate that mare volcanism on the lunar farside lasted longer than was previously considered and may have occurred episodically.

Unraveling the volcanic history of the Moon is essential for understanding the origin and thermal evolution of the Moon. Mare basalts are the most common volcanic features on the Moon. However, absolute dating of mare deposits is nontrivial because it ultimately requires radiogenic dating of samples, and only a limited number of lunar samples returned by Apollo and Luna missions are currently available from a few locations on the near lunar nearside. A second method, measurements of crater size-frequency distribution with the use of remotely acquired images, is widely used to date planetary surfaces from which we do not have sample information. This method is based on the general paradigm that a newly created surface will accumulate craters with time (*1*–*3*). A considerable number of lunar maria have been dated by using image data from Lunar Orbiters and Apollo missions (*4*–*10*); however, precise age determinations of most maria on the farside have not been performed because of the lack of spatially high-resolution images (*5*). Therefore, a systematic, high-resolution mapping of the entire lunar surface, particularly the farside, is one of the primary objectives of the SELENE (Selenological and Engineering Explorer) (Kaguya) mission.

The Terrain Camera (TC) carried on SELENE is a panchromatic push-broom imager with two optical heads (TC1 and TC2) to acquire stereo data for the entire surface of the Moon (*11*). The TC is an optical part of the Lunar Imager/Spectrometer (LISM) (*11*, *12*). The slant angles of TC1 and TC2 are ±15°, relative to the spacecraft flight direction from the nadir vector. Each head has a linear charge-coupled device sensor of 4096 pixels. The instantaneous field of view is 0.00553° and the sampling interval is 6.5 ms, corresponding to 10-m cross- and along-track resolutions, respectively, on the lunar surface at the SELENE nominal altitude of 100 km. The width of the TC swath is 35 km in the normal-swath mode, which provides sufficient overlaps in TC data of sequential observations for producing mosaic images.

SELENE was launched on 14 September 2007 and inserted into a lunar polar orbit on 4 October 2007. Since the beginning of the nominal mission phase on 21 December 2007, the TC has now obtained high-resolution images of most areas of the lunar farside under solar elevation angles lower than ~10°. These images cover previously undated mare deposits on the farside, including those in the South Pole–Aitken (SPA) basin and the Moscoviense basin (e.g., Figs. 1A and 1B are mosaicked images of Antoniadi crater in SPA and Mare Moscoviense, respectively; enlarged figures of Fig. 1, A and B, are shown in figs. S1 and S2).

By using the newly obtained TC data, we performed crater size-frequency distribution measurements for the mare deposits in both SPA and the Moscoviense basin (Table 1). For each region, craters were manually counted using images on computers. Because age determination is sensitive to any contamination by secondary craters (*13*), we excluded areas covered by long shadows from higher terrains and occupied by obvious secondary craters based on their morphological characteristics, such as chain craters, elliptical craters, and clusters (*14*). From these data, absolute model ages of conspicuous mare deposits on the farside of the Moon were obtained. Our results are complimentary to model ages of the nearside mare deposits, which have been studied extensively (*8*–*10*).

We assume that a lunar crate size-frequency distribution (CSFD) can be expressed by the following polynomial for 100 m < D < 200 km proposed by (*2*, *3*, *15*)

$$\log_{10}N(D) = a_0 + \sum_{n=1}^{11} a_n[\log_{10}(D)]^n \quad (1)$$

where D is the crater diameter in kilometers and $N(D)$ is the number of craters with a diameter ≥D per square kilometer. We assigned commonly accepted values for the coefficients of a_1 to a_{11} (*2*, *3*). By fitting an observed CSFD to Eq. 1, we derived the coefficient a_0 [where $a_0 = \log_{10}N(1)$], which gives the age t of the unit in billion years ago (Ga) by using the cratering chronology curve (*2*, *3*, *15*) expressed by

$$N(1) = 5.44 \times 10^{-14}[\exp(6.93 \times t) - 1] + 8.38 \times 10^{-4}t \quad (2)$$

The model ages from crater counts are principally limited by the statistic error (*16*). The quality of the age determination thus improves statistically when the number of countable craters increases and craters are measured over a wide diameter range. In other words, acquisition of images with high spatial resolution and under favorable illumination geometries is crucial, especially for determining model ages of small and/or young areas, where only a few craters are visible in lower-resolution

[1]Institute of Space and Astronautical Science, Japan Aerospace Exploration Agency, 3-1-1 Yoshino-dai, Sagamihara, Kanagawa 229-8510, Japan. [2]Center for Global Environmental Research, National Institute for Environmental Studies, 16-2 Onogawa, Tsukuba, Ibaraki 305-8506, Japan. [3]University Museum, University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan. [4]Department of Aerospace and Astronautics, University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8656, Japan. [5]Department of Geological Sciences, Brown University, Providence, RI 02912, USA. [6]University of Aizu, Ikki-machi, Aizu-wakamatsu, Fukushima 965-8580, Japan. [7]RISE Project Office, National Astronomical Observatory of Japan, 2-12 Hoshigaoka, Mizusawa, Oshu 023-0861, Japan. [8]Department of Earth and Space Science, Graduate School of Science, Osaka University, 1-1 Machikaneyama, Toyonaka 560-0043, Osaka, Japan. [9]Division of Earth and Planetary Sciences, Kyoto University, Sakyo-ku, Kyoto 606-8502, Japan. [10]Fujitsu, Chiba 261-8588, Japan. [11]Space Exploration Institute CP 774, CH-2002 Neuchatel, Switzerland.

*To whom correspondence should be addressed. E-mail: haruyama.junichi@jaxa.jp

images. The high-resolution images from the TC that are taken at low sun illumination geometries are particularly suitable for dating small deposits such as those in the SPA basin. The statistical error of individual data points in our crater frequency measurements is mostly <20% (1σ); the errors of ages based on the model (*2*, *3*) are ±0.02 Ga for >3.5 Ga, 0.01 to 0.03 Ga for between 3.2 and 3.0 Ga, and <20% for <3.0 Ga (*8*).

The SPA basin is the largest and deepest impact basin on the Moon: more than 8 km in depth and extending ~2500 km from the south pole to the crater Aitken, which is located 15° south of the equator (*5*). Although a few ancient cryptomaria were observed within the basin (*17*), based on stratigraphical analyses, most mare deposits in the SPA basin were thought to have formed during the late Imbrian epoch (3.85 to 3.8 Ga) (*5*, *18*–*21*). On the basis of our CSFD measurements, we find that most mare deposits are late Imbrian in age (Table 1), consistent with previous stratigraphical work (*5*, *18*, *19*). However, several mare deposits show younger model ages up to Eratosthenian. For example, in the 143-km diameter Antoniadi crater (70°S, 172°E), there is an apparently smooth, young-looking mare deposit (Fig. 1A), which is the southernmost mare deposit in the SPA basin. The derived model age of this deposit is as young as 2.58 Ga (Fig. 2A); hence, it is substantially younger than the Imbrian-Eratosthenian boundary (3.2 Ga).

For several mare deposits, we can fit multiple production function curves to our CSFD, similar to observations of some maria on the nearside (*22*). Apollo N, the easternmost mare deposit in the SPA basin, which has been previously classified as late Imbrian in age (*5*, *7*, *18*), is used as an example. The CSFD (Fig. 2B) for craters larger than 700 m in diameter gives a 3.51-Ga model age of the deposit, which is consistent with previous interpretations (*7*). However, when we assess craters <500 m in diameter by taking advantage of the TC's high-resolution, we derive a second model age of 2.49 Ga (Fig. 2B). It is difficult for this distribution to be generated by adding only a CSFD from an acceptable power-law distribution of secondary craters. A reasonable interpretation for the two distinct model ages is that Apollo N basalts have been formed by multiple eruptions, as observed for several maria on the nearside, including those in Oceanus Procellarum, Imbrium, and Tranquillitatis (*22*). In our interpretation, this Apollo N area has been resurfaced by a basaltic eruption at ~2.5 Ga, where the erupted lava did not completely obliterate preexisting craters on a deposit formed at ~3.5 Ga. The data suggest the size of obliterated craters is less than several hundred meters. Thus, if we consider that the diameter–to–rim-height ratio of a typical lunar crater of <1 km is <0.05 (*23*), the thickness of the younger deposit can be estimated as <40 m.

Mare Moscoviense fills a part of the 550-km diameter Moscoviense basin (27°N, 146°E) (Fig. 1B), which is in the northern hemisphere of the lunar farside. Within the mare, three units have been identified on the basis of color difference in the Clementine ultraviolet-visible multispectral images: (i) an oldest basaltic unit with low estimated FeO content in the southern part, (ii) a northwestern basaltic unit with low TiO_2 content, and (iii) an eastern younger unit with higher TiO_2 content (*24*). The southern unit appeared to be saturated for craters smaller than several hundred meters in diameter. Whereas the CSFD of the northwestern unit reveals a model age of 3.50 Ga, the eastern unit is younger and has a model age of 2.57 Ga (Fig. 2C). The model age of the eastern unit for craters larger than 1 km is close to that of the northwestern unit, which we interpret to indicate that the eastern unit is superposed on an older unit of 3.50 Ga in age. Because of their

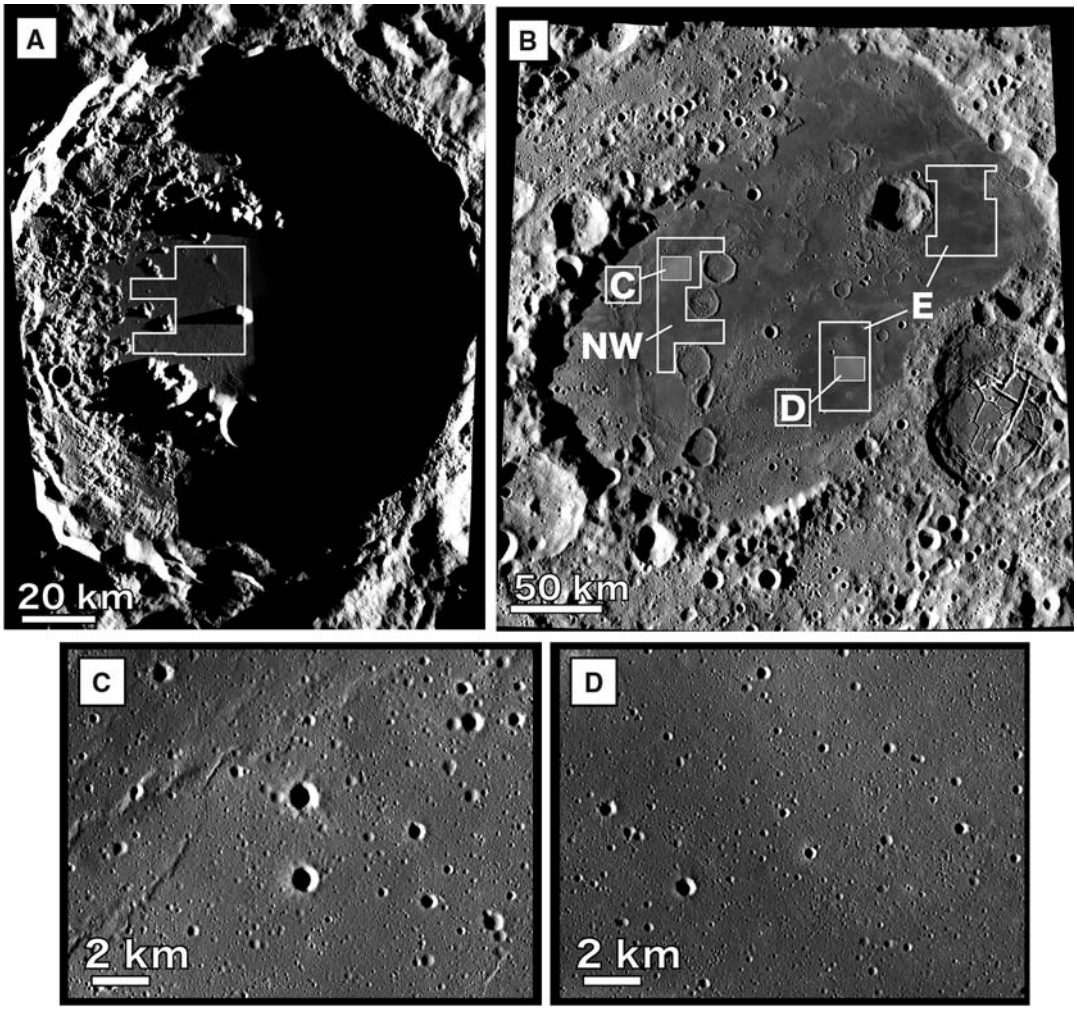


Fig. 1. Mosaics of TC images. (**A**) Antoniadi crater, which shows a dark and smooth mare deposit on its floor. Crater counting was performed for the area outlined by the white polygon. (**B**) Mare Moscoviense, which is filled with mare deposits within its 210-km-diameter inner ring. Crater counting was performed for areas outlined by the white polygons (Table 1). E and NW denote the eastern and northwestern areas, respectively. (**C** and **D**) Close-up images of the mare deposits in Mare Moscoviense. The locations of these deposits are shown as hatched gray boxes in (B). These images are Transverse Mercator map projections.

similar model ages, we propose that the eastern (later) basaltic eruption might have spread over parts of the northwestern unit but did not completely obliterate the previously formed craters on the underlying northwestern unit. We estimated the younger, uppermost unit is 30 to 50 m thick. Several other deposits on the farside show similar characteristics to those in Apollo N and Mare Moscoviense, indicating a potentially complex history of these farside deposits.

Frequency distributions of all estimated ages for maria on the lunar farside (Fig. 3A) from both our study (black bars) and previous studies (gray bars) (*6*–*8*) indicate that most mare units on the lunar farside formed before 3.0 Ga. Because most mare on the nearside are also older than ~3.0 Ga (*8*–*10*) (Fig. 3B), it appears that the global volcanic activity on the Moon declined markedly after 3.0 Ga. However, it is notable that some mare units on the farside show younger model ages clustered around 2.5 Ga (Fig. 3A). Basaltic volcanism also continued on the nearside, but in contrast, the youngest model ages of basaltic unit extend to nearly 1.0 Ga in Oceanus Procellarum.

The global cessation of most lunar mare volcanism by ~3.0 Ga suggests a similar early lunar thermal history for the nearside and farside. Thus, the internal structure of the Moon and its cooling and/or heating during this early period appear to be controlled by the same constraints on a global scale. In combination with previous CSFD analyses (*9*), our results indicate that the timing of the termination of mare basalt volcanism differs between the farside (2.5 Ga) and the nearside (1.2 Ga). This difference might be related to a larger crustal thickness on the lunar farside, which makes it harder for magma to reach the surface (*25*), and/or a deficiency of heat-producing radiogenic elements on the farside as compared with the nearside (*26*). It should be noted that the young mare deposits are widespread throughout the farside, from the Moscoviense basin in the

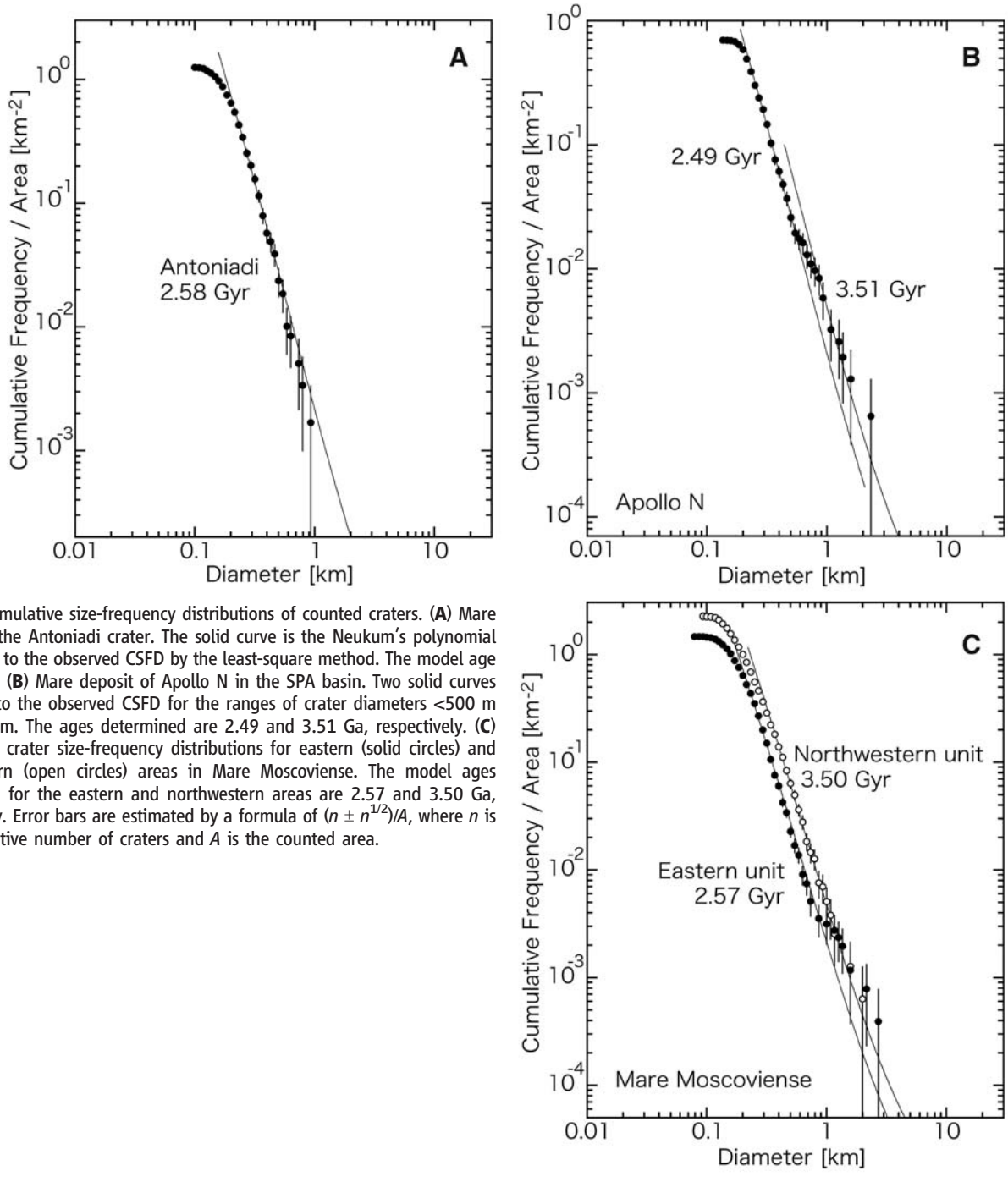


Fig. 2. Cumulative size-frequency distributions of counted craters. (**A**) Mare deposit in the Antoniadi crater. The solid curve is the Neukum's polynomial function fit to the observed CSFD by the least-square method. The model age is 2.58 Ga. (**B**) Mare deposit of Apollo N in the SPA basin. Two solid curves can be fit to the observed CSFD for the ranges of crater diameters <500 m and >700 m. The ages determined are 2.49 and 3.51 Ga, respectively. (**C**) Cumulative crater size-frequency distributions for eastern (solid circles) and northwestern (open circles) areas in Mare Moscoviense. The model ages determined for the eastern and northwestern areas are 2.57 and 3.50 Ga, respectively. Error bars are estimated by a formula of $(n \pm n^{1/2})/A$, where n is the cumulative number of craters and A is the counted area.

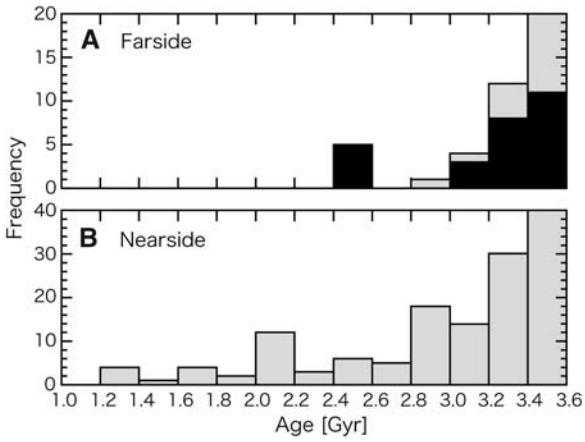


Fig. 3. Histograms of the model ages of mare deposits on the farside (**A**), including those in Mare Moscoviense, the SPA basin, Tsiolkovsky (*6*), Mare Orientale (*7*), and Mare Australe (*8*), and on the nearside (**B**), including those in Mare Imbrium, Serenitatis, Humorum, Tranquillitatis, Humboldtianum (*8*), Oceanus Procellarum, Mare Nubium, Cognitum, Insularum (*9*), and Mare Fecunditatis (*10*). Black bars indicate mare deposits, whose ages are newly determined by this study.

Table 1. Model ages for maria on the lunar farside.

Mare		Area (km^2)	SEA* (degrees)	$N(1)$† (km^{-2})	Age (Ga)	Previous studies‡
Poincare	S	734.0	2.1	0.00284	3.17	Im (*5*)
	NW	490.6	2.0	0.00433	3.46	Im (*5*)
	E	1067.2	2.5	0.00277	3.13	Im (*5*)
				0.00504	3.52	
Ingenii		1291.8	9.9	0.00292	3.20	Im (*5, 18*)
				0.00489	3.51	
Antoniadi		592.7	2.8	0.00217	2.58	Im (*5*); EIm (*19*)
Chretien		1183.1	2.3	0.00315	3.28	Im (*5, 18*)
				0.00427	3.46	
Jules Verne		815.2	11.1	0.00343	3.34	Im (*5, 18*)
Aitken		1257.1	10.9	0.00315	3.28	Im (*5, 18*); 3.8 Ga (*4*)
				0.00469	3.49	
Apollo N		1541.5	5.5	0.00209	2.49	Im (*5, 18*), 3.63 Ga (*7*)
				0.00491	3.51	
Apollo S		778.4	5.1	0.00205	2.44	Im (*5, 18*), 3.63 Ga (*7*)
Von Karman		1628.6	3.0	0.00349	3.35	Im (*5, 18*)
Von Karman M		1220.6	2.9	0.00364	3.38	Im (*5, 18*)
Leibnitz		926.0	3.6	0.00410	3.44	Im (*5, 18*)
				0.00473	3.50	
Rumford		239.0	4.5	0.00346	3.35	Im (*5, 18*)
Maksutov		615.3	4.3	0.00311	3.27	Im (*5, 18*)
				0.00451	3.48	
Nishina		596.7	4.1	0.00207	2.47	Im (*5*)
				0.00282	3.16	
				0.02406	3.85	
Moscoviense	NW	1576.1	11	0.00481	3.50	Im (*5, 18*)
	E	2543.5	11	0.00216	2.57	Im (*5, 18*)
				0.00562	3.55	

*Solar elevation angle (SEA) of image used for crater counting. †$N(1)$ is the number of craters with a diameter larger than or equal to 1 km per km^2, derived by fitting a polynomial proposed in (*2*, *3*) to observed CSFDs. ‡Im, Imbrian mare materials; EIm, Eratosthenian–Imbrian mare materials.

north to within the Antoniadi crater in the south. In addition, our data suggest these young basalts might have erupted during a late pulse of volcanic activities on the lunar farside near ~2.5 Ga. Alternatively, the lack of observed farside volcanism between 2.5 and 3.0 Ga might merely be explained by continuous superposition of younger deposits.

References and Notes

1. W. K. Hartmann, *Icarus* **5**, 406 (1966).
2. G. Neukum, *Meteoritenbombardment und Datierung Planetarer Oberflächen* (Ludwig Maximilians Univ., Munich, Germany, 1983).
3. G. Neukum, B. A. Ivanov, in *Hazards Due to Comets and Asteroids*, T. Gehrels, Ed. (Univ. of Arizona Press, Tucson, AZ, 1994), pp. 359–415.
4. A. S. Walker, F. El-Baz, *Moon Planets* **27**, 91 (1982).
5. D. E. Wilhelms, *The Geologic History of the Moon. U.S. Geological Survey Professional Paper* (U.S. Geological Survey, Washington, DC, 1987).
6. A. Tyrie, *Earth Moon Planets* **42**, 245 (1988).
7. R. Greeley *et al.*, *J. Geophys. Res.* **98**, 17183 (1993).
8. H. Hiesinger, R. Jaumann, G. Neukum, J. W. Head III, *J. Geophys. Res.* **105**, 29239 (2000).
9. H. Hiesinger *et al.*, *J. Geophys. Res.* **108**, 5065 (2003).
10. H. Hiesinger *et al.*, *37th Lunar Planet. Sci. Conf.*, no. 1151 (2006).
11. J. Haruyama *et al.*, *Earth Planets Space* **60**, 243 (2008).
12. LISM consists of the TC, Multi-Band Imager (MI), and Spectral Profiler (SP). MI is a multicolor imager with nine bands in the visible–to–near-infrared range, and SP is a continuous spectral profiling sensor in the same range. See (*11*) for LISM's specifications and operation strategy.
13. A. S. McEwen, E. B. Bierhaus, *Annu. Rev. Earth Planet. Sci.* **34**, 535 (2006).
14. H. J. Melosh, *Impact Cratering: A Geologic Process* (Oxford Univ. Press, New York, 1989).
15. We investigated model dependencies of adopted functions (Eqs. 1 and 2), as compared with two other well-established models. A lunar CSFD function model suggested by Hartmann (*27*), that includes three power laws for different diameter ranges, would yield different model ages over a wide range of crater diameters (*16*, *28*). However, in the diameter range we use (a few kilometers to a few hundred meters), the difference between the functions is <10% (*28*), which gives small differences in the estimated ages. More recently, a different chronology curve between age and number of craters was proposed (*29*). The newer model shows a steeper slope in the age range of >3.75 Ga, where the difference from the model we adopt becomes large. However, in the range of <3.6 Ga, ages from the newer model would differ less than 0.1 Ga from our results.
16. G. Neukum, B. A. Ivanov, W. K. Hartmann, *Space Sci. Rev.* **96**, 55 (2001).
17. C. M. Pieters, L. Gaddis, B. Jolliff, M. Duke, *J. Geophys. Res.* **106**, 28001 (2001).
18. D. E. Stuart-Alexander, *U.S. Geol. Surv. Map I-1047* (1978).
19. D. E. Wilhelms, K. A. Howard, H. G. Wilshire, *U.S. Geol. Surv. Map I-1162* (1979).
20. The lunar geologic periods are divided into pre-Nectarian (from 4.2 to 3.92 Ga), Nectarian (from 3.92 to 3.85 Ga), early Imbrian (from 3.85 to 3.8 Ga), late Imbrian (from 3.8 to 3.2 Ga), Eratosthenian (from 3.2 to 1.1 Ga), and Copernican (from 1.1 Ga to the present) (*5*).
21. Although a detailed documentation of farside cryptomaria is beyond the scope of this discussion, the role of these ancient basalts in early phases of lunar volcanism will become better defined as additional remote sensing data for the Moon are returned.
22. H. Hiesinger *et al.*, *Geophys. Res. Lett.* **29**, 1248 (2002).
23. R. J. Pike, in *Impact and Explosion Cratering*, D. J. Roddy, R. O. Pepin, R. B. Merrill, Eds. (Pergamon, New York, 1977), pp. 489–510.
24. G. Y. Kramer, B. L. Jolliff, C. R. Neal, *J. Geophys. Res.* **113**, E01002 (2008).
25. J. W. Head III, L. Wilson, *Geochim. Cosmochim. Acta* **56**, 2155 (1992).
26. B. L. Jolliff *et al.*, *J. Geophys. Res.* **105**, 4197 (2000).
27. W. K. Hartmann, *Meteorit. Planet. Sci.* **34**, 167 (1999).
28. B. A. Ivanov, *Space Sci. Rev.* **96**, 87 (2001).
29. D. Stöffler, G. Ryder, *Space Sci. Rev.* **96**, 9 (2001).

30. We thank all the contributors to the SELENE (Kaguya) project, especially members of the project management group (Y. Takizawa, M. Kato, S. Sasaki, R. Nagashima, K. Tsuruda, and H. Mizutani) and the Lunar Imager/Spectrometer working group (H. Otake, H. Kawasaki, R. Nakamura, S. Kodama, S. Minami, S. Takechi, A. Akiyama, M. Shirao, T. Yokota, H. Takeda, T. Arai, T. Sugihara, Y. Yamaguchi, T. Hiroi, T. Hashimoto, T. Michikami, K. Kitazato, M. Higa, P. Pinet, T. Nimura, T. Yamamoto, N. Harada, K. Iseki, T. Hodokuma, S. Kikuchi, S. Kawabe, S. Okuno, and T. Takayama), for their efforts in development, operation, and data processing of SELENE and LISM/TC. We also thank the three anonymous reviewers for their helpful comments. This work was supported by KAKENHI (grant 20540416 to J.H. and C.H.), (19204045 to H.D.), and (20•9211 to T.M.).

Supporting Online Material

www.sciencemag.org/cgi/content/full/1163382/DC1
SOM Text
Figs. S1 and S2
Table S1

17 July 2008; accepted 30 September 2008
Published online 6 November 2008;
10.1126/science.1163382
Include this information when citing this paper.

Lunar Radar Sounder Observations of Subsurface Layers Under the Nearside Maria of the Moon

Takayuki Ono,[1] Atsushi Kumamoto,[1] Hiromu Nakagawa,[1] Yasushi Yamaguchi,[2] Shoko Oshigami,[2] Atsushi Yamaji,[3] Takao Kobayashi,[4] Yoshiya Kasahara,[5] Hiroshi Oya[6]

Observations of the subsurface geology of the Moon help advance our understanding of lunar origin and evolution. Radar sounding from the Kaguya spacecraft has revealed subsurface layers at an apparent depth of several hundred meters in nearside maria. Comparison with the surface geology in the Serenitatis basin implies that the prominent echoes are probably from buried regolith layers accumulated during the depositional hiatus of mare basalts. The stratification indicates a tectonic quiescence between 3.55 and 2.84 billion years ago; mare ridges were formed subsequently. The basalts that accumulated during this quiet period have a total thickness of only a few hundred meters. These observations suggest that mascon loading did not produce the tectonics in Serenitatis after 3.55 billion years ago. Global cooling probably dominated the tectonics after 2.84 billion years ago.

Subsurface stratigraphic and tectonic features revealing the formation and evolution of a planetary body can be scanned by a ground-penetrating radar or radar sounder onboard a spacecraft. This was first realized by the Apollo Lunar Sounder Experiment (ALSE) onboard the Apollo 17 Command/Service Module (*1*), which revealed subhorizontal layering in Mare Serenitatis (*2*, *3*). Recently, most of Mars has been probed by radar soundings (*4*–*7*).

The Lunar Radar Sounder (LRS) onboard the Kaguya spacecraft (SELENE) has been exploring the lunar subsurface since 20 November 2007. The LRS uses the HF band (5 MHz), enabling subsurface data to be obtained to a depth of several kilometers (*8*, *9*). The LRS data reveal that most nearside maria have subsurface stratifications (Fig. 1). Many of the reflectors lie at apparent depths of several hundred meters below the surface. The deepest reflector was found in the northeastern Mare Imbrium at an apparent depth of 1050 m. Multiple-layer structures were observed in several mare regions, including the northeastern Imbrium, Crisium, and Oceanus Procellarum.

Subsurface echoes appear as apparent linear echo patterns a few hundred meters below the surface in B-scan display (*10*) (Fig. 2A and fig. S1). Similar echo patterns can also be produced by the range sidelobes of surface echoes. But the intensity of range sidelobes decreases so rapidly as their order increases that they should be negligible at a range of a few hundred meters from the main lobe. Spatially dense observations support the hypothesis that these echoes originate from subsurface horizons. Figure S2 shows B-scan displays of 40 adjacent orbit observations between 20° and 25.6°E. The constant appearance of these subsurface echoes indicates that there are subsurface boundary interfaces that reflect the LRS pulses. Estimating the dielectric constant of the lower subsurface-layer material provides a means for evaluating whether the detected echo

[1]Graduate School of Science, Tohoku University, Sendai 980-8578, Japan. [2]Graduate School of Environmental Studies, Nagoya University, Nagoya 464-8601, Japan. [3]Graduate School of Science, Kyoto University, Kyoto 606-8502, Japan. [4]Mineral Resources Research Division, Korea Institute of Geoscience and Mineral Resources, Daejeon 305-350, Korea. [5]Graduate School of Natural Science and Technology, Kanazawa University, Kanazawa 920-1192, Japan. [6]Department of Space Science and Technology, Fukui University of Technology, Fukui 910-8505, Japan.

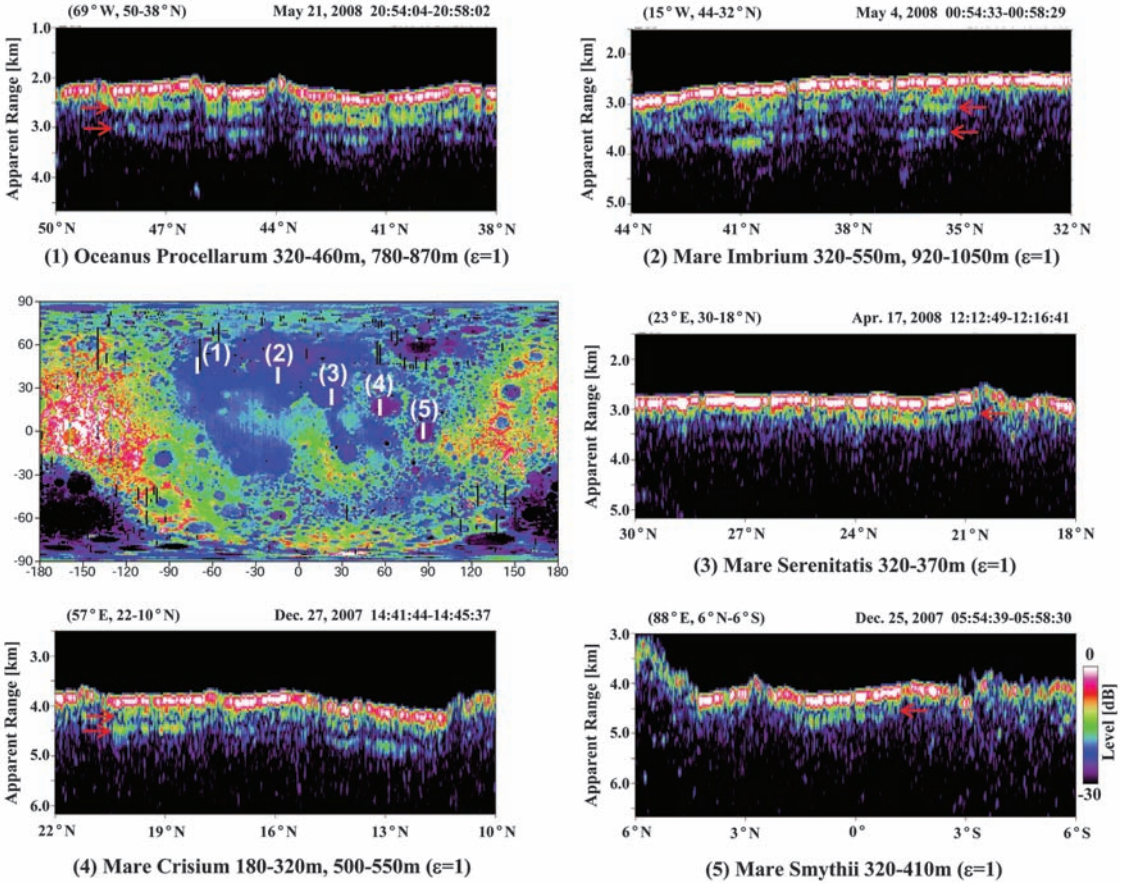


Fig. 1. B-scan (*10*) displays with obvious subsurface echoes (red arrows). Doppler focusing was not applied to the data. Observation ground tracks are plotted on the topographic map (center left panel) that was generated using the LRS as an altimeter. A spherical surface with a radius of 1737.4 km is used as the reference for the vertical coordinates.

is a subsurface reflection. Using a simple Fresnel reflection/refraction model (*11*), we estimated the loss tangent of the surface-layer material and the dielectric constant of the second-layer material (and the dielectric constant of lower-layer material if a second subsurface echo was detected) (Table 1) and found that they were consistent with the Apollo measurement results (*12*).

We attempted to identify the subsurface reflectors reported from the ALSE data for Mare Serenitatis (*2*). Two reflectors were inferred along east-west–trending tracks between 14° and 27°E at depths of ~0.9 and 1.6 km when ε, the relative dielectric constant of lunar rocks, was assumed to be 8.7 (apparent depths of ~2.7 and 4.7 km). Peeples *et al*. (*2*) distinguished subsurface echoes from noise by the multiple-orbit correlation technique, which compares echoes obtained from nearby orbits. However, no such reflectors were detected at apparent depths of ~2.7 and 4.7 km in the LRS radargrams (Fig. 2B and fig. S1) for Serenitatis, although a more detailed analysis is necessary to conclusively show whether the ALSE reflectors exist. Instead, we found prominent reflectors lying at apparent depths of a few hundred meters (e.g., Fig. 2A and fig. S1). Because the ALSE had a range resolution of ~1200 m in free space, it would be difficult to resolve shallow reflectors from the ALSE data. The actual resolution would be ~400 m for bedrock with ε = 8.7; a lower value of ε would reduce this resolution. By contrast, the LRS has a range resolution of ~75 m in free space, permitting detection of shallow reflectors.

Figure 3 and fig. S3 show profiles obtained by the LRS for Serenitatis, where the echo amplitudes were stretched to visualize subsurface features. Subhorizontal prominent reflectors are seen at apparent depths of several hundred meters. Linear topographic features, such as mare ridges parallel to the nadir tracks, can generate linear echo traces similar to those of subhorizontal reflectors in B-scan images, but we eliminated this possibility by estimating the off-nadir distances of the features in question.

The subsurface reflectors are predominantly horizontal, which suggests that they are interfaces within the mare fill; the base of the fill should have a topographic relief as large as that of the surface of Orientale, an impact basin approximately the same size as Serenitatis (*13*). Mare volcanism peaked at 3.6 to 3.7 billion years ago (Ga), several hundred million years after the basin-forming impact event (*14*). The heavy bombardment during this period should have enhanced the undulations. Thus, the reflectors are considered to indicate not the base of mare deposits, but layers within mare deposits. Thus, it is important to consider the stratigraphy of mare units evaluated by Hiesinger *et al*. (*14*). They subdivided the mare into spectrally homogeneous lithologic units, the ages of which were determined by crater counting.

The lithology responsible for these reflectors is unclear, but high-porosity deposits, such as buried regolith and pyroclastic and ejecta deposits, are possibilities. However, pyroclastic and ejecta deposits can be ruled out on the basis of the following observations. First, the reflectors have horizontal dimensions of >200 km. Second, layers represented by some reflectors appear to outcrop at the lunar surface. Also, the strata above and below the reflectors can be correlated to geological units on the surface. An excellent example is the wavy reflector in Fig. 3, which dives at ~17.8°N and extends to 24° to 25°N under the surface. The intersection between this reflector and the surface coincides with the boundary between mare unit S11 (*14*) and its overlying strata. Except at the margins of Serenitatis, the apparent depth of this horizon is 400 to 500 m (Fig. 3B), which corresponds to an actual depth of ~250 m or shallower if the dielectric constant is assumed to be 4 or higher (*12*). The reflector of Serenitatis in Table 1 is correlated by its apparent depth with the prominent reflector in profile d-d′ (fig. S3), which is adjacent to the locations given in Table 1. Thus, the top of S11 lies at an actual depth of ~175 m at around 22.0°N, 23.3°E. Hiesinger *et al*. (*14*) used crater chronology to estimate that this unit formed at 3.55 Ga, whereas the overlying units S15 and S22 were dated at 3.44 and 3.28 Ga, respectively. Considering that the apparent depth is overestimated by a factor of ~2, we see that the reflector identified with the top of S11 to the south of 18°N is an extension of the slope made by the

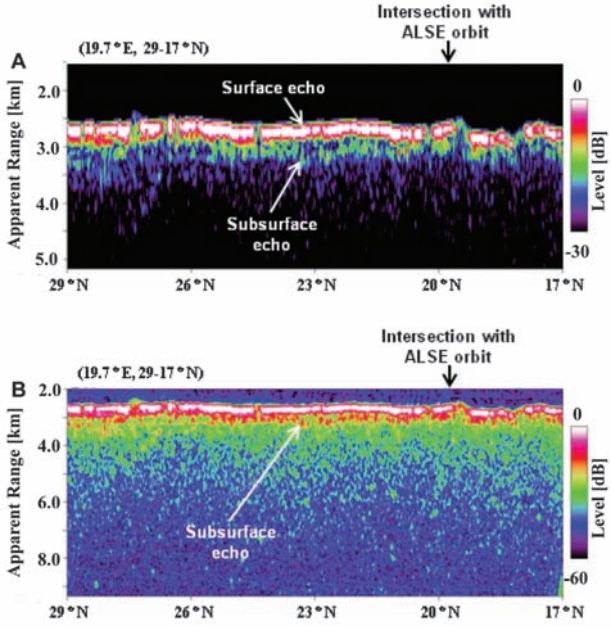


Fig. 2. (**A**) B-scan display (*10*) down to an apparent depth of 2.5 km in Mare Serenitatis. (**B**) Same as (A) but down to an apparent depth of 7 km, indicating the absence of prominent reflectors at apparent depths of ~2.7 and 4.7 km. Doppler focusing was not applied to the data.

Table 1. Mare subsurface stratification at five locations, as inferred from LRS data. Apparent depth was determined by the mean of a number of A-scope data (*10*). The electric properties were estimated by inversely solving a Fresnel model with the assumption that the top-layer material has a dielectric constant of 4. The loss tangent, tan δ, of the lowermost layer was not determined, as the attenuation in that layer is unknown. *d* and *D* are the apparent depth and estimated depth to the upper surface of each layer, respectively; *P* is echo power.

		Procellarum (44.5°N, 69.3°W)	Imbrium (36.0°N, 15.3°W)	Crisium (57.0°N, 15.5°E)	Serenitatis (22.0°N, 23.3°E)	Smythii (1.0°N, 87.4°E)
Data samples		**400**	**460**	**330**	**530**	**410**
Top layer	tan δ	0.001 to 0.019	0.001 to 0.016	0.002 to 0.005	0.001 to 0.019	0.001 to 0.019
Second layer	*d* (m)	320	500	290	350	500
	D (m)	160	250	145	175	250
	P (dB)	−12.2	−12.7	−12.7	−11.4	−12.5
	ε	5.86 to 8.08	5.77 to 9.04	6.45 to 6.76	6.09 to 8.09	5.82 to 10.68
	tan δ	0.001 to 0.010	0.001 to 0.008	0.002		
Third layer	*d* (m)	770	1000	660		
	D (m)	343 to 347	454 to 460	290 to 293		
	P (dB)	−16.0	−15.6	−13.0		
	ε	7.64 to 11.03	7.52 to 11.89	9.25 to 9.71		

surface of this unit to the south of ~17.8°N, where the reflector appears at the surface (Fig. 3C).

The depositional hiatuses represented by the reflectors may have lasted 0.17 to 0.11 billion years, which would allow the regolith to be thicker than 1 m (*15*). Accordingly, we interpret this reflector as being the buried regolith blanketing S11, although the thickness of the layer returning the prominent echoes remains unknown. An identical interface appears at a similar apparent depth in profile e-e′. This interface meets the surface at ~19°N, where S11 is overlain by S15 (fig. S3). In addition, S28 lies above the reflector. Consequently, the layers above S11 have ages between 3.56 and 2.84 Ga. Pyroclastic deposits are unlikely to be the source of the prominent reflectors that meet the surface, because there is no regional pyroclastic deposit with an age between 3.55 and 2.84 Ga in or around Serenitatis (*16*).

Ejecta blankets are unlikely to account for the extensive reflectors. Continuous ejecta extend about one crater radius from the crater rim (*17*). Thus, an ejecta blanket that is the same width as the reflectors would be accompanied by a basin-sized crater, but the tracks extend beyond the continuous ejecta from the large craters produced between 3.55 and 2.84 Ga. The top of S11 is at an actual depth of <250 m. Thus, the relationship between crater size and rim height (*18*) indicates that basin-sized craters cannot be concealed by strata above this depth.

The reflector identified with the top of S11 extends from the mare's southern margin to its central region. The deeper prominent reflector in profile a-a′ (fig. S3) is correlated with the top of this unit, because the apparent depth of this reflector is generally consistent with that of the interfaces identified with the horizon found in the LRS data from neighboring orbits. The basalts of S11 are known to have a higher TiO_2 content and a lower reflectivity than younger ones (Fig. 3) (*19*, *20*). The ejecta from the Bessel crater (21.8°N, 17.9°E) to the west of Deseilligny have similar spectral characteristics to those of S11, which suggests that the unit lies under the central region of Serenitatis (*21*).

The horizon at the top of S11 is equivalent to the interface between units I and III in the nomenclature of Solomon and Head (*22*), who estimated that the thickness of unit III was 1 km at the margin of the basin and 2 km at its center. They estimated that the total mare fill was 8.5 km thick at the center by assuming that before mare flooding, the basin had a topography similar to that of Orientale. They also estimated the thickness of the mare fill from the gravity anomaly. In contrast, the thicknesses of basalt fills in many maria, as inferred from partially flooded craters, are less than estimates obtained by assuming a prefill topography similar to the present topography of empty basins. De Hon *et al.* (*23*) showed that nearside mare fills are generally thinner than 1 km. The thinner estimate was recently verified by Clementine spectral data from the Humorum basin (*24*). Although the thickness of the mare fill under the top of S11 is unknown, the thickness above this horizon determined by the LRS is inconsistent with the thicker estimate but is closer to the thinner one.

The ALSE data were used to support the tectonic origin of mare ridges [e.g., (*25*)]. Peeples *et al.* (*2*) illustrated an anticline marked by reflectors under a mare ridge in Serenitatis. Although the existence of these reflectors was not verified by the LRS, our data support the hypothesis. The wavy reflector in Fig. 3 and that between 18° and 21°N in profile d-d′ (fig. S3) are subparallel to the surface undulations, which are parts of ridge systems in southern Serenitatis (*25*). The northward continuation of the latter reflector is disturbed by Dorsa Smirnov, a broad linear topographic swell. Thus, basaltic layers are folded under mare ridges.

The thicknesses of the strata above S11 appear to be independent of their positions relative to the ridges; this suggests a tectonic quiescence during the emplacement of the younger units (i.e., from 3.55 to 2.84 Ga). Mare basalts had very low viscosities (*26*), resulting in approximately horizontal lava fields. Their original horizontality should have resulted in thinning of strata toward anticlines if the folds grew during the deposition of the strata (fig. S4). Therefore, post-S11 strata were subject to horizontal shortening to form mare ridges predominantly after the emplacement of S28 (2.84 Ga).

The tectonic quiescence and the thin post-S11 strata are in conflict with the tectonic model of mascon loading (*22*). This model predicts that the accumulation of strata gradually bends the lithosphere by their weight, giving rise to syndepositional tectonics. Accordingly, mare fills were assumed to be several kilometers thick. The superstrata over S11 were thought to be 1 to 2 km thick to have sufficient weight for the model (*22*). However, Freed *et al.* have demonstrated the difficulty of explaining mare ridge formation after S11 emplacement by the strata, even if they were assumed to be 2 km thick (*27*). Instead, they conjecture that the mare ridges were formed by a combination of loading and compression due to global cooling. Our study reveals that the superstrata are about one-tenth the thickness of those predicted by the model. Thus, mascon loading was not responsible for forming the ridges after 3.55 Ga, and global cooling was probably the dominant cause of post–3.55 Ga tectonics, at least in southern Serenitatis.

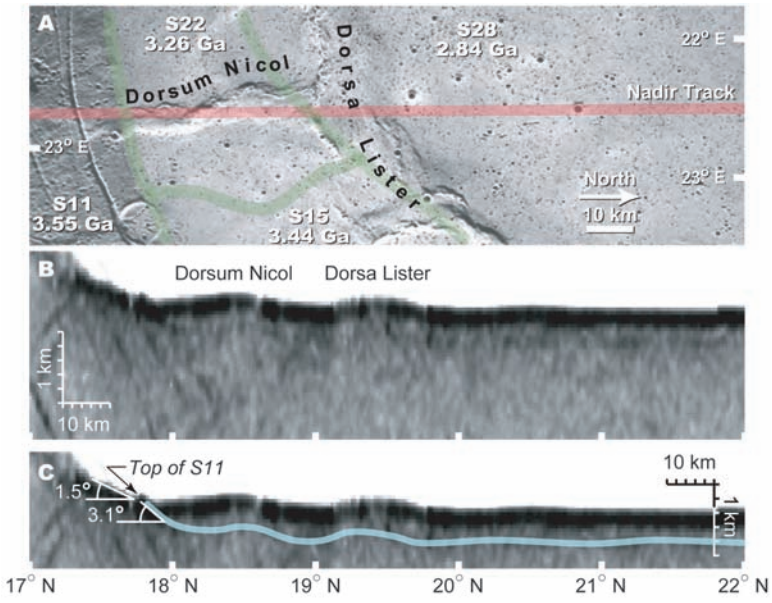


Fig. 3. LRS data along a north-south–trending track in southern Serenitatis. (**A**) Red and green lines show the nadir track of Kaguya and the boundaries of stratigraphic units S11, S15, S22, and S28 defined by (*14*). Solid triangles indicate the boundary between S11 and the other units. S15 and S22 are bounded largely by Dorsum Nicol and are separated from S28 by Dorsa Lister. [Apollo image AS17-M-0452] (**B**) LRS data along the track. This shows a close-up of the southern part of profile c-c′ in fig. S3. (**C**) Prominent reflector found in the profile. This reflector appears at the lunar surface at a latitude of ~17.8°. The apparent dip of the reflector is ~3.1° in the vicinity of this outcrop, giving a true dip of 1° to 2° when the dielectric constant of the rocks is accounted for. This value is consistent with the surface slope to the south of this latitude.

References and Notes

1. R.J. Phillips *et al., NASA Spec. Publ. 330* (1973), pp. 22-1–22-26.
2. W. J. Peeples *et al.*, *J. Geophys. Res.* **83**, 3459 (1978).
3. V. L. Sharpton, J. W. Head III, *J. Geophys. Res.* **87**, 10983 (1982).
4. G. Picardi *et al.*, *Science* **310**, 1925 (2005); published online 29 November 2005 (10.1126/science.1122165).
5. T. R. Watters *et al.*, *Science* **318**, 1125 (2007); published online 31 October 2007 (10.1126/science.1148112).
6. R. Seu *et al.*, *Planet. Space Sci.* **52**, 157 (2004).
7. R. Seu *et al.*, *Science* **317**, 1715 (2007).
8. T. Ono, H. Oya, *Earth Planets Space* **52**, 629 (2000).
9. T. Ono *et al.*, *Earth Planets Space* **60**, 321 (2008).
10. A-scope data are the data set of the observed signal amplitude versus range. B-scan data are the time series of A-scope data. B-scans are also referred to as radargrams.

11. Using three independent observed parameters (i.e., the surface echo amplitude, the subsurface echo amplitude, and the delay time of the subsurface echo relative to the surface echo), we need to determine four independent parameters (i.e., dielectric constants of two subsurface layers, the loss tangent of the top layer, and the depth of the subsurface reflector). To solve this ill-posed problem, we inversely solve a Fresnel model of three media (i.e., vacuum and two subsurface layers) assuming normal incidence and a moderate value of 4 for the dielectric constant of the top layer.
12. B. A. Campbell, *Radar Remote Sensing of Planetary Surfaces* (Cambridge Univ. Press, Cambridge, 2002).
13. J. W. Head III, *Moon Planets* **21**, 439 (1979).
14. H. Hiesinger *et al*., *J. Geophys. Res.* **105**, 29239 (2000).
15. W. L. Quaide, V. R. Oberbeck, *Moon* **13**, 27 (1975).
16. C. M. Weitz, J. W. Head, C. M. Pieters, *J. Geophys. Res.* **103**, 22725 (1998).
17. H. J. Moore, C. A. Hodges, D. H. Scott, *Lunar Sci. Conf.* **5**, 71 (1974).
18. R. J. Pike, in *Impact and Explosion Cratering*, D. J. Roddy, R. O. Pepin, R. B. Merrill, Eds. (Pergamon, New York, 1977), pp. 489–509.
19. S. Kodama, Y. Yamaguchi, *Meteorit. Planet. Sci.* **38**, 1461 (2003).
20. J. F. Bell III, B. R. Hawke, *Icarus* **118**, 51 (1995).
21. M. I. Staid, C. M. Pieters, *Icarus* **145**, 122 (2000).
22. S. C. Solomon, J. W. Head III, *J. Geophys. Res.* **84**, 1667 (1979).
23. R. A. De Hon, J. D. Waskom, *Lunar Planet. Sci. Conf.* **7**, 2729 (1976).
24. C. J. Budney, P. G. Lucey, *J. Geophys. Res.* **103**, 16855 (1998).
25. T. A. Maxwell, F. El-Baz, S. H. Ward, *Geol. Soc. Am. Bull.* **86**, 1273 (1975).
26. D. F. Weill, R. A. Grieve, I. S. McCallum, Y. Bottinga, *Lunar Sci. Conf.* **2**, 413 (1971).
27. A. M. Freed, H. J. Melosh, S. C. Solomon, *J. Geophys. Res.* **106**, 20603 (2001).

28. The Kaguya mission was conducted by the Japan Aerospace Exploration Agency (JAXA). We thank Y. Takizawa (project manager), S. Sasaki (project scientist), and M. Kato (science manager) for their efforts in accomplishing the Kaguya mission, and K. Tanaka, Y. Iijima, S. Nakazawa, H. Otake, S. Sobue, H. Hoshino, H. Okumura, Y. Yamamoto, and J. Kimura (team members of the Kaguya mission) for their helpful discussions on achieving a high reliability for the mission instruments.

Supporting Online Material

www.sciencemag.org/cgi/content/full/323/5916/909/DC1
Figs. S1 to S4
References and Notes

16 September 2008; accepted 21 January 2009
10.1126/science.1165988

Equilibrium Iron Isotope Fractionation at Core-Mantle Boundary Conditions

Veniamin B. Polyakov

The equilibrium iron isotope fractionation between lower mantle minerals and metallic iron at core-mantle boundary conditions can be evaluated from the high-pressure ^{57}Fe partial vibrational density of states determined by synchrotron inelastic nuclear resonant x-ray scattering spectroscopy using a diamond anvil. Ferropericlase [(Mg,Fe)O] and (Fe,Mg)SiO_3—post-perovskite are enriched in heavy iron isotopes relative to metallic iron at ultrahigh pressures, as opposed to the equilibrium iron isotope fractionation between these compounds at low pressure. The enrichment of Earth and Moon basalts in heavy iron isotopes relative to those from Mars and asteroid Vesta can be explained by the equilibrium iron isotope fractionation during the segregation of Earth's core and the assumption that Earth was already differentiated before the Moon-forming "giant impact."

Iron isotopes may provide important information for understanding the mechanisms of planetary differentiation and core formation. Basalts from Earth and the Moon are enriched in heavy Fe isotopes relative to those from Mars and asteroid Vesta and from chondrites, which are primitive undifferentiated meteorites (*1*–*3*) [supporting online material (SOM) text and fig. S1]. In addition, possible differences in the Fe isotope compositions between lunar and terrestrial basalts (*1*, *2*, *4*–*6*) (SOM text and fig. S2) are currently under debate. The mechanism for the heavy Fe isotope compositions of Earth and Moon basalts is unclear (*4*–*6*). The heavier Fe in Earth and Moon basalts has been suggested to result from evaporation and condensation during the "giant impact" that is thought to have created the Moon (*1*, *4*, *7*) or from Fe isotope fractionation caused by partial melting and magmatic differentiation (*5*, *8*). Core-mantle differentiation has not been considered to explain the enrichment of the bulk silicate Earth (BSE) in heavy Fe isotopes, because the Fe isotope composition of Fe metal in pallasites (differentiated stony-Fe meteorites) is heavy relative to that of the silicate fraction (olivine) (*2*, *7*, *9*) (fig. S3). This implies that the silicate-metallic Fe differentiation depleted the silicate fraction in heavy Fe isotopes and cannot be responsible for the heavier Fe isotope composition of terrestrial and lunar basalts. The pallasite data agree with equilibrium Fe isotope fractionation factors (SOM text) between Fe-metal and ferrous silicates at low pressures (*10*–*12*). Georg *et al*. (*13*) calculated small equilibrium Fe isotope fractionation between Earth's core and mantle [0.02 per mil (‰) at 2000 K] from low-pressure β factors (SOM text) derived from Mössbauer spectroscopy data (*10*). Here I present a method for determining the equilibrium Fe isotope fractionation at higher pressures (up to ~150 GPa) and use this method to estimate equilibrium Fe isotope fractionation between lower mantle minerals and Fe-metal at core-mantle boundary (CMB) conditions and show that core separation on Earth can explain the observed differences in the Fe isotope data from the Earth, Moon, and meteorites.

I determined the equilibrium Fe isotope fractionation between Fe-bearing minerals using their

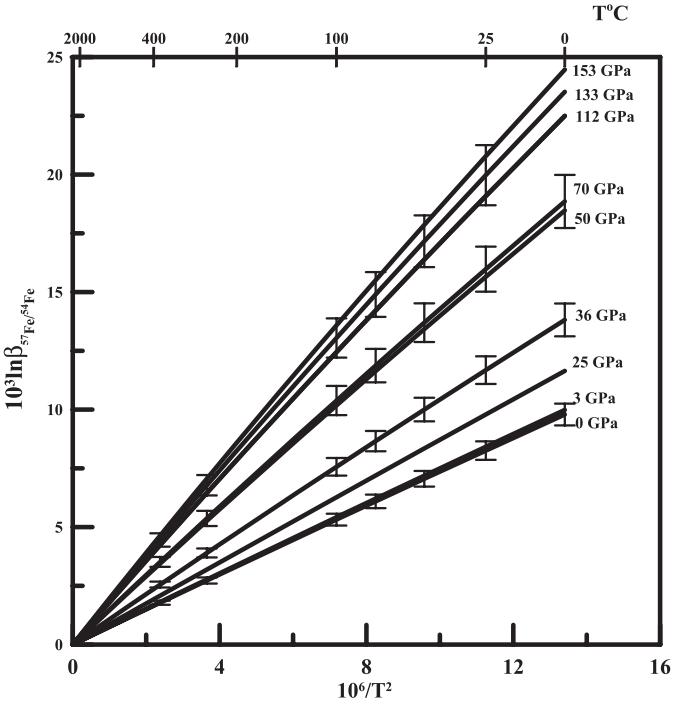


Fig. 1. Temperature and pressure dependence of the ^{57}Fe/^{54}Fe β factor for metallic Fe. The ^{57}Fe/^{54}Fe β factor of Fe metal is computed from the ^{57}Fe PVDOS obtained by Mao *et al*. (*14*) using the high-pressure synchrotron INRXS. The mathematical algorithm is described in the text and justified in (*11*, *17*). Error bars for lnβ are calculated from the experimental uncertainties in the ^{57}Fe PVDOS (*14*) for ambient pressures and at 36, 70, and 133 GPa using the Monte-Carlo technique (*17*) (SOM text). The significant increase of the Fe β factor with increasing pressure to ultrahigh pressures is not a unique feature of Fe metal (see text and Figs. 2 and 3)

Institute of Experimental Mineralogy, Russian Academy of Sciences, Chernogolovka, Moscow Region, Russia. E-mail: polyakov@iem.ac.ru

^{57}Fe partial vibrational densities of states (PVDOSs) derived from high-pressure synchrotron inelastic nuclear resonance x-ray scattering (INRXS) spectroscopy (*14*–*16*) and a calculation method allowing computation of the β factor from the PVDOS (*11*, *17*).

The synchrotron INRXS spectroscopy, in conjunction with a diamond anvil, allows an estimate of the ^{57}Fe PVDOS at different pressures (*14*–*16*), which are input data for computing Fe β factors. The kinetic energy of the thermal vibrations of the ^{57}Fe isotope can be calculated from the ^{57}Fe PVDOS with the following equation

$$K_{^{57}\mathrm{Fe}}(T) = \int g(e)E(e,T)de \qquad (1)$$

where $K_{^{57}\mathrm{Fe}}$ is the kinetic energy of the ^{57}Fe isotope per nucleus, $g(e)$ is the ^{57}Fe PVDOS at given pressure, and $E(e,T)$ is the Einstein function for energy of the single harmonic oscillator at temperature T

$$E(e) \equiv \frac{e}{\exp(e/kT) - 1} + 0.5e \qquad (2)$$

where k is the Boltzmann constant.

Starting from the ^{57}Fe kinetic energy, one can calculate the Fe β factor with the following equation (*18*)

$$\ln\beta = -\left(\frac{K_{^{57}\mathrm{Fe}}}{kT} - \frac{3}{2}\right)\frac{\Delta m}{m} \qquad (3)$$

where m is the mass of an iron isotope (the mass of the ^{54}Fe for ^{57}Fe/^{54}Fe isotope substitution); $\Delta m = m - m_{^{57}\mathrm{Fe}}$ (*19*). This technique does not require any additional assumptions except the harmonic approximation used in Eq. 1 for calculating the kinetic energy of the ^{57}Fe isotope (*10*, *11*). Anharmonic effects may be substantial at high temperatures, which decrease the lnβ (*20*). One can show that high pressure reduces the effect of anharmonicity: for instance, the anharmonic correction to iron lnβ at 130 GPa is about 2.5 times smaller than that at ambient pressure (SOM text). The method for determination of the Fe β factor at high pressure can be directly applied to the problem of equilibrium Fe isotope fractionation at CMB conditions.

Two main Fe-bearing minerals are thought to be present in the lower mantle at CMB pressure: ferropericlase [(Mg,Fe)O] and (Fe,Mg)SiO_3–post-perovskite. Because the equation relating the equilibrium isotopic shift (Δ) between phases A and B and appropriate isotope fractionation factors (α) and β factors (SOM text) is

$$\Delta_{A-B}(‰) \approx 10^3 \ln\alpha_{A-B} = 10^3 \ln\beta_A - 10^3 \ln\beta_B \qquad (4)$$

three Fe β factors for metallic Fe, ferropericlase, and (Fe,Mg)SiO_3–post-perovskite at CMB conditions—are needed for evaluating the equilibrium Fe isotope fractionation factors between the lower mantle minerals and the core.

For metallic Fe, ^{57}Fe/^{54}Fe lnβ determined from the ^{57}Fe PVDOS (*14*) at pressures corresponding to CMB conditions (>100 GPa) exceed those at ambient pressures by a factor of 3 (Fig. 1). ^{57}Fe/^{54}Fe β factors for ferropericlase ($Fe_{0.25}Mg_{0.75}O$) calculated from the ^{57}Fe PVDOS obtained from synchrotron INRXS experiments at different pressures (*16*) reveal an analogous pressure dependence (Fig. 2). ^{57}Fe PVDOS data are also available for ($Fe_{0.4}Mg_{0.6}$)SiO_3–post-perovskite at 130 GPa (*15*), and the calculated ^{57}Fe/^{54}Fe lnβ is higher than those of many other minerals at high pressures (Fig. 3).

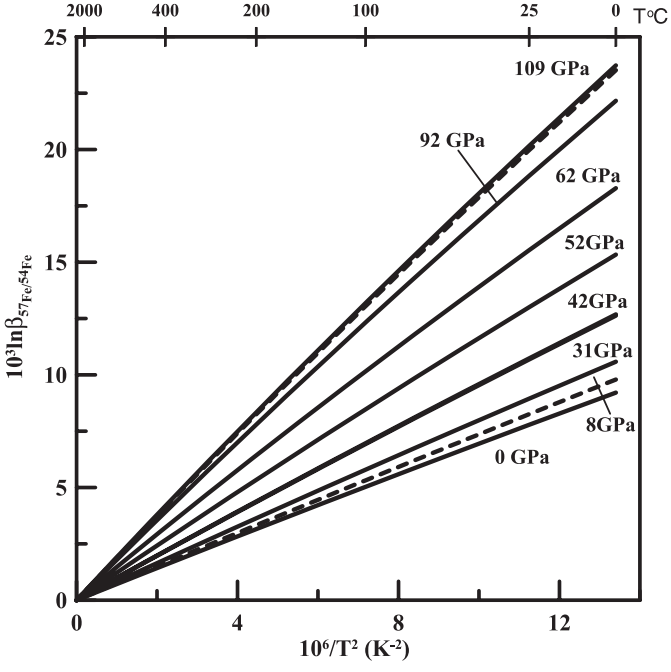


Fig. 2. Temperature and pressure dependence of the ^{57}Fe/^{54}Fe β factor for ($Fe_{0.25}Mg_{0.75}$)O-ferropericlase. The ^{57}Fe/^{54}Fe β factors for ($Fe_{0.25}Mg_{0.75}$)O-ferropericlase are computed using the ^{57}Fe PVDOS at different pressures from Lin *et al.* (*16*). The ^{57}Fe/^{54}Fe β factors for metallic Fe (dashed lines) are also presented at CMB and ambient pressures for comparisons. The Fe β factor for ferropericlase increases significantly with increasing pressure, similar to that found for the ^{57}Fe/^{54}Fe β factors for metallic Fe.

Equilibrium ^{57}Fe/^{54}Fe isotope fractionation factors between the lower mantle Fe-bearing minerals and metallic Fe have opposite signs at ambient and CMB pressures (Fig. 4). The equilibrium Fe isotope fractionation at CMB pressures between ferropericlase and Fe varies from 0.012 ± 0.029‰ at 2000 K to 0.0010 ± 0.0053‰ at 4000 K (which is very small), whereas that between ($Fe_{0.4}Mg_{0.6}$)SiO_3–post-perovskite and Fe is significant and varies from 0.24 ± 0.043‰ at 2000 K to 0.062 ± 0.012‰ at 4000 K. Following (*13*) and assuming that the total amount of Fe is equally distributed between post-perovskite and periclase phases at the CMB, the calculated equilibrium ^{57}Fe/^{54}Fe isotopic shift between the silicate mantle and the core of Earth is 0.13 ± 0.053‰ at 2000 K and 0.04 ± 0.014‰ at 4000 K. This estimate of the equilibrium ^{57}Fe/^{54}Fe isotope fractionation at the CMB, which significantly exceeds the previous estimate of ~0.02‰ at 2000 K (*13*), is in good agreement with the enrichment of terrestrial basalts in heavy Fe isotopes of ~0.1‰ for ^{57}Fe/^{54}Fe (*1*, *2*).

The equilibrium Fe isotope fractionation factor between silicate and Fe metal is negative at low pressures (Fig. 4) (*10*–*12*) but changes its sign and becomes positive at ultrahigh pressures of ~100 GPa (Fig. 4). The sign of the equilibrium Fe isotope fractionation factor between metallic Fe and ferrous silicates may thus be considered as the indicator of ultrahigh pressure in silicate-metal differentiation processes.

The estimated Fe isotope fractionation at CMB conditions implies that the enrichment of terrestrial basalts in heavy Fe isotopes relative to those from Mars, Vesta, and chondrites was caused by core-mantle differentiation in Earth occurring mainly at high pressure. Core formation in Mars and Vesta at low-pressure conditions would have resulted in small isotope enrichments of silicates in light Fe isotopes. The BSE is also enriched in the heavy isotopes of Si. These data can be explained by the preferable accumulation of the light Si isotopes in Earth's core (*13*). Thus, analogous isotope fractionation of Fe and Si might have occurred during the core segregation. In contrast, no measurable differences among planets and meteorites in isotope compositions were detected for light elements such as Mg and Li (*21*, *22*), which are not likely to be present in Earth's core. The absence of such differences for Mg and Li isotopes, which have volatilities higher than (Li) or comparable to (Mg) that of Fe, also renders evaporation unlikely as an explanation of the interplanetary differences in Fe isotope composition (*13*).

One can explain the enrichment of lunar basalts in heavy Fe isotopes, similar in magnitude to those of terrestrial basalts, in terms of the equilibrium Fe isotope fractionation during Earth's core-mantle differentiation, if Earth was already differentiated before the Moon-forming giant impact. However, the putative enrichment of lunar basalts in heavy Fe isotopes relative to those from the Earth (*1*, *4*) (SOM text and fig. S2) requires other mechanisms.

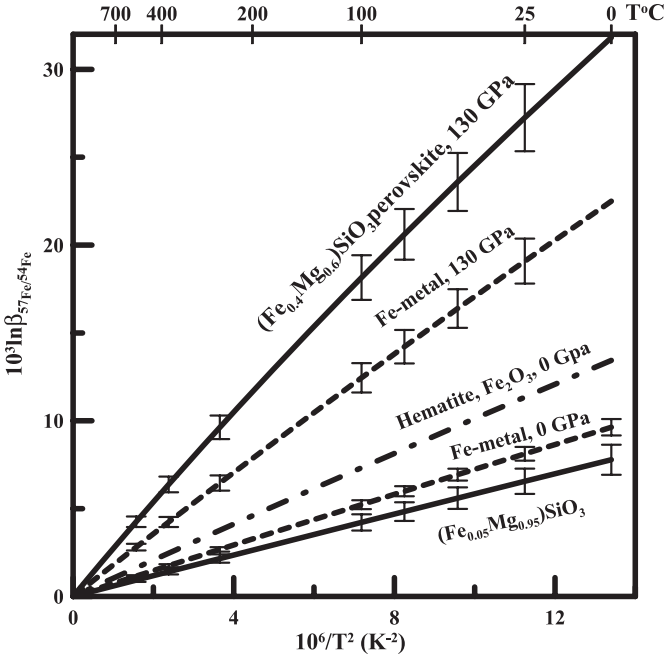


Fig. 3. The Fe β factor for the post-perovskite and perovskite phases at CMB and ambient pressures, respectively. The Fe β factor for $(Fe_{0.4}Mg_{0.6})SiO_3$—post-perovskite at 130 GPa was calculated from the ^{57}Fe PVDOS obtained by the synchrotron INRXS (*15*). Error bars are calculated from experimental uncertainties using the Monte-Carlo technique (*17*) (SOM text). The β factor for $(Fe_{0.05}Mg_{0.95})SiO_3$-perovskite was calculated from the Mössbauer second-order Doppler shift (*27*) by a method described elsewhere (*10*, *18*). The β factors for metallic Fe (Fig. 1) and hematite (*10*, *11*) are also shown for comparisons. The β factor for hematite at ambient pressure from (*10*) was used as an analog for the perovskite phase by Georg *et al*. (*13*).

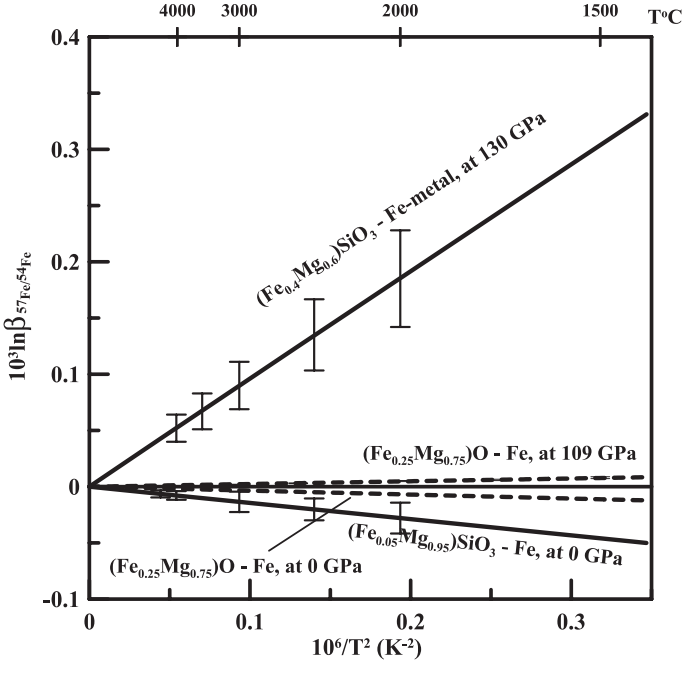


Fig. 4. Equilibrium Fe isotope fractionations between lower mantle minerals and metallic Fe at CMB and ambient pressures. The equilibrium Fe isotope fractionations are calculated from β factors presented in Figs. 1 to 3 using Eq. 4. One can see the different directions of the equilibrium Fe isotope fractionation between ferrous (Fe^{2+}) lower mantle minerals and metallic Fe (Fe^0) at ambient and CMB pressures.

The Fe isotope fractionation between core and mantle estimated in this study reflects equilibrium Fe isotope fractionation between lower-mantle Fe^{2+}-bearing minerals and Fe^0-metal at the CMB, not an isotope fractionation between Fe^{3+} and Fe^0 (*23*). If only 2 to 3% of the lower mantle Fe is ferric (*24*), one can conclude, from isotope mass balance, that $^{57}Fe/^{54}Fe$ isotope fractionation between ferric and ferrous Fe would have to range from 3 to 5‰ to provide the isotope shift of as much as 0.1‰ in the total Fe. Such a large $^{57}Fe/^{54}Fe$ isotope fractionation seems to be unrealistic at temperatures in the lower mantle.

Teng *et al*. (*25*) provided evidence that Fe isotopes fractionate during low-pressure magmatic differentiation, accompanied by a change of the Fe oxidation state. This process is considered as a possible explanation of the differences in Fe isotope composition of basalts from different planetary bodies (*26*). However, it would require that the high-temperature magmatic differentiation processes leading to basalt formation on Earth and the Moon be different from those on Mars and Vesta. It is more likely that the magmatic differentiation is responsible for the range of variations of Fe isotope compositions within a single planetary body and that the primary interplanetary differences reflect the influence of pressure on core formation.

The method developed in the present study for determining Fe β factors at ultrahigh pressures can be directly applied to any chemical elements that have at least one Mössbauer-sensitive isotope, if the synchrotron INRXS technique is available for them.

References and Notes

1. F. Poitrasson, A. N. Halliday, D. C. Lee, S. Levasseur, N. Teutsch, *Earth Planet. Sci. Lett.* **223**, 253 (2004).
2. S. Weyer *et al*., *Earth Planet. Sci. Lett.* **240**, 251 (2005).
3. R. Schoenberg, F. von Blanckenburg, *Earth Planet. Sci. Lett.* **252**, 342 (2006).
4. F. Poitrasson, *Earth Planet. Sci. Lett.* **256**, 484 (2007).
5. S. Weyer *et al*., *Earth Planet. Sci. Lett.* **256**, 638 (2007).
6. B. L. Beard, C. M. Johnson, *Earth Planet. Sci. Lett.* **256**, 633 (2007).
7. F. Poitrasson, S. Levasseur, N. Teutsch, *Earth Planet. Sci. Lett.* **234**, 151 (2005).
8. S. Weyer, D. A. Ionov, *Earth Planet. Sci. Lett.* **259**, 119 (2007).
9. X. K. Zhu *et al*., *Earth Planet. Sci. Lett.* **200**, 47 (2002).
10. V. B. Polyakov, S. D. Mineev, *Geochim. Cosmochim. Acta* **64**, 849 (2000).
11. V. B. Polyakov, R. N. Clayton, J. Horita, S. D. Mineev, *Geochim. Cosmochim. Acta* **71**, 3833 (2007).
12. M. Roskosz, B. Luais, H. C. Watson, M. J. Toplis, C. M. O'D. Alexander, B. O. Mysen, *Earth Planet. Sci. Lett.* **248**, 851 (2006).
13. R. B. Georg, A. N. Halliday, E. A. Schauble, B. C. Reynolds, *Nature* **447**, 1102 (2007).
14. H. K. Mao *et al*., *Science* **292**, 914 (2001).
15. W. L. Mao *et al*., *Science* **312**, 564 (2006).
16. J.-F. Lin *et al*., *Geophys. Res. Lett.* **33**, L22304 (2006).
17. V. B. Polyakov, S. D. Mineev, R. N. Clayton, G. Hu, K. S. Mineev, *Geochim. Cosmochim. Acta* **69**, 5531 (2005).
18. V. B. Polyakov, *Geochim. Cosmochim. Acta* **61**, 4213 (1997).
19. Equation 3 allows calculation of the β factor for Fe isotope fractionation between a given Fe isotope and the ^{57}Fe isotope. For example, $m = m_{^{54}Fe} \approx 54$ atomic units (au) and $\Delta m = m_{^{54}Fe} - m_{^{57}Fe} \approx -3$ au in the case of $^{57}Fe \rightarrow {}^{54}Fe$ isotope substitution. However, in fact this does not impose a limitation on the application of Eq. 2 to Fe isotope fractionation, because the following equation holds: $\ln^{m_2/m_2}\beta = \ln^{m_1/57}\beta - \ln^{m_2/57}\beta$. Here, the $^{m_1/m_2}\beta$ notation is used for the β factor relating to $^{m_2}Fe \rightarrow {}^{m_1}Fe$ isotope substitution. In particular, $\ln^{57/54}\beta = -\ln^{54/57}\beta$; $\ln^{56/54}\beta = \ln^{57/54}\beta - \ln^{57/56}\beta$. See (*18*) for details.
20. V. B. Polyakov, *Geochim. Cosmochim. Acta* **62**, 3077 (1998).
21. U. Wiechert, A. N. Halliday, *Earth Planet. Sci. Lett.* **256**, 360 (2007).
22. T. Magna, U. Wiechert, A. N. Halliday, *Earth Planet. Sci. Lett.* **243**, 336 (2006).
23. H. Williams *et al*., *Earth Planet. Sci. Lett.* **250**, 486 (2006).
24. D. J. Frost, C. A. McCammon, *Annu. Rev. Earth Planet. Sci.* **36**, 389 (2008).
25. F.-Z. Teng, N. Dauphas, R. T. Helz, *Science* **320**, 1620 (2008).
26. S. Weyer, *Science* **320**, 1600 (2008).
27. C. A. McCammon, *Phys. Chem. Miner.* **25**, 292 (1998).

28. Discussions with J. Horita, E. D. Young, R. N. Clayton, A. A. Ariskin, and S. D. Mineev are appreciated. I thank three anonymous reviewers for constructive comments.

Supporting Online Material

www.sciencemag.org/cgi/content/full/323/5916/912/DC1
SOM Text
Figs. S1 to S3
References

23 September 2008; accepted 22 December 2008
10.1126/science.1166329

Skyrmion Lattice in a Chiral Magnet

S. Mühlbauer,[1,2] B. Binz,[3] F. Jonietz,[1] C. Pfleiderer,[1]* A. Rosch,[3]
A. Neubauer,[1] R. Georgii,[1,2] P. Böni[1]

Skyrmions represent topologically stable field configurations with particle-like properties. We used neutron scattering to observe the spontaneous formation of a two-dimensional lattice of skyrmion lines, a type of magnetic vortex, in the chiral itinerant-electron magnet MnSi. The skyrmion lattice stabilizes at the border between paramagnetism and long-range helimagnetic order perpendicular to a small applied magnetic field regardless of the direction of the magnetic field relative to the atomic lattice. Our study experimentally establishes magnetic materials lacking inversion symmetry as an arena for new forms of crystalline order composed of topologically stable spin states.

The formation of most crystals is related to three main aspects. First is the interplay of local repulsion and long-range attraction of atoms leading to an instability of the liquid with correlations that are completely isotropic and without preferred direction. Second, three particle collisions lower the energy and dominate the formation of the crystal out of the isotropic density fluctuations. Third, atoms are quantized with an integer number of them in a unit cell.

For the spin structures in magnetic materials, it was believed that some or all of the three above-mentioned mechanisms could not take place. Consider, for instance, the formation of a magnetically ordered state out of a paramagnetic metal. The underlying atomic lattice or the Fermi surface strongly breaks the rotational and translational invariance. Furthermore, for magnetic fluctuations in the paramagnet, three particle collisions are forbidden by time-reversal symmetry. Lastly, the question arises as to which quantized entities may play the role of the atoms in a solid. Possible candidates are topologically stable objects such as skyrmions, hedgehogs, or merons, which have received great theoretical interest ranging from magnetic monopoles in particle physics (*1*, *2*) to the emergence of gauge theories in condensed matter (*3*). In analogy to the Abrikosov vortex lattice in superconductors, these objects may be considered as the building blocks of novel forms of electronic order akin to crystal structures. However, the experimental evidence for these building blocks is scarce.

We report the observation of the formation of a magnetic structure with hexagonal symmetry perpendicular to a small applied magnetic field in the cubic B20 compound MnSi. We show that this structure can approximately be visualized by a simple superposition of three helical states in the presence of a uniform field, where the superimposed state with the lowest energy can be viewed as a lattice of antiskyrmion lines, that is, magnetic vortices for which the magnetization in the center is antiparallel to the applied field. All three mechanisms mentioned above play an important role in explaining the magnetic structure. The lack of space-inversion symmetry in the atomic crystal of MnSi results in weak spin-orbit coupling that generates slow rotations of all magnetic structures such that they decouple from the underlying atomic lattice very efficiently. Further, in the presence of an external magnetic field, which breaks time-reversal symmetry, three-particle interactions of the magnetic excitations do occur. Lastly, for the magnetic state that emerges under these conditions skyrmion lines, that is, certain topologically protected knots in the magnetic structure, take over the role of the atoms in usual crystals.

At ambient pressure and zero applied magnetic field, MnSi develops helical magnetic order below a transition temperature T_c = 29.5 K that is the result of three hierarchical energy scales. The strongest scale is ferromagnetic exchange favoring a uniform spin polarization (spin alignment). The lack of inversion symmetry of the cubic B20 crystal structure results in chiral spin-orbit interactions, which may be described by the rotationally invariant Dzyaloshinsky Moriya (DM) interaction. The ferromagnetic exchange together with the chiral spin-orbit coupling lead to a rotation of the spins with a periodicity $\lambda_h \approx 190$ Å that is large compared with the lattice constant, $a \approx 4.56$ Å. This large separation of length scales implies an efficient decoupling of the magnetic and atomic structures. Therefore, the alignment of the helical spin spiral along the cubic space diagonal ⟨111⟩ is weak and is only fourth power in the small spin-orbit coupling. These crystalline field interactions, which break the rotational symmetry, are by far the weakest scale in the system.

Our study was partly inspired by recent work on the pressure dependence of the properties of MnSi (*4*–*6*). As shown in Fig. 1A, well below T_c an applied magnetic field, **B**, unpins the helical order and aligns its wave vector **Q** parallel to the field, **Q** ∥ **B**, for a field exceeding $B_{c1} \approx 0.1$ T (*7*–*9*). This state is referred to as the conical phase. For a magnetic field exceeding $B_{c2} \approx 0.6$ T, the effects of the DM interaction are suppressed, giving way to a spin-aligned (ferromagnetic) state. For temperatures just below T_c, an additional phase, referred to as the A phase, is stabilized in a finite field interval shown for **B** ∥ ⟨100⟩ in Fig. 1 (*7*, *8*). It had been believed that the A phase was explained by a single-**Q** helix, where **Q** is perpendicular to the applied field (*8*, *9*). The specific heat exhibits a tiny peak at the border of the A phase, whereas the susceptibility discontinuously assumes a lower value (*10*). When taken together with the discontinuous change of the scattering intensity seen in neutron scattering, this suggests that the A phase represents a distinct magnetic phase, where the transition from the conical order to the A phase is discontinuous (first order).

Because **Q** tends to align parallel to an applied magnetic field, neutron scattering as a function of **B** has been reported for setups in which the magnetic field was perpendicular to the incident neutron beam (*11*). In contrast, we chose the incident neutron beam to be parallel to the applied magnetic field (Fig. 1B). Two samples were studied. Sample 1 refers to a disk of 19-mm diameter and a thickness of 3 mm, where the vector normal to the disc was slightly misaligned by 11° with respect to a ⟨110⟩ axis. Sample 2 is a small parallelepiped, with dimensions 1.5 mm by 1.5 mm by 14 mm, where a ⟨110⟩ axis corresponded to the long axis. To search for higher-order peaks and double scattering, we increased the neutron flux in our measurements of sample 1, accepting a larger beam divergence. In contrast, the beam divergence was reduced for sample 2 to improve the resolution in rocking scans. All data at finite magnetic field were measured after zero-field cooling to the desired temperature, followed by a field ramp to the desired field value. However, the results for the A phase were identical when recorded after field cooling. For further details of the experimental setup, we refer to (*11*).

Figure 2 shows typical data recorded, where the spot sizes represent the resolution limit. All data shown represent sums over rocking scans of typically ±8° (*11*). Figure 2, A to C, shows data for sample 1, whereas Fig. 2, D to F, shows data for sample 2. Figure 2A shows the scattering intensity of sample 1 in a zero-field-cooled state at a temperature of 27 K for a ⟨110⟩ scattering plane, and the ⟨110⟩ axis is indicated in the figure. The pattern is consistent with previously published data and helical magnetic order along ⟨111⟩. Figure 2B shows the intensity pattern of sample 1 in the A phase. Six spots emerge on a regular hexagon.

We tested the variation of the intensity pattern on the orientation of the field relative to the crystal axes in both samples. The field was always parallel to the incident beam, whereas the sample was rotated for a large number of different orientations. Typical data are shown in Fig. 2C for sample 1, where the sample was rotated with respect to the vertical axis into a

[1]Physik Department E21, Technische Universität München, D-85748 Garching, Germany. [2]Forschungsneutronenquelle Heinz Maier-Leibnitz (FRM II) Technische Universität München, D-85748 Garching, Germany. [3]Institute for Theoretical Physics, University of Cologne, Zülpicher Straße 77, D-50937 Cologne, Germany.

*To whom correspondence should be addressed. E-mail: christian.pfleiderer@frm2.tum.de

random position. For both samples and all crystal orientations, the scattering pattern always exhibited the sixfold symmetry. In case the scattering plane contained a ⟨110⟩ direction, two of the peaks of the sixfold pattern coincided with this direction. As for Fig. 2C, the scattering plane did not contain a ⟨110⟩ direction. For sample 2, the intensities along the vertical direction, which coincided with the ⟨110⟩ direction, were systematically weaker. This may be explained by the demagnetizing fields caused by the large aspect ratio, which implies that part of the scattering intensity was not captured in the rocking scans [see also (*11*)]. The main result of our study is that, for all orientations of the magnetic field with respect to the atomic lattice, six Bragg reflections are observed on a regular hexagon that is strictly perpendicular to the magnetic field.

We performed rocking scans to test whether the A phase has long-range order. Typical data are presented in (*11*). In the helical state, the half-width of the rocking scans corresponded to a magnetic mosaicity $\eta_m \approx 3.5°$ consistent with previous work and long-range order (*12*, *13*). Remarkably, in the A phase the half-width of the rocking scans corresponded to a reduced magnetic mosaicity $\eta_m \approx 1.75°$, implying an even longer correlation length of at least $\xi \approx 5500$ Å, when allowing for demagnetizing fields (*11*).

To test for consistency with previous work, we also measured the emergence of the A phase as a function of temperature for magnetic field perpendicular to the neutron beam, where the vertical axis was the same ⟨110⟩ axis as before and the low-symmetry horizontal axis containing spots 6 and 8 in Fig. 2F was perpendicular to the magnetic field and incident neutron beam. Data were recorded after (i) zero-field-cooling the sample to a temperature well below T_c, (ii) increasing the magnetic field to 0.19 T, and (iii) measuring the neutron scattering pattern for selected increasing temperatures (Fig. 2F shows data for $T = 27.7$ K). Well below T_c we first observe the two spots parallel to the field direction labeled 9 and 10, characteristic of the conical state. When entering the A phase, the intensity of the spots of the conical phase becomes very weak but does not vanish, whereas strong scattering intensity appears in the perpendicular direction (spots 6 and 8). This is consistent with previous work and may signal a phase coexistence, as expected of a weak first-order transition with possible extra effects of the demagnetizing fields added.

The key results of our neutron scattering data may be summarized as follows: (i) the helical wave vector aligns perpendicular to the applied magnetic field; (ii) the fundamental symmetry of the intensity pattern is sixfold, suggesting a multi-**Q** structure; and (iii) the A phase stabilizes in a magnetic field strength of order $B_{c2}/2$. Moreover, the pattern aligns very weakly with respect to the ⟨110⟩ orientation. We can readily account for these features in the framework of standard Landau-Ginzburg theory in the mean-field approximation by taking fluctuations into account. Near T_c the Ginzburg-Landau energy functional can be written as (*14*, *15*)

$$F[\mathbf{M}] = \int d^3r [r_0\mathbf{M}^2 + J(\nabla\mathbf{M})^2 + 2D\mathbf{M} \cdot (\nabla \times \mathbf{M}) + U\mathbf{M}^4 - \mathbf{B} \cdot \mathbf{M}] \quad (1)$$

The first and second terms represent the usual quadratic contribution with the conventional gradient term; the third term, the Dzyaloshinsky-Moriya interaction; and the last term, the coupling to an external magnetic field **B**. The quartic term accounts in lowest order for the effects of mode-mode interactions and stabilizes the magnetic order. We neglect higher-order spin-orbit coupling terms describing anisotropy effects (*14*, *15*). The free energy is given by $\exp(-G) = \int D\mathbf{M} \exp(-F[\mathbf{M}])$ (throughout the paper, we use a dimensionless free energy). Within mean-field approximation, $G(\mathbf{B})$ is equal to $\min F[\mathbf{M}]$, and one minimizes F with respect to the spin structure $\mathbf{M}(\mathbf{r})$.

To explain the A phase, we evoke strong analogies with the crystal formation of ordinary solids out of the liquid state. The latter is in most cases driven by the cubic interactions of density waves (*16*), which in momentum space can be written as

$$\sum_{\mathbf{q}_1,\mathbf{q}_2,\mathbf{q}_3} \rho_{\mathbf{q}_1}\rho_{\mathbf{q}_2}\rho_{\mathbf{q}_3}\delta(\mathbf{q}_1 + \mathbf{q}_2 + \mathbf{q}_3)$$

The ordered state can gain energy from this term only when three ordering vectors of the crystal structure add up to zero. Accordingly, in many cases (exceptions can arise only for strong first-order transitions) the ordered phase, which forms first out of a liquid state, is of body-centered cubic symmetry (*16*), which is the crystal structure with the largest number of such triples of reciprocal lattice vectors.

In the presence of a finite uniform component of the magnetization, $\mathbf{M}_f$, a similar mechanism can also occur in MnSi. From the quartic term in Eq. 1, we obtain terms that are effectively cubic in the modulated moment amplitudes

$$\sum_{\mathbf{q}_1,\mathbf{q}_2,\mathbf{q}_3} (\mathbf{M}_f \cdot \mathbf{m}_{\mathbf{q}_1})(\mathbf{m}_{\mathbf{q}_2} \cdot \mathbf{m}_{\mathbf{q}_3})\delta(\mathbf{q}_1 + \mathbf{q}_2 + \mathbf{q}_3) \quad (2)$$

where $\mathbf{m}_\mathbf{q}$ is the Fourier transform of $\mathbf{M}(\mathbf{r})$. As in the case of an ordinary crystal, one can gain energy from this term for a structure with three **Q** vectors adding up to zero. These vectors have a fixed modulus determined by the interplay of the two gradient terms in Eq. 1. Therefore, these three vectors have relative angles of 120° (Fig. 3A) and define a plane characterized by a normal vector, $\hat{n}$. By symmetry, the energy change is proportional to $\mathbf{M}_f \cdot \hat{n}$, and therefore the three **Q** vectors must be perpendicular with respect to the external magnetic field. Our qualitative arguments already explain the two main experimental observations in the A phase: The

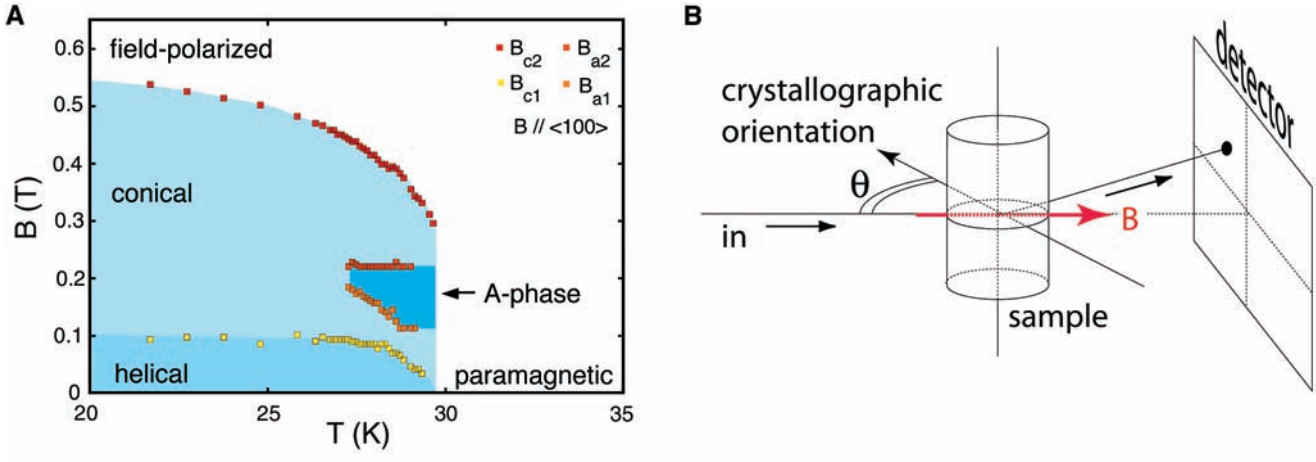


Fig. 1. (**A**) Magnetic phase diagram of MnSi. For $B = 0$, helimagnetic order develops below $T_c = 29.5$ K. Under magnetic field, the helical order unpins and aligns along the field above B_{c1}; above B_{c2}, the helical modulation collapses. In the conical phase, the helix is aligned parallel to the magnetic field. The transition fields shown here have been inferred from the AC susceptibility, where the DC and AC fields were parallel to ⟨100⟩ (*10*). (**B**) Neutron scattering setup used in our study; the applied magnetic field **B** was parallel to the incident neutron beam.

Bragg spots are located in the plane perpendicular to **B** and display a sixfold symmetry (because both **Q** and −**Q** give a Bragg reflection) independent of the orientation of the underlying lattice. We therefore suggest that the A phase is a chiral spin crystal, the A crystal, approximately characterized by the magnetization

$$\mathbf{M}(r) \approx \mathbf{M}_{\mathrm{f}} + \sum_{i=1}^{3} \mathbf{M}^{h}_{\mathbf{Q}_i}(\mathbf{r} + \Delta\mathbf{r}_i) \qquad (3)$$

where $\mathbf{M}^{h}_{\mathbf{Q}_i}(\mathbf{r}) = A[\mathbf{n}_{i1}\cos(\mathbf{Q}_i\mathbf{r}) + \mathbf{n}_{i2}\sin(\mathbf{Q}_i\mathbf{r})]$ is the magnetization of a single chiral helix with amplitude A, wave vector $\mathbf{Q}_i$, and two unit vectors, $\mathbf{n}_{i1}$ and $\mathbf{n}_{i2}$, orthogonal to each other and to $\mathbf{Q}_i$. All three helices have the same chirality; that is, all $\mathbf{Q}_i(\mathbf{n}_{i1} \times \mathbf{n}_{i2})$ have the same sign. More precisely, one has to add further higher-order Fourier components to Eq. 3 when minimizing $F[M]$. However, these terms remain small close to T_c. The relative shifts, $\Delta\mathbf{r}_i$, of the helices, which we calculate theoretically, determine whether the A crystal can be described as a lattice of skyrmions (see below).

One also has to take into account that an external magnetic field favors helices with **Q** vectors parallel to **B**, because this way the spins may easily tilt parallel to the field to form a conical structure. Within mean-field theory, this conical phase always has the lowest energy (*11*). However, in the parameter range, where the A phase occurs experimentally, that is, close to T_c, at intermediate magnetic fields (Fig. 1A), the energy difference between the two phases becomes very small as shown in the Fig. 3B inset. The origin of the energy minimum of the A crystal for moderate magnetic field can be traced back to the size of the modulations of the magnetization amplitude, $|\mathbf{M}(\mathbf{r})|$, which is minimal close to $B \approx 0.4B_{c2}$ (*11*). In our mean-field Landau-Ginzburg theory, the A crystal thus appears as a metastable phase, which becomes extremely close in energy to the conical phase for intermediate fields $B \approx 0.4B_{c2}$.

It turns out that, when we consider thermal fluctuations around the mean-field solution, these stabilize the A crystal. To show this, we consider the leading correction to mean-field theory arising from Gaussian fluctuations

$$\mathbf{G} \approx F[\mathbf{M}_0] + \frac{1}{2}\log\ \det\left(\frac{\delta^2 F}{\delta\mathbf{M}\delta\mathbf{M}}\right)\Bigg|_{\mathbf{M}_0} \qquad (4)$$

where $\mathbf{M}_0$ is the mean-field spin configuration for either the A phase or the conical phase. To make Eq. 4 well defined, one has to specify a cutoff scheme for short length scales. We use a cutoff in momentum space, $k < 2\pi/a$, where a is the lattice spacing of the MnSi crystal. Because of the long pitch of the helix, it turns out that most contributions arise from fluctuations on short length scales with the exception of temperatures extremely close to T_c [see (*11*) for a detailed discussion], but both short-range and long-range fluctuations favor the A crystal for intermediate magnetic fields. As shown in the Fig. 3B inset, the fluctuations indeed stabilize the A crystal. A typical phase diagram resulting from Eq. 4 is shown in Fig. 3B. The theoretical phase diagram catches the main characteristics of the experimental phase diagram. The A crystal is stable at intermediate fields not too far from T_c. When interpreting the theoretical result, one has to take into account that Eq. 4 is only valid for small fluctuations and therefore cannot be applied too close to T_c. Indeed it is expected (*17*, *18*) that fluctuations ultimately drive the transition first order and that such strong fluctuations will substantially shift the transition line to the paramagnet. We estimate the strength of fluctuations by calculating the leading correction to the order parameter for both the conical phase and the A crystal (*11*). In the shaded area of Fig. 3B, these corrections are small (less than 20%), which justifies the use of Eq. 4.

As can be seen from Fig. 3C, the magnetic structure of the A crystal obtained by minimizing $F[\mathbf{M}]$ is characterized by a pattern of magnetic vortices. To elucidate their nature, we compute the skyrmion density given by (*19*):

$$\phi = \frac{1}{4\pi}\mathbf{n} \cdot \frac{\partial\mathbf{n}}{\partial x} \times \frac{\partial\mathbf{n}}{\partial y} \qquad (5)$$

where x and y are the coordinates perpendicular to **B** and $\mathbf{n} = \mathbf{M}(\mathbf{r})/|\mathbf{M}(\mathbf{r})|$ is the orientation of the magnetization. ϕ is a measure of the winding of the magnetization profile. If ϕ integrates to 1 or

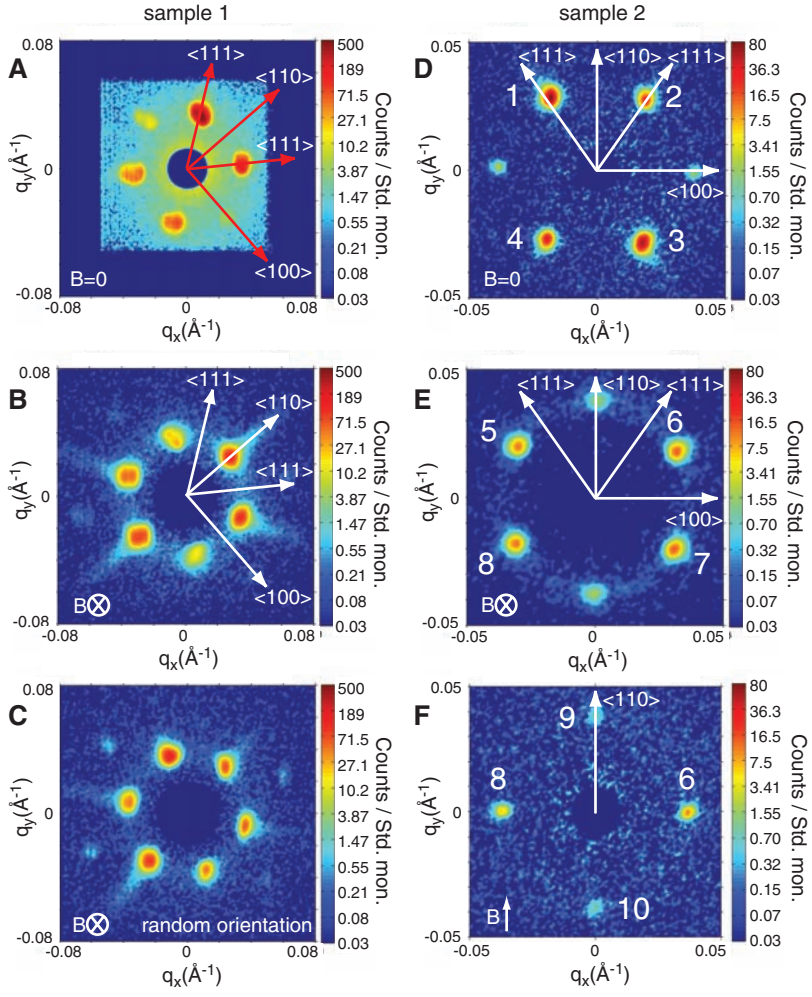


Fig. 2. Typical neutron small angle scattering intensities; note that the color scale is logarithmic to make weak features visible. Data represent the sum over rocking scans with respect to the vertical axis through the sample. (A) to (C) show data for sample 1 and (D) to (F) for sample 2. Backgrounds measured above T_c and $B = 0$ were subtracted in all panels except for (A) (light blue square). Spots are labeled for reference; for the intensity of these spots as a function of rocking angle, see (*11*). (**A**) Helical order in sample 1 in the zero-field-cooled state at $T = 27$ K and $B = 0$. (**B**) Sixfold intensity pattern in the A phase in sample 1; same orientation as in (A); $T = 26.45$ K, $B = 0.164$ T. (**C**) Sixfold intensity pattern in the A phase for random orientation of sample 1 (see text for details); $T = 26.77$ K, $B = 0.164$ T. (**D**) Helical order in sample 2 in the zero-field-cooled state at $T = 16$ K and $B = 0$. (**E**) A phase in sample 2, same orientation as in (D); $T = 27.7$ K, $B = 0.162$ T. (**F**) A phase as measured in conventional setup [compare Fig. 1A in (*11*)], where data in all other panels were measured in the configuration shown in Fig. 1B; $T = 27.7$ K, $B = 0.190$ T. A small residual intensity due to the conical phase is observed (spots 9 and 10), whereas spots 6 and 8 correspond to those in (E).

−1, a topologically stable knot exists in the magnetization. As shown in Fig. 3D, the skyrmion density is finite and oscillates between positive and negative as compared with the normal helical or conical phases, where it is zero. Moreover, the skyrmion number $\Phi = \int \phi(\mathbf{r}) d^2\mathbf{r}$ per two-dimensional unit cell is quantized and adds up to −1. Taken together this implies that the A crystal can be interpreted as a crystal made out of quantized objects, the skyrmion lines, with a magnetization at their core that is antiparallel to the applied magnetic field and $\mathbf{M}_0$.

Because of its symmetry and the cubic interactions, the A crystal has to be separated by two first-order phase transitions both from the conical and the paramagnetic phases. This is consistent with the experimental observations. Two additional features in the scattering patterns that account for less than 1% of the total integrated scattering intensity are, first, weak continuous streaks of intensity emerging radially outward from the six main spots and, second, the coexistence of conical order and the spin crystal. Both may be the result of weak heterogeneities resulting from these generic first-order boundaries of the A crystal, possibly in combination with demagnetizing fields.

Many years ago in a seminal study, Bogdanov and collaborators used a mean-field model to predict the existence of skyrmion lattices for anisotropic noncentrosymmetric magnetic materials under the application of a magnetic field (*20*, *21*). The authors also pointed out that, within their mean-field theory for cubic materials such as MnSi, the skyrmion lattice would always be metastable. Moreover, in the absence of a magnetic field, it has been shown theoretically that certain crystalline spin structures can be stabilized by long-range interactions or an additional phenomenological parameter (*22*–*24*). In contrast to these predictions, we find in this study that it is sufficient to include the effects of Gaussian thermal fluctuations to stabilize skyrmion lattices in a magnetic field in cubic materials.

It is instructive to search for analogies of the A crystal in other condensed matter systems. Because it is a multi-**Q** structure, we note that previously known multi-**Q** structures, for example, in the rare earths, involve large values of **Q** and exhibit very strong pinning to the atomic lattice (*25*, *26*), whereas for MnSi we observe that the sixfold pattern of the A crystal exists independently of the underlying lattice and Q is quite tiny. Although flux lines in superconductors and the magnetic skyrmion lines observed are topologically completely different objects, there is nevertheless an intimate similarity of the Abrikosov lattice of superconducting flux lines and the hexagonal symmetry of the A crystal (*20*, *21*). Moreover, the A crystal is characterized by broken translation symmetry in the plane perpendicular to **B** only. Therefore the A phase is similar to the chiral columnar phase of liquid crystals (*16*, *27*). Further, the spin structure of the A crystal is topologically equivalent to theoretical predictions of the spin structure of the ferromagnetic quantum Hall state near 1/2 filling (*28*), where, however, the underlying energetics is completely different. Lastly, individual magnetic vortices attract also great interest as a micromagnetic phenomenon, which arises when conventional domain walls in soft ferromagnets are made to meet (*29*).

The skyrmion lattice in the chiral magnet MnSi reported here represents an example where an electronic liquid forms a spin crystal made from topologically nontrivial entities. This provides a glimpse of the large variety of magnetic states that may be expected from the particle-like magnetic objects currently discussed in the literature.

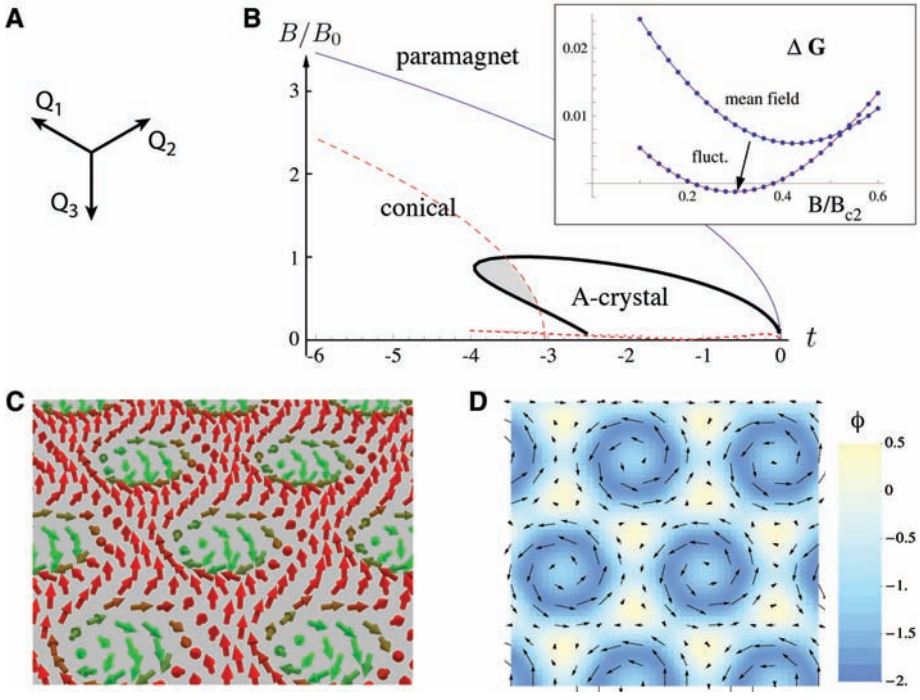


Fig. 3. (**A**) Depiction of the hexagonal basis vectors of the crystalline spin order in the A phase. (**B**) Theoretical phase diagram as a function of magnetic field B/B_0 with $B_0 = \sqrt{(JQ^2)^3}/U$ and the parameter $t = r_0 J/D^2 - 1$, which is roughly proportional to $T - T_c$. We use the model parameter $\gamma = JD/U = 5$ and a momentum-space cutoff $k < 40D/J$. Smaller values of γ increase the A phase regime. For most values of field, the A phase is either metastable or stable, but at low fields below the dotted line it becomes unstable. Above and to the right of the red dashed line, the fluctuation correction to the size of the order parameter becomes larger than 20%, and our theoretical analysis becomes uncontrolled. Therefore in the shaded gray region, we have reliably established stability of the A phase within our model [see (*11*) for details]. (Inset) Energy difference between A phase and conical phase as a function of field for the same parameters and $t = -3.5$, both in the mean-field approximation and with fluctuation corrections. Fluctuations stabilize the A phase at intermediate fields. (**C**) Real space depiction of the spin arrangement in the A phase in the x-y plane. Note that this spin arrangement is translation-invariant along the z axis, which is parallel to the magnetic field. (**D**) Skyrmion density per unit cell area as calculated for the A phase as shown in (C). The integrated skyrmion density per unit cell is finite, $\Phi = -1$. The arrows represent the magnetization component perpendicular to the line of sight.

References and Notes

1. G. t'Hooft, *Nucl. Phys. B* **79**, 276 (1974).
2. A. M. Polyakov, *JETP Lett.* **22**, 194 (1974).
3. M. Levin, T. Senthil, *Phys. Rev. B* **70**, 220403 (2004).
4. C. Pfleiderer, S. R. Julian, G. G. Lonzarich, *Nature* **414**, 427 (2001).
5. C. Pfleiderer *et al.*, *Nature* **427**, 227 (2004).
6. C. Pfleiderer, P. Böni, T. Keller, U. K. Rößler, A. Rosch, *Science* **316**, 1871 (2007).
7. Y. Ishikawa, M. Arai, *J. Phys. Soc. Jpn.* **53**, 2726 (1984).
8. B. Lebech, *Recent Advances in Magnetism of Transition Metal Compounds* (World Scientific, Singapore, 1993), p. 167.
9. S. V. Grigoriev, S. V. Maleyev, A. I. Okorokov, Y. O. Chetverikov, H. Eckerlebe, *Phys. Rev. B* **73**, 224440 (2006).
10. C. Thessieu, C. Pfleiderer, A. N. Stepanov, J. Flouquet, *J. Phys. Cond. Matter* **9**, 6677 (1997).
11. Materials and methods are available as supporting material on *Science* Online.
12. B. Lebech *et al.*, *J. Magn. Magn. Mater.* **140–144**, 119 (1995).
13. C. Pfleiderer, D. Reznik, L. Pintschovius, J. Haug, *Phys. Rev. Lett.* **99**, 156406 (2007).
14. O. Nakanishi, A. Yanase, A. Hasegawa, M. Kataoka, *Solid State Commun.* **35**, 995 (1980).
15. P. Bak, M. H. Jensen, *J. Phys. C* **13**, L881 (1980).
16. P. M. Chaikin, T. C. Lubensky, *Principles of Condensed Matter Physics* (Cambridge Univ. Press, Cambridge, 1995).
17. S. A. Brazovskii, *Sov. Phys. JETP* **41**, 85 (1975).
18. J. Schmalian, M. Turlakov, *Phys. Rev. Lett.* **93**, 036405 (2004).
19. B. Binz, A. Vishwanath, *Physica B* **403**, 1336 (2008).
20. A. Bogdanov, A. Hubert, *J. Magn. Magn. Mater.* **138**, 255 (1994).
21. A. N. Bogdanov, D. A. Yablonskii, *Sov. Phys. JETP* **68**, 101 (1989).
22. B. Binz, A. Vishwanath, V. Aji, *Phys. Rev. Lett.* **96**, 207202 (2006).
23. U. K. Rößler, A. N. Bogdanov, C. Pfleiderer, *Nature* **442**, 797 (2006).

24. I. Fischer, N. Shah, A. Rosch, *Phys. Rev. B* **77**, 024415 (2008).
25. E. M. Forgan, E. P. Gibbons, K. A. McEwen, D. Fort, *Phys. Rev. Lett.* **62**, 470 (1989).
26. D. Gignoux, D. Schmitt, *J. Magn. Magn. Mater.* **100**, 99 (1991).
27. E. Grelet, *Phys. Rev. Lett.* **100**, 168301 (2008).
28. L. Brey, H. A. Fertig, R. Cote, A. H. MacDonald, *Phys. Rev. Lett.* **75**, 2562 (1995).
29. S. D. Bader, *Rev. Mod. Phys.* **78**, 1 (2006).

30. We thank A. N. Bogdanov, S. Dunsiger, E. M. Forgan, C. Franz, M. Garst, M. Janoschek, H. Kolb, M. Laver, S. Legl, T. Lorenz, W. Münzer, T. Nattermann, J. Peters, U. K. Rößler, B. Russ, R. Schwikowski, A. Vishwanath, M. Vojta, W. Zwerger, and the team of FRM II for discussions and support and K. von Bergmann for stimulating discussions and help in preparing Fig. 3C. We gratefully acknowledge financial support by SFB608 and the Alexander-von-Humboldt foundation.

Supporting Online Material
www.sciencemag.org/cgi/content/full/323/5916/915/DC1
Materials and Methods
Figs. S1 to S6
References

3 October 2008; accepted 9 December 2008
10.1126/science.1166767

Observation of Unconventional Quantum Spin Textures in Topological Insulators

D. Hsieh,[1] Y. Xia,[1,2] L. Wray,[1,3] D. Qian,[1] A. Pal,[1] J. H. Dil,[4,5] J. Osterwalder,[5] F. Meier,[4,5] G. Bihlmayer,[6] C. L. Kane,[7] Y. S. Hor,[8] R. J. Cava,[8] M. Z. Hasan[1,2]*

A topologically ordered material is characterized by a rare quantum organization of electrons that evades the conventional spontaneously broken symmetry–based classification of condensed matter. Exotic spin-transport phenomena, such as the dissipationless quantum spin Hall effect, have been speculated to originate from a topological order whose identification requires a spin-sensitive measurement, which does not exist to this date in any system. Using Mott polarimetry, we probed the spin degrees of freedom and demonstrated that topological quantum numbers are completely determined from spin texture–imaging measurements. Applying this method to Sb and $Bi_{1-x}Sb_x$, we identified the origin of its topological order and unusual chiral properties. These results taken together constitute the first observation of surface electrons collectively carrying a topological quantum Berry's phase and definite spin chirality, which are the key electronic properties component for realizing topological quantum computing bits with intrinsic spin Hall–like topological phenomena.

Ordered phases of matter such as a superfluid or a ferromagnet are usually associated with the breaking of a symmetry and are characterized by a local order parameter (*1*), and the typical experimental probes of these systems are sensitive to order parameters. The discovery of the quantum Hall effects in the 1980s revealed a new and rare type of order that is derived from an organized collective quantum motion of electrons (*2*–*4*). These so-called "topologically ordered phases" do not exhibit any symmetry-breaking and are characterized by a topological number (*5*) as opposed to a local order parameter. The classic experimental probe of topological quantum numbers is magneto-transport, in which measurements of the quantization of Hall conductivity $\sigma_{xy} = ne^2/h$ (where e is the electric charge and h is Planck's constant) reveals the value of the topological number n that characterizes the quantum Hall effect state (*6*).

Recent theoretical and experimental studies suggest that a new class of quantum Hall–like topological phases can exist in spin-orbit materials without external magnetic fields, with interest centering on two examples: the "quantum spin Hall insulator" (*7*–*9*) and the "strong topological insulator" (*10*, *11*). Their topological order is believed to give rise to unconventional spin physics at the sample edges or surfaces, with potential applications ranging from dissipationless spin currents (*12*) to topological (fault-tolerant) quantum computing (*13*). However, unlike conventional quantum Hall systems, these previously unmeasured topological phases do not necessarily exhibit a quantized charge or spin response ($\sigma_{xy} \neq ne^2/h$) (*14*, *15*). In fact, the spin polarization is not a conserved quantity in a spin-orbit material. Thus, their topological quantum numbers, the analogs of n, cannot be measured via the classic von Klitzing–type (*2*) transport methods.

Here we show that spin-resolved angle-resolved photoemission spectroscopy (spin-ARPES) can perform analogous measurements for topological insulators and quantum spin Hall materials. We measured all of the topological numbers for $Bi_{1-x}Sb_x$ and provide an identification of its spin texture, which heretofore was unmeasured despite its surface states (SSs) having been observed (*10*). The measured spin texture reveals the existence of a nonzero geometrical quantum phase [Berry's phase (*16*, *17*)] and the handedness or chiral properties. More importantly, this technique enables us to investigate aspects of the metallic regime of the $Bi_{1-x}Sb_x$ series, such as spin properties in pure Sb, which are necessary to determine the microscopic origin of topological order. Our measurements on pure metallic Sb show that its surface carries a π geometrical (Berry's) phase and chirality property unlike the conventional spin-orbit metals such as gold (Au), which has zero net Berry's phase and no topological chirality (*18*).

Strong topological materials are distinguished from ordinary materials such as Au by a topological quantum number $\nu_0 = 1$ or 0, respectively (*14*, *15*). For $Bi_{1-x}Sb_x$, theory has shown that ν_0 is determined solely by the character of the bulk electronic wave functions at the L point in the three-dimensional (3D) Brillouin zone (BZ). When the lowest energy-conduction band state consists of an antisymmetric combination of atomic p-type orbitals (L_a) and the highest energy valence band state consists of a symmetric combination (L_s), then $\nu_0 = 1$, and vice versa for $\nu_0 = 0$ (*11*). Although the bonding nature (parity) of the states at L is not revealed in a measurement of the bulk band structure, the value of ν_0 can be determined from the spin textures of the surface bands that form when the bulk is terminated. In particular, a $\nu_0 = 1$ topology requires the terminated surface to have a Fermi surface (FS) (*11*) that supports a nonzero Berry's phase (an odd as opposed to an even multiple of π), which is not realizable in an ordinary spin-orbit material.

In a general inversion symmetric spin-orbit insulator, the bulk states are spin-degenerate because of a combination of space-inversion symmetry $[E(\vec{k},\uparrow) = E(-\vec{k},\uparrow)]$ and time-reversal symmetry $[E(\vec{k},\uparrow) = E(-\vec{k},\downarrow)]$, where E is the energy and $\vec{k}$ is the electron momentum. Because space-inversion symmetry is broken at the terminated surface, the spin degeneracy of surface bands can be lifted by the spin-orbit interaction (*19*–*21*). However, according to Kramers theorem (*16*), they must remain spin-degenerate at four special time-reversal invariant momenta ($\vec{k}_T = \overline{\Gamma}, \overline{M}$) in the surface BZ (*11*), which for the (111) surface of $Bi_{1-x}Sb_x$ are located at $\overline{\Gamma}$ and three equivalent $\overline{M}$ points (Fig. 1A).

Depending on whether ν_0 equals 0 or 1, the FS pockets formed by the surface bands will enclose the four $\vec{k}_T$ an even or odd number of times, respectively. If a FS pocket does not enclose $\vec{k}_T$ (= $\overline{\Gamma}$, $\overline{M}$), it is irrelevant for the topology (*11*, *20*). Because the wave function of a single electron spin acquires a geometric phase

[1]Joseph Henry Laboratories of Physics, Department of Physics, Princeton University, Princeton, NJ 08544, USA. [2]Princeton Center for Complex Materials, Princeton Institute for the Science and Technology of Materials, Princeton University, Princeton, NJ 08544, USA. [3]Advanced Light Source, Lawrence Berkeley National Laboratory, Berkeley, CA 94305, USA. [4]Swiss Light Source, Paul Scherrer Institute, CH-5232 Villigen, Switzerland. [5]Physik-Institut, Universität Zürich-Irchel, 8057 Zürich, Switzerland. [6]Institut für Festkörperforschung, Forschungszentrum Jülich, D-52425 Jülich, Germany. [7]Department of Physics and Astronomy, University of Pennsylvania, Philadelphia, PA 19104, USA. [8]Department of Chemistry, Princeton University, Princeton, NJ 08544, USA.

*To whom correspondence should be addressed. E-mail: mzhasan@princeton.edu

factor of π (*16*) because it evolves by 360° in momentum space along a Fermi contour enclosing a $\vec{k}_T$, an odd number of Fermi pockets enclosing $\vec{k}_T$ in total implies a π geometrical (Berry's) phase (*11*). In order to realize a π Berry's phase, the surface bands must be spin-polarized and exhibit a partner-switching (*11*) dispersion behavior between a pair of $\vec{k}_T$. This means that any pair of spin-polarized surface bands that are degenerate at $\bar{\Gamma}$ must not reconnect at $\bar{M}$ or must separately connect to the bulk valence and conduction band in between $\bar{\Gamma}$ and $\bar{M}$. The partner-switching behavior is realized in Fig. 1C because the spin-down band connects to and is degenerate with different spin-up bands at $\bar{\Gamma}$ and $\bar{M}$. The partner-switching behavior is realized in Fig. 2A because the spin-up and spin-down bands emerging from $\bar{\Gamma}$ separately merge into the bulk valence and conduction bands, respectively, between $\bar{\Gamma}$ and $\bar{M}$.

We first investigated the spin properties of the topological insulator phase. Spin-integrated ARPES (*19*) intensity maps of the (111) SSs of insulating $Bi_{1-x}Sb_x$ taken at the Fermi level (E_F) (Fig. 1, D and E) show that a hexagonal FS encloses $\bar{\Gamma}$, whereas dumbbell-shaped FS pockets that are much weaker in intensity enclose $\bar{M}$. By examining the surface-band dispersion below the Fermi level (Fig. 1F), it is clear that the central hexagonal FS is formed by a single band (Fermi crossing 1), whereas the dumbbell-shaped FSs are formed by the merger of two bands (Fermi crossings 4 and 5) (*10*).

This band dispersion resembles the partner-switching dispersion behavior characteristic of topological insulators. To check this scenario and determine the topological index ν_0, we have carried out spin-resolved photoemission spectroscopy. Figure 1G shows a spin-resolved momentum distribution curve taken along the $\bar{\Gamma}-\bar{M}$ direction at a binding energy $E_B = -25$ meV (Fig. 1G). The data reveal a clear difference between the spin-up and spin-down intensities of bands 1, 2, and 3 and show that bands 1 and 2 have opposite spin whereas bands 2 and 3 have the same spin. The former observation confirms that bands 1 and 2 form a spin-orbit split pair, and the latter observation suggests that bands 2 and 3 (as opposed to bands 1 and 3) are connected above the Fermi level and form one band. This is further confirmed by directly imaging the bands through raising the chemical potential via doping [supporting online material (SOM) text] (*22*). Irrelevance of bands 2 and 3 to the topology is consistent with the fact that the FS pocket they form does not enclose any $\vec{k}_T$. Because of a dramatic intrinsic weakening of signal intensity near crossings 4 and 5, and the small energy and momentum splittings of bands 4 and 5 lying at the resolution limit of modern spin-ARPES spectrometers, no conclusive spin information about these two bands can be drawn from the methods employed in obtaining the data sets in Fig. 1, G and H. However, whether bands 4 and 5 are both singly or doubly degenerate does not change the fact that an odd number of spin-polarized FSs enclose the $\vec{k}_T$, which provides evidence that $Bi_{1-x}Sb_x$ has $\nu_0 = 1$ and that its surface supports a nontrivial Berry's phase of topological origin.

We investigated the quantum origin of topological order in this class of materials. It has been theoretically speculated that the novel topological order originates from the parities of the electrons in pure Sb (*11*, *23*). The origin of the topological effects can only be tested by measuring the spin texture of the Sb surface (*20*), which has not been measured. Based on quantum oscillation and magneto-optical studies, the bulk band structure of Sb is known to evolve from that of insulating $Bi_{1-x}Sb_x$ through the holelike band at *H* rising above E_F and the electron-like band at *L* sinking below E_F (*23*). The relative energy ordering of the L_a and L_s states in Sb again determines whether the SS pair emerging from $\bar{\Gamma}$ switches partners (Fig. 2A) or not (Fig. 2B) between $\bar{\Gamma}$ and $\bar{M}$, and in turn determines whether they support a nonzero Berry's phase.

In a conventional spin-orbit metal such as Au, a free-electron–like SS is split into two parabolic spin-polarized sub-bands that are shifted in $\vec{k}$-space relative to each other (*18*). Two concentric spin-polarized FSs are created, one having an opposite sense of in-plane spin rotation from the other, that enclose $\bar{\Gamma}$. Such a FS arrangement, like the schematic shown in Fig. 2B, does not support a nonzero Berry's phase because the $\vec{k}_T$ are enclosed an even number of times (two times for most known materials).

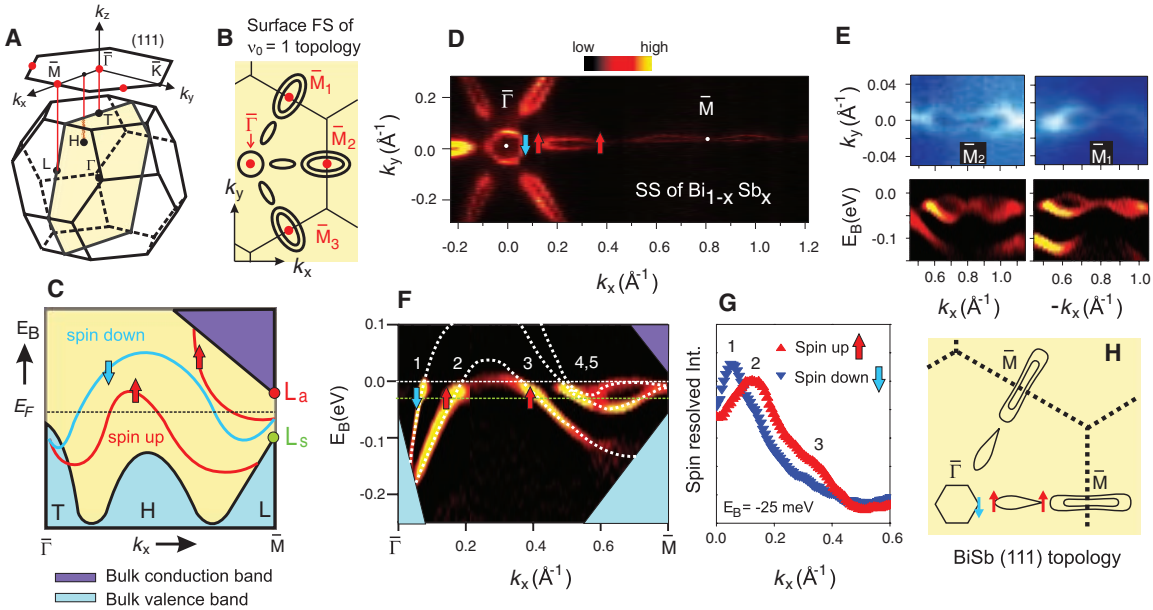


Fig. 1. Spin spectrum of a topological insulator and spin-resolved spectroscopy results. (**A**) Schematic sketches of the bulk BZ and (111) surface BZ of the $Bi_{1-x}Sb_x$ crystal series. The high symmetry points (*L*, *H*, *T*, Γ, $\bar{\Gamma}$, $\bar{M}$, $\bar{K}$) are identified. (**B**) Schematic of FS pockets formed by the SSs of a topological insulator that carries a π Berry's phase. (**C**) Partner-switching band structure topology: schematic of spin-polarized SS dispersion and connectivity between $\bar{\Gamma}$ and $\bar{M}$ required to realize the FS pockets shown in (B). L_a and L_s label bulk states at *L* that are antisymmetric and symmetric, respectively, under a parity transformation. (**D**) Spin-integrated ARPES intensity map of the SS of $Bi_{0.91}Sb_{0.09}$ at E_F. Arrows point in the measured direction of the spin. (**E**) High-resolution ARPES intensity map of the SS at E_F that enclose the $\bar{M}_1$ and $\bar{M}_2$ points. Corresponding band dispersion (second derivative images) are shown below. The left-right asymmetry of the band dispersions are due to the slight offset of the alignment from the $\bar{\Gamma}-\bar{M}_1$ ($\bar{M}_2$) direction. (**F**) Surface band–dispersion image along the $\bar{\Gamma}-\bar{M}$ direction showing five Fermi-level crossings. The intensity of bands 4 and 5 is scaled up for clarity (the dashed white lines are guides to the eye). The schematic projection of the bulk valence and conduction bands are shown in shaded blue and purple areas. (**G**) Spin-resolved momentum distribution curves presented at $E_B = -25$ meV showing single-spin degeneracy of bands at 1, 2, and 3. Spin up and down correspond to spin pointing along the $+\hat{y}$ and $-\hat{y}$ direction, respectively. (**H**) Schematic of the spin-polarized surface FS observed in our experiments. It is consistent with a $\nu_0 = 1$ topology [compare (B) and (H)].

However, for Sb, this is not the case. Figure 2C shows a spin-integrated ARPES intensity map of Sb(111) from $\bar{\Gamma}$ to $\bar{M}$. By performing a systematic incident photon energy–dependence study of such spectra, previously unavailable with He lamp sources (*24*), it is possible to identify two V-shaped SSs centered at $\bar{\Gamma}$, a bulk state located near k_x = –0.25 Å^{-1} and resonance states centered about k_x = 0.25 Å^{-1} and $\bar{M}$ that are hybrid states formed by surface and bulk states (SOM text) (*19*, *22*). An examination of the ARPES intensity map of the Sb(111) surface and resonance states at E_F (Fig. 2E) reveals that the central surface FS enclosing $\bar{\Gamma}$ is formed only by the inner V-shaped SS. The outer V-shaped SS, however, forms part of a teardrop-shaped FS that does not enclose $\bar{\Gamma}$, unlike the case in Au. This teardrop-shaped FS is formed partly by the outer V-shaped SS and partly by the holelike resonance state. The electron-like resonance state FS enclosing $\bar{M}$ does not affect the determination of ν_0 because it must be doubly spin-degenerate (SOM text) (*22*). Such a FS geometry (Fig. 2G) suggests that the V-shaped SS pair may undergo a partner-switching behavior as expected in Fig. 2A. This behavior is most clearly seen in a cut taken along the $\bar{\Gamma}$ – $\bar{K}$ direction because the top of the bulk valence band is well below E_F (Fig. 2F), showing only the inner V-shaped SS crossing E_F whereas the outer V-shaped SS bends back toward the bulk valence band near k_x = 0.1 Å^{-1} before reaching E_F. The additional support for this band-dispersion behavior comes from tight binding-surface calculations on Sb (Fig. 2D), which closely match with experimental data below E_F. Our observation of a single-surface band forming a FS enclosing $\bar{\Gamma}$ suggests that pure Sb is probably described by ν_0 = 1 and that its surface may support a Berry's phase.

Confirmation of a surface π Berry's phase rests critically on a measurement of the relative spin orientations (up or down) of the SS bands near $\bar{\Gamma}$ so that the partner switching is indeed realized, which cannot be done without spin resolution. We achieved spin resolution by using a Mott polarimeter that measures two orthogonal spin components of a photoemitted electron (*25*, *26*). These two components are along the y' and z' directions of the Mott coordinate frame, which lie predominantly in and out of the sample (111) plane respectively. Each of these two directions represents a normal to a scattering plane defined by the photoelectron incidence direction on a Au foil and two electron detectors mounted on either side (left and right) (Fig. 3A). Strong spin-orbit coupling of atomic Au is known to create an asymmetry in the scattering of a photoelectron off the Au foil that depends on its spin component normal to the scattering plane (*26*). This leads to an asymmetry $A_{y',z'}$ between the left intensity ($I^L_{y',z'}$) and right intensity ($I^R_{y',z'}$) given by $A_{y',z'} = [(I^L_{y',z'} - I^R_{y',z'})/(I^L_{y',z'} + I^R_{y',z'})]$, which is related to the spin polarization $P_{y',z'} = (1/S_{\text{eff}}) \times A_{y',z'}$ through the Sherman function S_{eff} = 0.085 (*25*, *26*). Spin-resolved momentum-distribution curve data sets of the SS bands along the $-\bar{M}-\bar{\Gamma}-\bar{M}$ cut at E_B = –30 meV (Fig. 3B) are shown for maximal intensity. Figure 3D displays both y' and z' polarization components along this cut, showing clear evidence that the bands are spin polarized with spins pointing largely in the (111) plane. In order to estimate the full 3D spin-polarization vectors from a two-component measurement (which is not required to prove the topological partner switching or the Berry's phase), we fit a model polarization curve (*27*), which takes the polarization directions associated with each momentum-distribution curve peak (Fig. 3C) as input parameters, with the constraint that each polarization vector has length one (in angular momentum units of $\hbar/2$, $\hbar$ being h divided by 2π). Our fitted polarization vectors are displayed in the sample (x, y, z) coordinate frame (Fig. 3F), from which we derived the spin-resolved momentum distribution curves for the spin components parallel ($I_y^{\uparrow}$) and anti-parallel ($I_y^{\downarrow}$) to the y direction as shown in Fig. 3E (SOM text) (*22*). There is a clear difference in $I_y^{\uparrow}$ and $I_y^{\downarrow}$ at each of the four momentum-distribution curve peaks, indicating that the SS bands are spin-polarized (Fig. 3E), which is possible to conclude even without a full 3D fitting. Each of the pairs *l*2/*l*1 and *r*1/*r*2 have opposite spin, which is consistent with the behavior of a spin-split pair, and the spin polarization of these bands is reversed on either side of $\bar{\Gamma}$ in accordance with the system being time-reversal symmetric [$E(\vec{k},\uparrow) = E(-\vec{k},\downarrow)$] (Fig. 3F). The measured spin texture of the Sb(111) SSs (Fig. 3), together with the connectivity of the surface bands (Fig. 2), distinctively determines its belonging to the ν_0 = 1 class. Therefore, the surface of Sb carries a nonzero (π) Berry's phase via the inner V-shaped band, and pure Sb can be regarded as the parent matrix of the $Bi_{1-x}Sb_x$ topological insulator class; the topological order originates from the parity set of Sb wave functions.

Our spin-polarized measurement methods (Figs. 1 and 3) uncover a new type of topological quantum number n_M that provides information about the chirality properties. Topological theory suggests that the bulk electronic states in the mirror (k_y = 0) plane can be classified in terms of a number n_M (= ±1) that describes the handedness (either left- or right-handed) or chirality of the surface spins that can be directly measured or seen

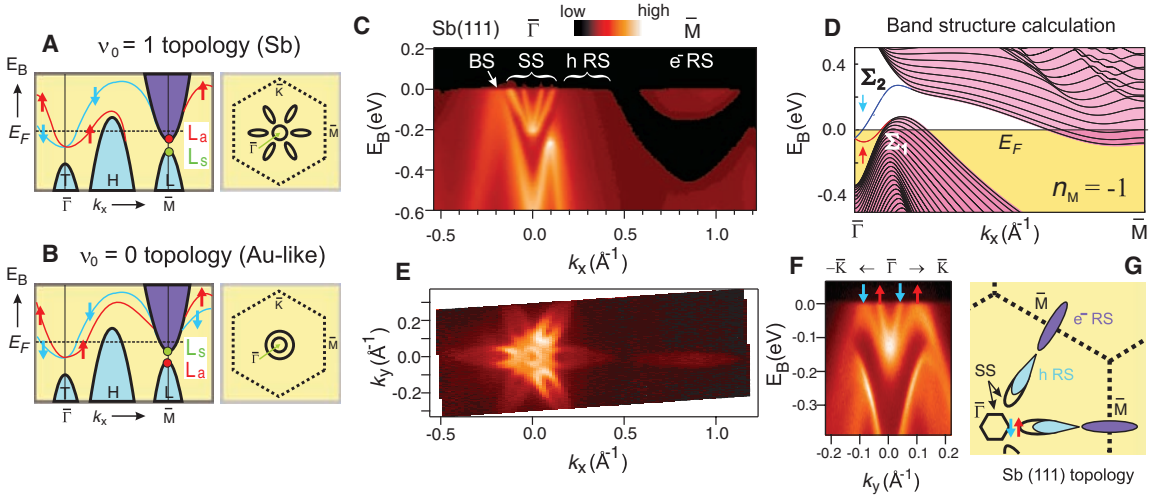


Fig. 2. Topological character of pure Sb revealed on the (111) SSs. Schematic of the bulk-band structure (shaded areas) and surface-band structure (red and blue lines) of Sb near E_F for (**A**) topologically nontrivial and (**B**) topologically trivial (Au-like) case, together with their corresponding surface FSs are shown. (**C**) Spin-integrated ARPES spectrum of Sb(111) along the $\bar{\Gamma}-\bar{M}$ direction. The surface states are denoted by SS, bulk states by BS, and the holelike resonance states and electron-like resonance states by hRS and e$^-$RS, respectively. (**D**) Calculated SS band structure of Sb(111) based on the methods in (*20*, *30*). The continuum bulk-energy bands are represented with pink-shaded regions, and the lines show the discrete bands of a 100-layer slab. The red and blue single bands, denoted Σ_1 and Σ_2, are the SS bands with spin polarization $\langle \vec{P} \rangle \propto +\hat{y}$ and $\langle \vec{P} \rangle \propto -\hat{y}$, respectively. (**E**) ARPES intensity map of Sb(111) at E_F in the k_x–k_y plane. The only one FS encircling $\bar{\Gamma}$ seen in the data is formed by the inner V-shaped SS band seen in (C) and (F). The outer V-shaped band bends back toward the bulk band that is best seen in data in (F). (**F**) ARPES spectrum of Sb(111) along the $\bar{\Gamma}-\bar{K}$ direction shows that the outer V-shaped SS band merges with the bulk band. (**G**) Schematic of the surface FS of Sb(111) showing the pockets formed by the SSs (unfilled) and the resonant states (blue and purple). The purely SS Fermi pocket encloses only one $\vec{k}_T$, namely $\bar{\Gamma}(=\vec{k}_T)$, and is therefore consistent with the ν_0 = 1 topological classification of Sb, which is different from Au [compare (B) and (G)]. As discussed in the text, the *h*RS and e^-RS count trivially.

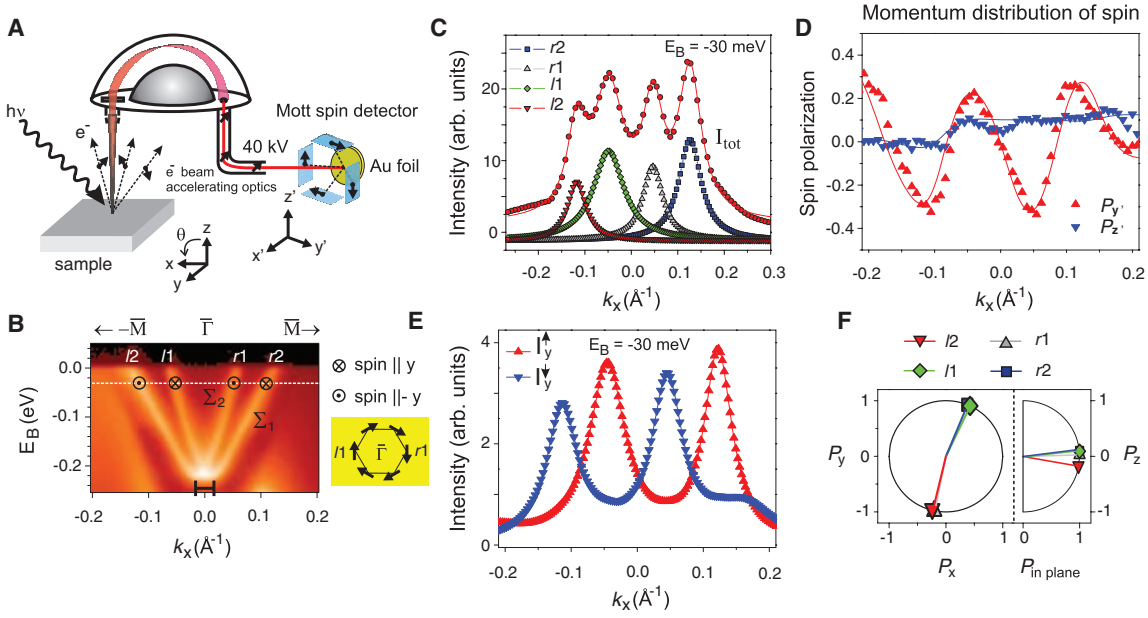


Fig. 3. Spin texture of topological surface-edge states and chirality. (**A**) Experimental geometry of the spin-ARPES study. At normal emission ($\theta = 0°$), the sensitive y' axis of the Mott detector is rotated by 45° from the sample $\bar{\Gamma}$ to $-\bar{M}$ ($\| -\hat{x}$) direction, and the sensitive z' axis of the Mott detector is parallel to the sample normal ($\| \hat{z}$). (**B**) Spin-integrated ARPES spectrum of Sb(111) along the $-\bar{M}$–$\bar{\Gamma}$–$\bar{M}$ direction. The momentum splitting between the band minima is indicated by the black bar and is approximately 0.03 Å^{-1}. A schematic of the spin chirality of the central FS based on the spin-ARPES results is shown on the right. (**C**) Momentum-distribution curve of the spin-averaged spectrum at $E_B = -30$ meV [shown in (B) by white line] together with the Lorentzian peaks of the fit. (**D**) Measured spin-polarization curves (red and blue triangles) for the detector y' and z' components together with the fitted lines by using the two-step fitting routine (*27*). (**E**) Spin-resolved spectra for the sample y component based on the fitted spin-polarization curves shown in (D). Up triangles represent a spin direction along the $+\hat{y}$ direction, and down triangles represent a spin direction along the $-\hat{y}$ direction. (**F**) The in-plane and out-of-plane spin polarization components in the sample coordinate frame obtained from the spin-polarization fit. Overall spin-resolved data and the fact that the surface band that forms the central electron pocket has $\langle \vec{P} \rangle \propto -\hat{y}$ along the $+k_x$ direction, as in (E), suggest a left-handed chirality [schematic in (B)].

in spin-resolved experiments (*20*). We next determined the value of n_M from our data: From Fig. 1, it is seen that a single- (one-) surface band, which switches partners at $\bar{M}$, connects the bulk valence and conduction bands, so $|n_M| = 1$ (SOM text) (*22*). The sign of n_M is related to the direction of the spin polarization $\langle \vec{P} \rangle$ of this band (*20*), which is constrained by mirror symmetry to point along $\pm\hat{y}$. Because the central electron-like FS enclosing $\bar{\Gamma}$ intersects six mirror invariant points (Fig. 3B), the sign of n_M distinguishes two distinct types of handedness for this spin-polarized FS. Figures 1F and 3 show that for both $Bi_{1-x}Sb_x$ and Sb, the surface band that forms this electron pocket has $\langle \vec{P} \rangle \propto -\hat{y}$ along the k_x direction, which suggests a left-handed rotation sense for the spins around this central FS and thus $n_M = -1$. Therefore, both insulating $Bi_{1-x}Sb_x$ and pure Sb possess equivalent chirality properties: a definite spin-rotation sense (left-handed chirality) (Fig. 3B) and a topological Berry's phase.

These spin-resolved experimental measurements reveal an intimate and straightforward connection between the topological numbers (ν_0, n_M) and the physical observables. Although ν_0 determines whether the surface electrons support a nontrivial Berry's phase, if they do, the n_M determines the spin-handedness of the FS that manifests this Berry's phase. The 2D topological Berry's phase is a critical signature of group Z_2 topological order and is not realizable in isolated 2D electron systems nor on the surfaces of conventional spin-orbit or exchange-coupled magnetic materials. A nonzero Berry's phase is known, theoretically, to protect an electron system against the almost universal weak-localization behavior in their low-temperature transport (*11*, *13*) and is expected to form the key element for fault-tolerant quantum computation schemes (*13*, *28*, *29*). Its remarkable realization on the $Bi_{1-x}Sb_x$ surface represents an unprecedented example of a 2D π Berry's phase and opens the possibility for building realistic prototype systems to test quantum computing modules. In general, our results demonstrate that spin-ARPES is a powerful probe of topological order and quantum spin Hall physics, which opens up a new search front for topological materials for novel spin devices and fault-tolerant quantum computing.

References and Notes

1. N. W. Ashcroft, N. D. Mermin, *Solid State Physics* (Holt, Rinehart and Winston, NY, 1976).
2. K. von Klitzing, G. Dorda, M. Pepper, *Phys. Rev. Lett.* **45**, 494 (1980).
3. R. B. Laughlin, *Phys. Rev. Lett.* **50**, 1395 (1983).
4. X.-G. Wen, *Int. J. Mod. Phys. B* **4**, 239 (1990).
5. D. J. Thouless *et al.*, *Phys. Rev. Lett.* **49**, 405 (1982).
6. J. E. Avron, D. Osadchy, R. Seiler, *Phys. Today* **56**, 38 (2003).
7. M. König *et al.*, *Science* **318**, 766 (2007).
8. B. A. Bernevig, T. L. Hughes, S.-C. Zhang, *Science* **314**, 1757 (2006).
9. N. Nagaosa, *Science* **318**, 758 (2007).
10. D. Hsieh *et al.*, *Nature* **452**, 970 (2008).
11. L. Fu, C. L. Kane, *Phys. Rev. B* **76**, 045302 (2007).
12. X.-L. Qi, T. Hughes, S.-C. Zhang, *Phys. Rev. B* **78**, 195424 (2008).
13. L. Fu, C. L. Kane, *Phys. Rev. Lett.* **100**, 096407 (2008).
14. C. L. Kane, E. J. Mele, *Phys. Rev. Lett.* **95**, 146802 (2005).
15. J. E. Moore, L. Balents, *Phys. Rev. B* **75**, 121306 (2007).
16. J. J. Sakurai, *Modern Quantum Mechanics* (Addison-Wesley, NY, 1994).
17. M. V. Berry, *Proc. R. Soc. London Ser. A* **392**, 45 (1984).
18. M. Hoesch *et al.*, *Phys. Rev. B* **69**, 241401 (2004).
19. S. Hufner, *Photoelectron Spectroscopy* (Springer-Verlag, Berlin, 1995).
20. J. C. Y. Teo, L. Fu, C. L. Kane, *Phys. Rev. B* **78**, 045426 (2008).
21. T. Hirahara *et al.*, *Phys. Rev. B* **76**, 153305 (2007).
22. Materials and methods, including details of sample characterization, data analysis, and theoretical background, are available as supporting material on *Science* Online.
23. B. Lenoir *et al.*, *Int. Conf. Thermoelec.* **15**, 1 (1996).
24. K. Sugawara *et al.*, *Phys. Rev. Lett.* **96**, 046411 (2006).
25. M. Hoesch *et al.*, *J. Electron Spectrosc. Relat. Phenom.* **124**, 263 (2002).
26. T. J. Gay, F. B. Dunning, *Rev. Sci. Instrum.* **63**, 1635 (1992).
27. F. Meier *et al.*, *Phys. Rev. B* **77**, 165431 (2008).
28. P. J. Leek *et al.*, *Science* **318**, 1889 (2007).
29. A. Kitaev, *Ann. Phys. (NY)* **303**, 2 (2003).
30. Y. Liu, E. Allen, *Phys. Rev. B* **52**, 1566 (1995).

31. We thank J. Teo for providing the SS band calculations of antimony (Sb); A. Fedorov, L. Patthey, and D.-H. Lu for beamline assistance; and D. Haldane, B. I. Halperin, N. P. Ong, D. A. Huse, F. Wilczek, P. W. Anderson, D. C. Tsui, J. E. Moore, L. Fu, L. Balents, D.-H. Lee, S. Sachdev, P. A. Lee, and X.-G. Wen for stimulating discussions. C.L.K. was supported by NSF grant DMR-0605066. The spin-resolved ARPES experiments are supported by NSF through the Center for Complex Materials (DMR-0819860) and Princeton University; the use of synchrotron X-ray facilities (ALS-LBNL, Berkeley, and SSRL-SLAC, Stanford) is supported by the Basic Energy Sciences of the U.S. Department of Energy (DE-FG-02–05ER46200) and by the Swiss Light Source, Paul Scherrer Institute, Villigen, Switzerland.

Supporting Online Material

www.sciencemag.org/cgi/content/full/323/5916/919/DC1
Materials and Methods
SOM Text
Figs. S1 to S6
References

27 October 2008; accepted 7 January 2009
10.1126/science.1167733

Coherence Factors in a High-T_c Cuprate Probed by Quasi-Particle Scattering Off Vortices

T. Hanaguri,[1,2]* Y. Kohsaka,[1,2] M. Ono,[1,2] M. Maltseva,[3] P. Coleman,[3] I. Yamada,[4]† M. Azuma,[4] M. Takano,[4]‡ K. Ohishi,[5]§ H. Takagi[1,6]

When electrons pair in a superconductor, quasi-particles develop an acute sensitivity to different types of scattering potential that is described by the appearance of coherence factors in the scattering amplitudes. Although the effects of coherence factors are well established in isotropic superconductors, they are much harder to detect in their anisotropic counterparts, such as high-superconducting-transition-temperature cuprates. We demonstrate an approach that highlights the momentum-dependent coherence factors in $Ca_{2-x}Na_xCuO_2Cl_2$. We used Fourier-transform scanning tunneling spectroscopy to reveal a magnetic-field dependence in quasi-particle scattering interference patterns that is sensitive to the sign of the anisotropic gap. This result is associated with the *d*-wave coherence factors and quasi-particle scattering off vortices. Our technique thus provides insights into the nature of electron pairing as well as quasi-particle scattering processes in unconventional superconductors.

Superconductivity is characterized by macroscopic phase coherence, as exemplified by the development of a complex order parameter. In this respect, superconductors are similar to bosonic superfluids, such as ^{4}He, and many phenomenological features are common to both, such as vortex quantization and the Josephson effect (*1*). However, superconductors are microscopically distinct from bosonic superfluids; whereas bosons individually condense, electrons condense as Cooper pairs.

The internal structure of the condensed pairs strongly influences the properties of quasi-particle excitations, enforcing coherence between two quasi-particle scattering processes, $\mathbf{k}_i \rightarrow \mathbf{k}_f$ and $-\mathbf{k}_f \rightarrow -\mathbf{k}_i$. ($\mathbf{k}_i$ and $\mathbf{k}_f$ denote initial and final state momenta, respectively.) This manifests itself as a coherence factor $C(\mathbf{k}_i, \mathbf{k}_f)$ in the scattering matrix element, which is sensitive to the momentum-dependent phase of the superconducting (SC) order parameter and the time-reversal symmetry of the scattering potential (*2*).

In isotropic *s*-wave superconductors, the effect of coherence factors is manifested in the temperature dependence of various measurable quantities associated with quasi-particle scattering. For example, the nuclear-spin relaxation rate exhibits an enhancement just below the SC transition temperature, T_c, called the Hebel-Slichter peak (*3*), which is a consequence of coherence factors. In unconventional superconductors, however, the effect of coherence factors is suppressed by the strong momentum, $\mathbf{k}$, dependence of the anisotropic SC gap (*4*).

We have developed a technique that highlights the $\mathbf{k}$-dependent coherence factors in anisotropic superconductors by introducing vortices as controllable scatterers. The vortices induced by a magnetic field, B, generate extra quasi-particle scatterings around them, in which coherence factor effects can be detected. Quasi-particle scattering can be examined through the Fourier analysis of quasi-particle interference (QPI) patterns imaged by spectroscopic-imaging scanning tunneling microscopy (SI-STM) (*5*–*8*).

We performed experiments at 1.6 K on a high-T_c cuprate, $Ca_{2-x}Na_xCuO_2Cl_2$ ($x \sim 0.14$ and $T_c \sim 28$ K), with a *d*-wave SC gap. QPI patterns in the high-T_c cuprates have been observed in the absence of a field (*5*–*8*) and analyzed in terms of the so-called octet model (*6*), which is based only on the joint density of states (DOS) in $\mathbf{k}$ space. This model predicts that seven distinct scattering wavevectors $\mathbf{q}_i$ ($i = 1 \sim 7$) connecting the ends of the banana-shaped constant energy contours in $\mathbf{k}$-space (Fig. 1A) will dominate QPI (*9*).

We show that the coherence of quasi-particle scattering induced by pair formation causes the scattering probability to acquire an additional $\mathbf{k}$ dependence determined by the coherence factor, $C(\mathbf{k}_i, \mathbf{k}_f)$. $C(\mathbf{k}_i, \mathbf{k}_f)$ is given by a combination of Bogoliubov coefficients, $u_\mathbf{k} = \mathrm{sgn}[\Delta(\mathbf{k})]\sqrt{[1 + \varepsilon(\mathbf{k})/E(\mathbf{k})]/2}$ and $v_\mathbf{k} = \sqrt{1 - u_\mathbf{k}^2}$, where $\varepsilon(\mathbf{k})$ and $\Delta(\mathbf{k})$ are

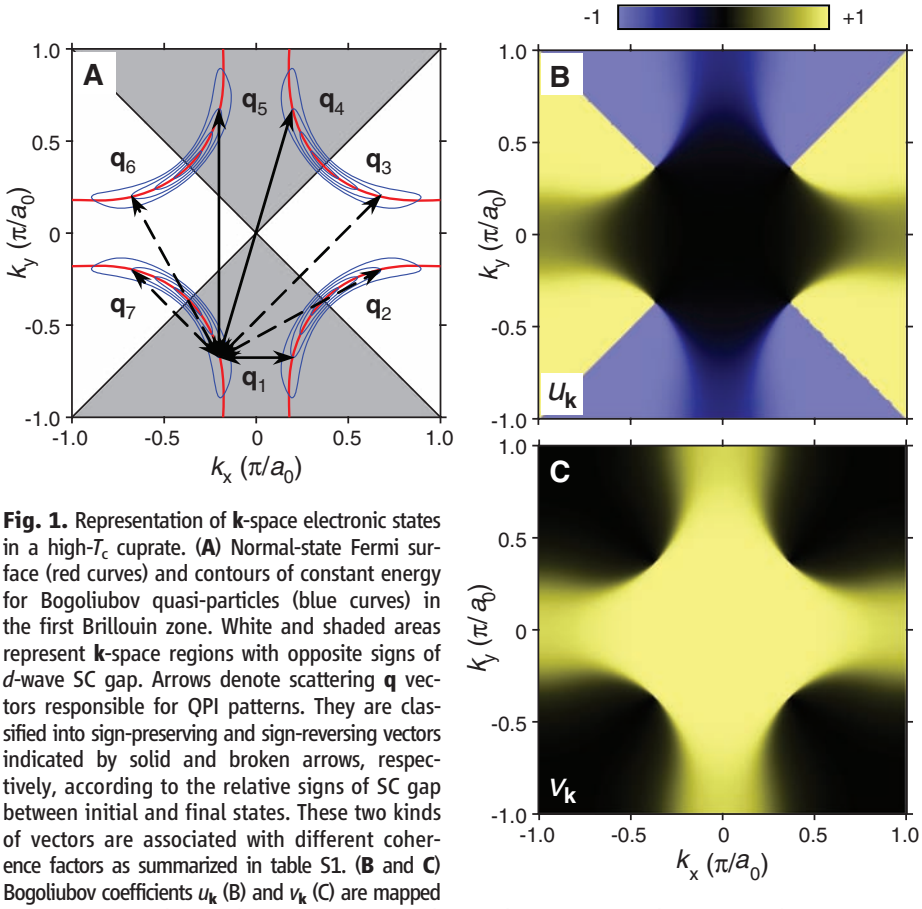


Fig. 1. Representation of $\mathbf{k}$-space electronic states in a high-T_c cuprate. (**A**) Normal-state Fermi surface (red curves) and contours of constant energy for Bogoliubov quasi-particles (blue curves) in the first Brillouin zone. White and shaded areas represent $\mathbf{k}$-space regions with opposite signs of *d*-wave SC gap. Arrows denote scattering $\mathbf{q}$ vectors responsible for QPI patterns. They are classified into sign-preserving and sign-reversing vectors indicated by solid and broken arrows, respectively, according to the relative signs of SC gap between initial and final states. These two kinds of vectors are associated with different coherence factors as summarized in table S1. (**B** and **C**) Bogoliubov coefficients $u_\mathbf{k}$ (B) and $v_\mathbf{k}$ (C) are mapped in $\mathbf{k}$ space. Note that $u_\mathbf{k}$ changes its sign according to that of SC gap, whereas $v_\mathbf{k}$ is always positive.

[1]Magnetic Materials Laboratory, RIKEN Advanced Science Institute, Wako 351-0198, Japan. [2]Core Research for Evolutional Science and Techonology (CREST), Japan Science and Technology Agency, Kawaguchi 332–0012, Japan. [3]Department of Physics and Astronomy, Rutgers University, Piscataway, NJ 08854–8019, USA. [4]Institute for Chemical Research, Kyoto University, Uji 601–0011, Japan. [5]Advanced Science Research Center, Japan Atomic Energy Agency, Ibaraki 319–1195, Japan. [6]Department of Advanced Materials, University of Tokyo, Kashiwa 277–8561, Japan.

*To whom correspondence should be addressed. E-mail: hanaguri@riken.jp

†Present address: Department of Chemistry, Ehime Universiy, Matsuyama 790–8577, Japan.

‡Present address: Institute for Integrated Cell-Material Sciences, Kyoto University, care of Research Institute for Production Development, Kyoto 606-0805 Japan.

§Present address: Advanced Meson Science Laboratory, RIKEN, Wishina Center for Accelerator-Based Science, Wako 351-0198, Japan.

dispersion relations of normal-state band and SC gap, respectively, and $E(\mathbf{k}) = \sqrt{\varepsilon(\mathbf{k})^2 + \Delta(\mathbf{k})^2}$. The detailed form of $C(\mathbf{k}_i, \mathbf{k}_f)$ depends on the nature of the scatterer (table S1). For a scalar potential, which is even under time reversal, $C(\mathbf{k}_i, \mathbf{k}_f) = (u_{\mathbf{k}_i} u_{\mathbf{k}_f} - v_{\mathbf{k}_i} v_{\mathbf{k}_f})^2$; for a magnetic scattering potential, which is odd under time reversal, $C(\mathbf{k}_i, \mathbf{k}_f) = (u_{\mathbf{k}_i} u_{\mathbf{k}_f} + v_{\mathbf{k}_i} v_{\mathbf{k}_f})^2$. Scattering off inhomogeneities in the SC gap amplitude, which is a kind of inhomogeneous Andreev reflection process that converts electrons into holes as they are scattered, gives rise to the coherence factor $C(\mathbf{k}_i, \mathbf{k}_f) = (u_{\mathbf{k}_i} v_{\mathbf{k}_f} + v_{\mathbf{k}_i} u_{\mathbf{k}_f})(u_{\mathbf{k}_i} u_{\mathbf{k}_f} + v_{\mathbf{k}_i} v_{\mathbf{k}_f}) \propto \Delta(\mathbf{k}_i) + \Delta(\mathbf{k}_f)$ (*2*, *10*–*14*).

As shown in Fig. 1, B and C, $u_{\mathbf{k}}$ changes its sign in the same way as $\Delta(\mathbf{k})$, whereas $v_{\mathbf{k}}$ is always positive. This leads to a systematic "extinction rule" that depends on the nature of the scatterer. In the case of weak scalar potential scattering, the coherence factor $C(\mathbf{k}_i, \mathbf{k}_f)$ is heavily suppressed for those $\mathbf{q}_i$ that preserve the sign of the SC gap $\Delta(\mathbf{k})$, namely, $\mathbf{q}_1$, $\mathbf{q}_4$, and $\mathbf{q}_5$. By contrast, for scattering off magnetic impurities or gap inhomogeneities, $C(\mathbf{k}_i, \mathbf{k}_f)$ is heavily suppressed for those $\mathbf{q}_i$ that reverse the sign of $\Delta(\mathbf{k})$. Thus, depending on the type of disorder, sign-reversing or sign-preserving scatterings will dominate (*10*–*14*). In this way, QPI patterns can shed light onto the underlying nature of the quasi-particle scattering processes. However, in experiments carried out to date (*5*–*8*), each of the octet wave vectors $\mathbf{q}_i$ has been featured with comparable intensity, which indicates that more than one kind of scatterers is present in the samples, hiding the underlying effects of $C(\mathbf{k}_i, \mathbf{k}_f)$.

The introduction of vortices by applying a magnetic field B provides the system with scatterers with definite scattering wavevector $\mathbf{q} \equiv \mathbf{k}_f - \mathbf{k}_i$ selectivity. The phase of the SC gap precesses by 2π about each vortex, whereas the amplitude of the gap vanishes at its core. Both the phase gradient, proportional to the superfluid velocity, and the inhomogeneity in the SC gap amplitude, induced by the vortex core, can scatter quasi-particles. The inhomogeneous superflow about a vortex produces Doppler-shift scattering (*15*) that is odd under time reversal with $C(\mathbf{k}_i, \mathbf{k}_f) = (u_{\mathbf{k}_i} u_{\mathbf{k}_f} + v_{\mathbf{k}_i} v_{\mathbf{k}_f})^2$ like magnetic impurities. The spatial inhomogeneity in the SC gap amplitude causes inhomogeneous Andreev scattering with $C(\mathbf{k}_i, \mathbf{k}_f) = (u_{\mathbf{k}_i} v_{\mathbf{k}_f} + v_{\mathbf{k}_i} u_{\mathbf{k}_f})(u_{\mathbf{k}_i} u_{\mathbf{k}_f} + v_{\mathbf{k}_i} v_{\mathbf{k}_f})$. It should be noted that both of these scatterers selectively activate the sign-preserving $\mathbf{q}$ points. Therefore, the effect of $C(\mathbf{k}_i, \mathbf{k}_f)$ and the nature of quasi-particle scattering off vortices can be revealed through the $\mathbf{q}$ dependence of the B-induced change of QPI.

To image vortices and QPI patterns, we map the tunneling conductance, $g(\mathbf{r}, E) \equiv dI/dV(\mathbf{r}, E)$, where I and V are tunneling current and bias voltage, respectively. $g(\mathbf{r}, E)$ is a measure of the local DOS at location $\mathbf{r}$ and energy E (*9*). Vortices can be identified as regions where the gap in the local DOS is locally suppressed (Fig. 2, A to C). The number of these regions is consistent with the expected number of vortices.

In $Ca_{2-x}Na_xCuO_2Cl_2$, the raw $g(\mathbf{r}, E)$ map is dominated by checkerboard modulations (*16*) that mask the underlying QPI signal. As we reported earlier (*7*), the QPI signal is enhanced by taking the ratio $Z(\mathbf{r}, E) \equiv g(\mathbf{r}, E)/g(\mathbf{r}, -E)$. This procedure almost completely suppresses the checkerboard signal. It has an additional advantage of eliminating extrinsic effects associated with the scanning feedback loop, and thus $Z(\mathbf{r}, E)$ faithfully represents the local DOS ratio (*7*–*9*). Detailed characteristics of $Z(\mathbf{r}, E)$ have been discussed both theoretically and experimentally (*9*, *14*, *17*).

Figure 2, D and G, shows the zero-field $Z(\mathbf{r}, E = 4.4\ \text{meV})$ and its Fourier transform $|Z(\mathbf{q}, E = 4.4\ \text{meV})|$, which display the full set of discrete $\mathbf{q}$ points expected in the octet model. These discrete $\mathbf{q}$ points disperse with E in a fashion consistent with a d-wave SC gap up to 10 to 15 meV, which marks an upper limit for the detection of well-defined Bogoliubov quasiparticles (*7*, *8*) (fig. S1).

Repeating the measurements at $B = 5$ T and 11 T, we observed a remarkable field dependence in the intensities of the QPI patterns. The B dependence of $Z(\mathbf{r}, E)$ (Fig. 2, D to F) and the corresponding Fourier-transformed data (Fig. 2, G to I) show that field does not induce additional $\mathbf{q}$ vectors and that the positions of the peaks in $|Z(\mathbf{q}, E)|$ are only weakly B dependent. By contrast, there is a substantial B dependence in the intensities of the peaks. Depending on $\mathbf{q}_i$, intensity of peak is either enhanced or suppressed.

To explore the details of the field-induced intensity variations, we subtract the zero-field data

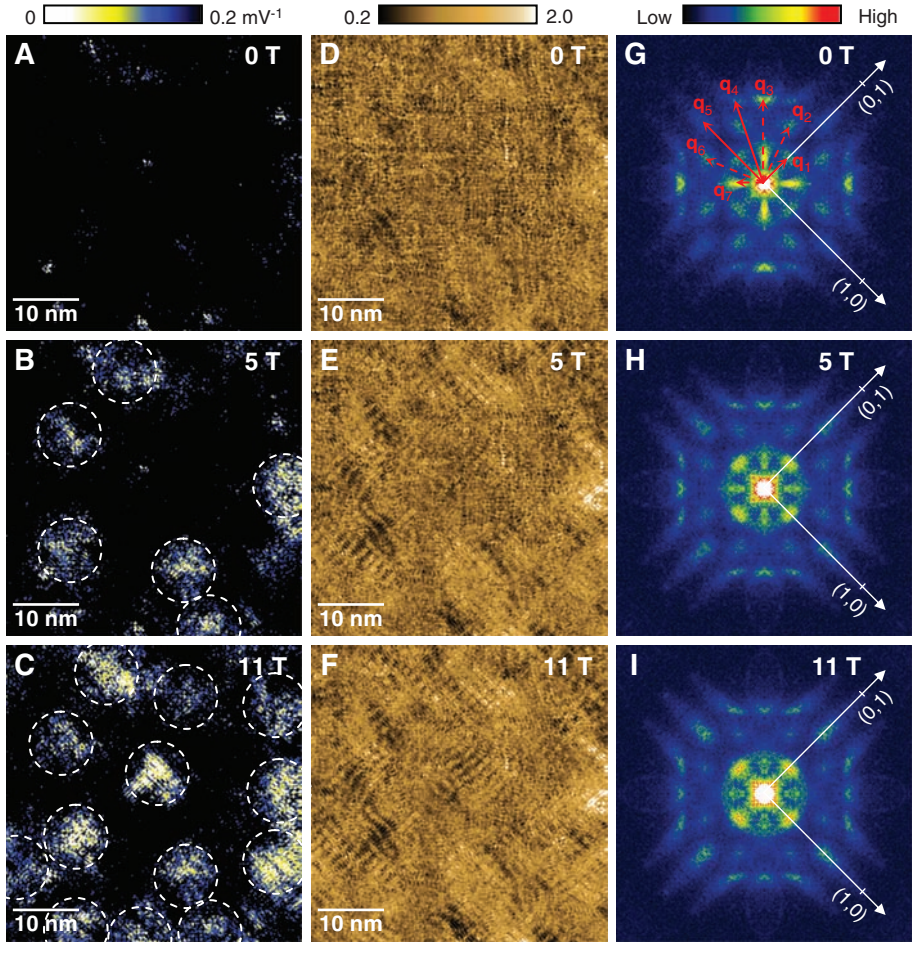


Fig. 2. Imaging vortices and QPI patterns in $Ca_{2-x}Na_xCuO_2Cl_2$ ($x \sim 0.14$ and $T_c \sim 28$ K) at different magnetic fields. All the data were collected with a setup condition of sample bias voltage $V_s = -100$ mV and tunneling current $I_t = 100$ pA. (**A** to **C**) Vortices imaged by mapping a function $s(\mathbf{r}, E) \equiv [dg(\mathbf{r}, +E)/dV]/g(\mathbf{r}, +E) - [dg(\mathbf{r}, -E)/dV]/g(\mathbf{r}, -E)$ at $E = 4.4$ meV. If there is a gap in the spectrum $g(\mathbf{r}, E)$, the function $s(\mathbf{r}, E)$ below the gap energy takes a larger value as gap structure becomes deeper, whereas it is almost zero if $g(\mathbf{r},E)$ is structureless. Vortices are imaged as shallower-gap regions [smaller $s(\mathbf{r}, E)$] shown in brighter color. Broken circles are guides to the eye. (**D** to **F**) Real-space QPI patterns at $E = 4.4$ meV imaged by mapping the conductance ratio $Z(\mathbf{r}, E)$. (**G** to **I**) $|Z(\mathbf{q}, E)|$ obtained by Fourier-transforming $Z(\mathbf{r}, E)$ shown in (D) to (F). In order to enhance the signal-to-noise ratio, each $|Z(\mathbf{q}, E)|$ map is averaged by folding it so as to superpose all the crystallographically equivalent $\mathbf{q}$ positions. Arrows in (G) correspond to those in Fig. 1A. All data were collected at 1.6 K.

$|Z(\mathbf{q}, E, B = 0)|$ from $|Z(\mathbf{q}, E, B)|$. The corresponding difference map is shown in Fig. 3A for $B = 11$ T. The $\mathbf{q}$ points can be classified into two groups: for $\mathbf{q}_1$, $\mathbf{q}_4$, and $\mathbf{q}_5$ the intensity is field-enhanced, whereas for $\mathbf{q}_2$, $\mathbf{q}_3$, $\mathbf{q}_6$, and $\mathbf{q}_7$ the intensity is depressed. These two groups are the sign-preserving and sign-reversing $\mathbf{q}$ points discussed earlier, and their selective enhancement and suppression imply the activation of coherence factors $C(\mathbf{k}_i, \mathbf{k}_f)$ induced by vortices.

The spatial resolution of SI-STM allows us to spatially resolve the origin of the $\mathbf{q}$-selective enhancement and suppression of the QPI signal and thereby to examine its relationship with the location of the vortices. For this purpose, we restricted the field of view to the vicinity of vortices (as indicated by the "vortex region" inside the blue lines in Fig. 3B) or to regions far from vortices (as indicated by the "matrix region" inside the red lines in Fig. 3B) and performed Fourier analyses separately for each region (fig. S2). As shown in Fig. 3,C and D, the enhanced sign-preserving QPI signals are concentrated in the vortex region, indicating that this signal is associated with quasi-particle scattering effects induced by vortices. By contrast, the suppressed sign-reversing QPI signals are apparently weak near vortices but are distributed throughout the matrix region far from the vortices.

We note as an aside that the enhanced sign-preserving $\mathbf{q}$ points $\mathbf{q}_1$ and $\mathbf{q}_5$ are very close to those $\mathbf{q}$ vectors that characterize the checkerboard electronic modulation observed in the vicinity of vortices in $Bi_2Sr_2CaCu_2O_y$ (*18*, *19*). Various charge- and spin-density-wave scenarios have been advanced to account for the vortex checkerboard modulation (*18*, *19*). The field-enhanced sign-preserving $\mathbf{q}$ values offer an alternate perspective on the origin of the vortex checkerboard; QPI may participate in the checkerboard formation with (or without) the possible electronic orders. Further studies of the relation between QPI and vortex checkerboard are needed to elucidate the electronic structure of vortices in high-T_c cuprates.

To account for the reduction of the scattering at the sign-reversing vectors $\mathbf{q}_2$, $\mathbf{q}_3$, $\mathbf{q}_6$, and $\mathbf{q}_7$ in the matrix region, we postulate the Doppler shift of quasi-particle energies induced by the superflow around vortices (*15*). The Doppler shift in the quasi-particle energies deforms the banana-shaped contours of constant energy by an amount proportional to the superfluid velocity. This has the effect of smearing the quasi-particle interference peaks, reducing their amplitude. This effect may have no momentum selectivity and will tend to uniformly depress the scattering at all octet momenta. We hypothesize that this effect is unable to completely mask out the strong enhancement effects induced by the vortices, so that the smearing effect is only observed at the sign-reversing momenta. A similar scenario has also been proposed by Pereg-Barnea and Franz (*14*).

Lastly, we examine the B dependence of Fermi surface topology and SC gap dispersion $\Delta(\mathbf{k})$ by analyzing the E dependence of $|Z(\mathbf{q}, E, B)|$. As shown in Fig. 4A, the Fermi surface displays no measurable B dependence up to 11 T as expected, because the corresponding Zeeman energy to B is negligibly small (<1 meV) compared with the hopping amplitude (~0.1 eV). On the other hand, $\Delta(\mathbf{k})$ shows a small but distinct field dependence (Fig. 4B). At $B = 0$, a linear extrapolation of $\Delta(\mathbf{k})$ from high E does not intercept the node at $\theta_\mathbf{k} = 45°$, where $\theta_\mathbf{k}$ is a Fermi surface angle around (π, π), and there is an apparent gapless region around the node. This gapless region expands in a field, indicating that zero-energy quasi-particles are generated by introducing vortices. In accordance with this interpretation, the observed DOS at the Fermi energy, given by the average value of $g(\mathbf{r}, E = 0)$, increases (Fig. 4C) (*20*) and is seen to follow a $B\log B$ behavior (*21*), as expected in a dirty d-wave super-

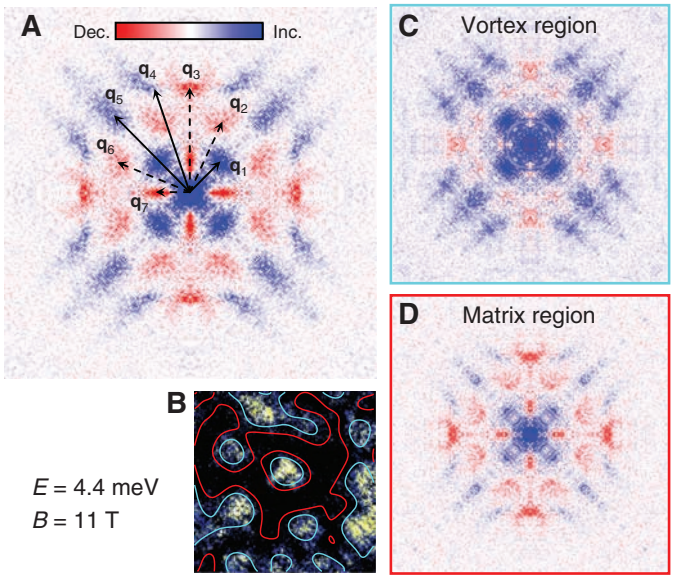


Fig. 3. Magnetic-field-induced weight transfer in $|Z(\mathbf{q}, E)|$ at $E = 4.4$ meV. (**A**) The difference map $|Z(\mathbf{q}, E, B)| - |Z(\mathbf{q}, E, B = 0)|$ for $B = 11$ T (namely, difference between Fig. 2I and Fig. 2G). Intensities of sign-preserving $\mathbf{q}$ points are field-enhanced, whereas those of sign-reversing ones are field-suppressed. (**B**) Vortex image reproduced from Fig. 1C showing the restricted field of views. Blue and red lines surround vortex and matrix regions, respectively (fig. S2). Magnetic-field-induced weight transfers are deduced separately for vortex and matrix regions as shown in (**C**) and (**D**), respectively. Intensities are normalized according to the area. Enhancement of sign-preserving scatterings at $\mathbf{q}_1$, $\mathbf{q}_4$, and $\mathbf{q}_5$ is remarkable near the vortices, whereas it is weak in the matrix region.

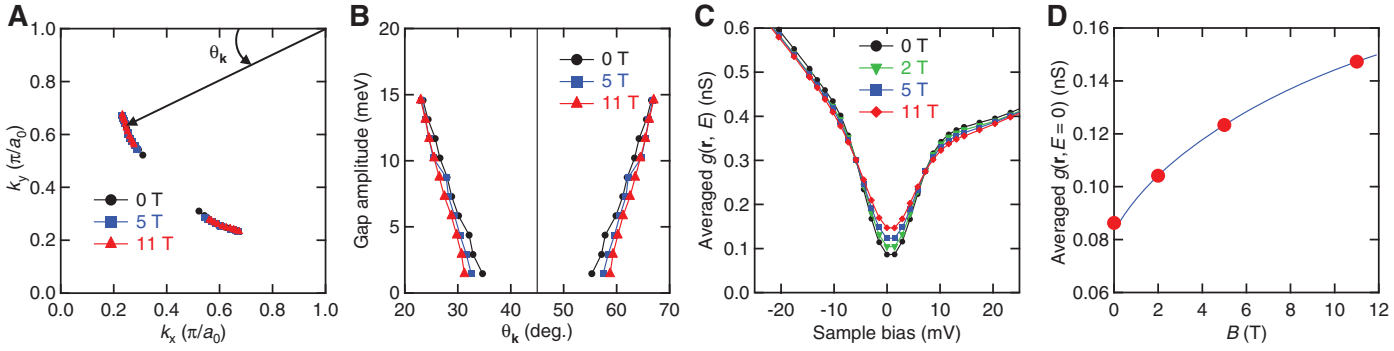


Fig. 4. Magnetic-field B effects on the electronic states of in $Ca_{2-x}Na_xCuO_2Cl_2$ ($x \sim 0.14$ and $T_c \sim 28$ K). (**A**) Loci of octet ends of constant energy contours at different B representing the underlying Fermi surface. Four independent $\mathbf{q}_4(E) = [\pm 2\mathbf{k}_x(E), 2\mathbf{k}_y(E)]$, $[2\mathbf{k}_y(E), \pm 2\mathbf{k}_x(E)]$ were used for analysis. No measurable B-induced change is found in the Fermi surface. (**B**) B-induced renormalization of the d-wave SC gap dispersion. B enlarges the apparent gapless region around the gap node, whereas the dispersion at higher energy is relatively insensitive to B. (**C**) Tunneling spectra averaged over the field of view. Gaplike feature below about 10 meV gets shallower and DOS at Fermi energy ($E = 0$) increases with increasing B. Spectrum at 2 T was averaged in the slightly different field of view for other fields. (**D**) The B dependence of spatially averaged $g(\mathbf{r}, E = 0)$. Blue line denotes $B\log B$ behavior expected in a dirty d-wave superconductor.

conductor (Fig. 4D). These results may provide a spectroscopic basis for the field-induced DOS observed by specific heat (*22*) and nuclear magnetic resonance (*23*) measurements and are consistent with the Volovik effect (*15*), which predicts field-induced gapless excitations around the SC gap nodes in **k** space and outside the vortex core in real space.

The detection of a *d*-wave coherence factor of a high-T_c cuprate using vortices as controllable quasi-particle scattering centers establishes that vortices selectively activate those quasi-particle scattering channels that preserve the sign of the SC gap in **k** space. Moreover, our method provides a simple phase-sensitive probe of gap anisotropy that can be applied to other superconductors with other forms of anisotropic gap, such as *p*-wave and extended *s*-wave superconductors. Another variant on this method is to examine the coherence factors for scattering off conventional impurities at temperatures above T_c: This approach may provide a viable way to probe the nature of the order that develops in the pseudogap normal state (*24*, *25*). Lastly, we note that Fourier-transform SI-STM is currently the only method to study the evolution of **k**-dependent electronic states as a function of magnetic field and, in this respect, offers a useful tool for the study of a wide range of field-induced quantum phenomena (*26*).

References and Notes

1. D. R. Tilley, J. Tilley, *Superfluidity and Superconductivity* (IOP, Bristol, UK, ed. 3, 1990).
2. J. R. Schrieffer, *Theory of Superconductivity* (Perseus, Reading, MA, 1999).
3. L. C. Hebel, C. P. Slichter, *Phys. Rev.* **113**, 1504 (1959).
4. N. Bulut, D. J. Scalapino, *Phys. Rev. B* **45**, 2371 (1992).
5. J. E. Hoffman *et al*., *Science* **297**, 1148 (2002).
6. K. McElroy *et al*., *Nature* **422**, 592 (2003).
7. T. Hanaguri *et al*., *Nature Phys.* **3**, 865 (2007).
8. Y. Kohsaka *et al*., *Nature* **454**, 1072 (2008).
9. Materials and methods are available as supporting online material on *Science* Online.
10. Q. Wang, D.-H. Lee, *Phys. Rev. B* **67**, 020511 (2003).
11. R. S. Markiewicz, *Phys. Rev. B* **69**, 214517 (2004).
12. T. S. Nunner, W. Chen, B. M. Andersen, A. Melikyan, P. J. Hirschfeld, *Phys. Rev. B* **73**, 104511 (2006).
13. J. E. Hoffman, thesis, University of California, Berkeley, CA (2003).
14. T. Pereg-Barnea, M. Franz, *Phys. Rev. B* **78**, 020509(R) (2008).
15. G. E. Volovik, *JETP Lett.* **58**, 469 (1993).
16. T. Hanaguri *et al*., *Nature* **430**, 1001 (2004).
17. K. Fujita *et al*., *Phys. Rev. B* **78**, 054510 (2008).
18. J. E. Hoffman *et al*., *Science* **295**, 466 (2002).
19. K. Matsuba *et al*., *J. Phys. Soc. Jpn.* **76**, 063704 (2007).
20. The presence of residual DOS at $B = 0$ indicates that a pair-breaking effect is not negligible in the present sample, although the microscopic nature of pair-breaking remains to be determined. We postulate that the QPI peaks in $|Z(\mathbf{q}, E, B)|$ track the quasi-particle dispersion even in the presence of such pair-breaking. This assumption may be supported by the fact that *B*-induced change of quasi-particle dispersion determined by QPI is consistent with the increase of DOS in *B*.
21. C. Kübert, P. J. Hirschfeld, *Solid State Commun.* **105**, 459 (1998).
22. K. A. Moler *et al*., *Phys. Rev. Lett.* **73**, 2744 (1994).
23. G.-q. Zheng *et al*., *Phys. Rev. Lett.* **88**, 077003 (2002).
24. T. Pereg-Barnea, M. Franz, *Phys. Rev. B* **68**, 180506(R) (2003).
25. T. Timusk, B. Statt, *Rep. Prog. Phys.* **62**, 61 (1999).
26. K. Iwaya *et al*., *Phys. Rev. Lett.* **99**, 057208 (2007).

27. The authors thank J. C. Davis, H. Eisaki, M. Franz, C. -M. Ho, K. Machida, T. Pereg-Barnea, and P. Wahl for valuable discussions and comments. T.H., M.T., and H.T. are supported by Grant-in-Aid for Scientific Research from the Ministry of Education, Culture, Sports, Science, and Technology of Japan.

Supporting Online Material

www.sciencemag.org/cgi/content/full/1166138/DC1
Materials and Methods
Figs. S1 and S2
Table S1
References

18 September 2008; accepted 19 December 2008
Published online 22 January 2009;
10.1126/science.1166138
Include this information when citing this paper.

Polydnaviruses of Braconid Wasps Derive from an Ancestral Nudivirus

Annie Bézier,[1] Marc Annaheim,[2] Juline Herbinière,[1] Christoph Wetterwald,[2] Gabor Gyapay,[4] Sylvie Bernard-Samain,[4] Patrick Wincker,[4] Isabel Roditi,[2] Manfred Heller,[3] Maya Belghazi,[5] Rita Pfister-Wilhem,[2] Georges Periquet,[1] Catherine Dupuy,[1] Elisabeth Huguet,[1] Anne-Nathalie Volkoff,[6] Beatrice Lanzrein,[2] Jean-Michel Drezen[1]*

Many species of parasitoid wasps inject polydnavirus particles in order to manipulate host defenses and development. Because the DNA packaged in these particles encodes almost no viral structural proteins, their relation to viruses has been debated. Characterization of complementary DNAs derived from braconid wasp ovaries identified genes encoding subunits of a viral RNA polymerase and structural components of polydnavirus particles related most closely to those of nudiviruses—a sister group of baculoviruses. The conservation of this viral machinery in different braconid wasp lineages sharing polydnaviruses suggests that parasitoid wasps incorporated a nudivirus-related genome into their own genetic material. We found that the nudiviral genes themselves are no longer packaged but are actively transcribed and produce particles used to deliver genes essential for successful parasitism in lepidopteran hosts.

Comparative genomic studies have highlighted the role of symbiotic associations in evolution (*1*). Polydnaviruses (PDVs) are virus-like particles associated with wasp species that parasitize lepidopteran larvae. PDV particles are injected along with the eggs of the wasp into the lepidopteran larvae (or eggs) and express proteins that interfere with host immune defenses, development, and physiology; this interference enables wasp larvae to survive and develop within the host (*2*). Viral particle production occurs exclusively in a specialized region of the wasp ovaries (the calyx), and the vertically transmitted virus does not initiate particle production in the infected host tissues (*3*). The viral genome packaged in the particles is composed of multiple double-stranded DNA (dsDNA) circles, and it is surprising that it encodes almost no viral structural proteins, although it harbors immunosuppressive genes that are expressed in the host and are essential for successful parasitism (*4*, *5*) (see PDV description at www.ictvonline.org). Because of this lack of genes coding for structural proteins, it has been debated whether PDVs are of viral origin or a "genetic secretion" of the wasp (*6*, *7*).

PDVs are classified as either bracoviruses or ichnoviruses, when associated with braconid or ichneumonid wasps, respectively. Detailed phylogenetic studies have shown that the bracovirus-associated wasps form a monophyletic group known as the microgastroid complex (*8*), and it has been hypothesized that there has been a single integration event of a viral genome, as a provirus, in the microgastroid lineage. This predicts that vertically transmitted viral DNA may have been maintained because of its contribution to successful parasitism and that PDVs have contributed to the diversification of the microgastroid complex of at least 17,500 species (*8*).

The sequence of the DNA packaged in *Cotesia congregata* bracovirus (CcBV) comprises 560 kb, organized in 30 circles of dsDNA (*4*), and encodes several products functionally resembling virulence factors used by parasites or bacterial pathogens of vertebrates (*9*–*11*). However, the origin of the protein components of PDV particles remains unknown. The few CcBV sequences with significant similarity to

[1]Institut de Recherche sur la Biologie de l'Insecte, CNRS UMR 6035, Université François Rabelais, Parc de Grandmont, 37200 Tours, France. [2]Institute of Cell Biology, University of Bern, Baltzerstrasse 4, CH-3012 Bern, Switzerland. [3]Department of Clinical Research, Murtenstrasse 35, CH-3010 Bern, University of Bern, Switzerland. [4]Commissariat à l'Énergie Atomique (CEA), Genoscope (Centre National de Séquençage), 2 rue Gaston Crémieux, CP 5706, 91057 Evry Cedex, France. [5]Institut Federatif de Recherche J. Roche (IFR11), Inserm, Centre d'Analyse Protéomique de Marseille, Faculté de Médecine, 51 boulevard Pierre Dramard, 13916 Marseille Cedex 20, France. [6]Biologie Intégrative et Virologie des Insectes, INRA UMR 1231, Université Montpellier II, Place Eugène Bataillon, 34095 Montpellier Cedex 5, France.

*To whom correspondence should be addressed. E-mail: drezen@univ-tours.fr

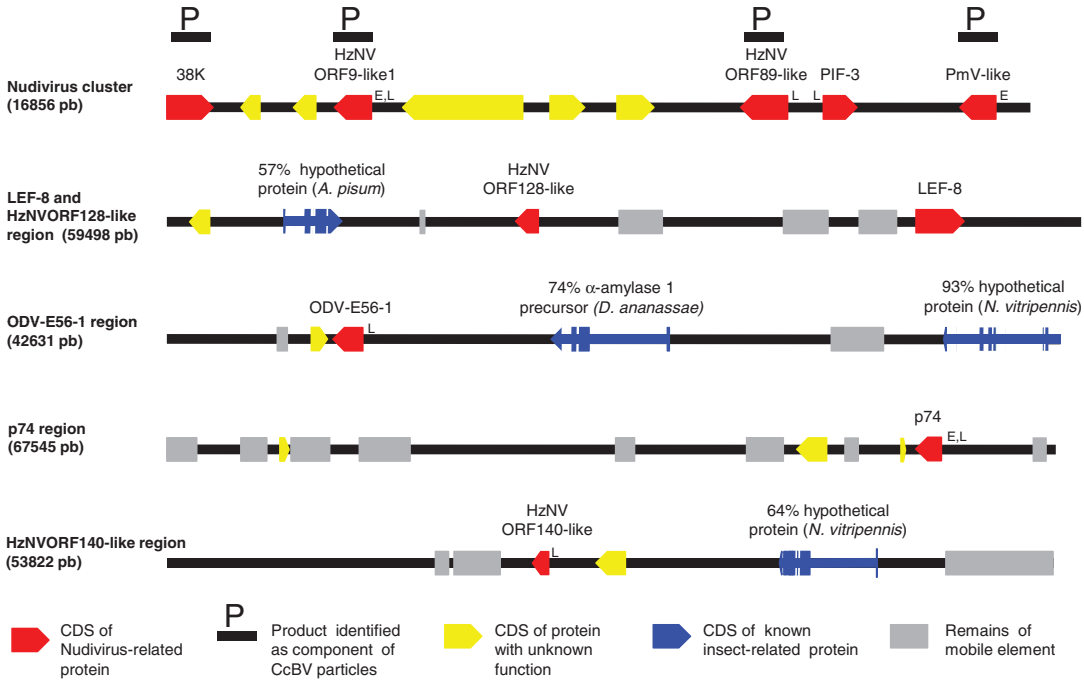


Fig. 1. Organization of nudivirus-related genes in the wasp genome (*Cotesia congregata*). Five genes (*38K*, *HzNVorf9-like1*, *HzNVorf89-like*, *pif-3*, and *PmV-like*) are organized as a cluster that is likely to constitute a remnant of the virus integrated into the ancestral wasp genome. Two genes (*HzNVorf128-like* and *lef-8)* are located within the same 60-kb chromosomal region; other genes (*odv-e56-1*, *p74*, and *HzNVorf140-like)* are dispersed and located in regions containing *Cotesia congregata* putative homologs of insect genes (in blue) and/or remnants of mobile elements (in gray). For nudivirus-related genes, the presence of sequences found in promoters of baculovirus genes transcribed by the cellular RNA polymerase (E) or baculovirus RNA polymerase (L) is indicated (*19*). E indicates the presence of a TATAA sequence with a CA(T/G)T or a CGTGC transcription start site 20 to 40 nucleotides downstream. L indicates the presence of a (A/T/G)TAAG motif within 300 nucleotides upstream of the translation start codon. CDS, coding sequence; ORF, open reading frame. *Acyrthosiphon pisum* (pea aphid), *Drosophila ananassae* (fruit fly), and *Nasonia vitripennis* (parasitoid wasp).

viral genes correspond mostly to remnants of mobile elements inserted randomly into the chromosomal form of CcBV and are thus not informative about the origin of bracoviruses (*4*). A viral genome without genes involved in particle production seems paradoxical. However, it may be that the genes involved in particle morphogenesis have lost the ability to be incorporated into the particles injected into the host, because the particles are exclusively produced in wasp ovaries and PDVs are only transmitted in their integrated chromosomal form. In accordance with this theory, a gene encoding an ichnovirus structural protein was identified in the wasp genome (*12*).

Regardless of where the genes are located, structural proteins of the particles are expressed in virus-producing cells. We therefore searched for virus-related genes expressed in wasp pupal ovaries at times when virus particle production is highest. We studied the braconid wasps *Chelonus inanitus* (Cheloninae) and *Cotesia congregata* (Microgastrinae) belonging to different bracovirus-associated subfamilies (*13*). We also sequenced cDNAs from the wasp *Hyposoter didymator,* associated with an ichnovirus but belonging to the same superfamily (Ichneumonoidea). We sequenced 5000 expressed sequence tags (ESTs) from the ovaries of each species and identified a set of nudivirus-related genes expressed by the braconid wasp ovaries (Table 1, A and B). The predicted products of these genes are related to 22 nudivirus genes (e values from $1e^{-65}$ to 1.1, see Table 1), 13 of which are conserved between nudiviruses and baculoviruses. In ichneumonid wasp ovaries, we detected mRNAs coding for characterized ichnovirus proteins but no genes related to known viruses. This suggests that the viral machinery of ichnovirus production may differ from that of bracoviruses. The recent discovery of a possible new lineage of insect viruses, represented by the *Glossina pallidipes* salivary gland hypertrophy virus (*14*), illustrates that the diver-

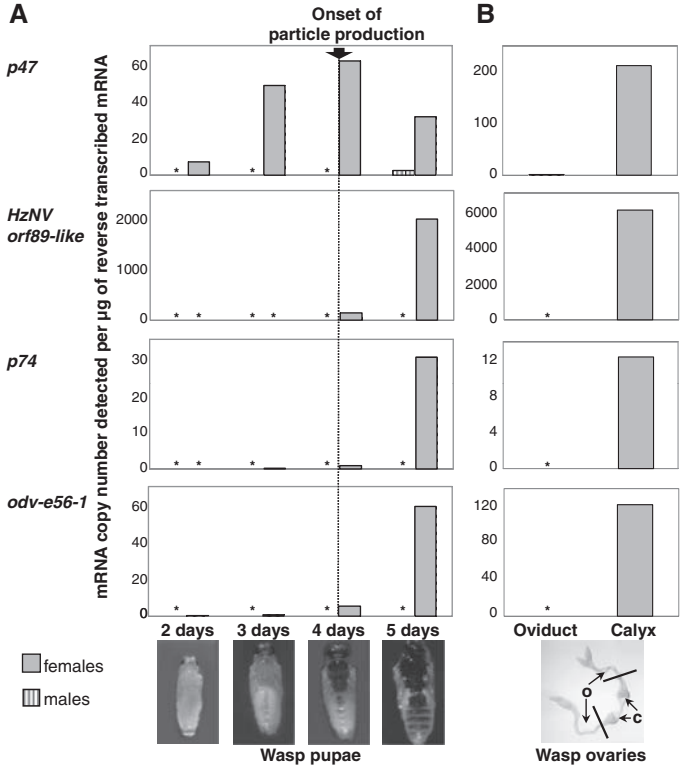


Fig. 2. Nudivirus-related gene expression correlates temporally with bracovirus particle production and occurs in the same tissue. (**A**) Real-time RT-PCR indicates that three nudivirus-related genes (*HzNVorf89-like* coding for a structural component of CcBV particles, *p74*, and *odv-e56-1*) are induced in female pupae from day 4, coincident with the initiation of particle production as detected by transmission electron microscopy (TEM) and PCR (*26*). The induction of *p47*, coding for one of the viral RNA polymerase subunits, occurs earlier, which would be expected if the viral RNA polymerase controls the expression of some nudivirus-related genes. Although baculovirus regulatory sequences are present in some promoters (see Fig. 1), they appear to be too short and, therefore, would lack the specificity necessary to selectively express chromosomally integrated genes. The four nudivirus-related genes are specifically expressed in the calyx region of the ovaries where bracovirus particles are produced, but not in the oviducts (**B**) nor in males (A). Asterisk indicates expression too low to be indicated. On wasp ovaries: o, oviduct; c, calyx.

sity of insect viruses is not completely known. This could explain our inability to identify viral genes in ichneumonid sequences. Few nudivirus genomes have been sequenced, and little is known about the function of their proteins, except by inference from studies on baculoviruses. Nudiviruses share half of the core set of essential genes conserved among baculoviruses (*15*, *16*). The nudivirus-related gene products we identified in the braconid ovaries have been associated with different functions in baculoviruses (*17*–*20*) (Table 1) including subunits of the viral RNA polymerase (LEF-4, LEF-8, and p47), proteins involved in particle assembly and packaging [38K (Ac98), VLF-1, and VP91], and envelope proteins of particles released after the death of infected insects (p74, PIF-1, PIF-2, and ODV-E56). In addition, we characterized several genes coding for variants

Table 1. Nudivirus-related genes expressed in braconid wasp ovaries and the presence of their products in bracovirus particles. Genes characterized in both *Cotesia congregata* (A) and *Chelonus inanitus* (B) are shaded in gray. Gene products identified as components of purified CcBV and CiBV particles by proteomic analyses are shown in red. HzNV-1, *Heliothis zea* nudivirus-1; GbNV, *Gryllus bimaculatus* nudivirus; OrNV, *Oryctes rhinoceros* nudivirus; PmV, *Penaeus monodon* virus (shrimp nudivirus); AdorNPV, *Adoxophyes orana* nucleopolyhedrovirus. Dash indicates a transcript not found among the cDNAs sequenced. "No hit" indicates a protein having no blast hit but similar to a nudivirus-related protein of the other species. See table S1 for more details.

A

Gene type and function	Genes expressed in wasp ovaries producing bracovirus particles	Best blast: Nudivirus/ baculovirus	Best blast: Acc. no.	*C. congregata* protein: Name	Acc. no.	e value	Similarity (%)
Nudivirus/baculovirus core genes							
Transcription of viral genes	*RNA polymerase subunit lef-4*	HzNV-1	22788805	–	–	–	–
	RNA polymerase subunit lef-8	HzNV-1	22788797	CcLEF-8	CAR31572.1	$1e^{-60}$	51
	RNA polymerase subunit p47	HzNV-1	22788782	Ccp47	CAR31573.1	$2e^{-09}$	43
	Transcript. initiation factor lef-5	AdorNPV	209978847	CcLEF-5*	CAT00573.1	0.008	54
Packaging and assembly	*38K (Ac98)*	HzNV-1	22788860	Cc38K	CAR31574.1	$1e^{-21}$	57
	Very late factor-1 Vlf-1	PmV	160432005	–	–	–	–
	vp91	GbNV	134303400	–	–	–	–
ODV envelope components	*Per os infectivity factor p74*	HzNV-1	22788862	Ccp74	CAR31575.1	$2e^{-54}$	47
	Per os infectivity factor pif-1	HzNV-1	22788762	–	–	–	–
	Per os infectivity factor pif-2	HzNV-1	22788829	–	–	–	–
	Per os infectivity factor pif-3	HzNV-1	22788795	CcPIF-3	CAR31576.1	$3e^{-15}$	51
	odv-e56	HzNV-1	22788783	CcODV-E56-1	CAR31577.1	$3e^{-12}$	41
				CcODV-E56-2	CAR31578.1	$8e^{-28}$	48
Unknown	*19 kDa (Ac96)*	GbNV	134303485	Cc19kDa	CAR31579.1	$7e^{-10}$	51
Nudivirus/lepidopteran baculovirus core gene							
ODV envelope component	*odv-e66*	OrNV	108515114	CcODV-E66-1	CAR31580.1	No hit	
				CcODV-E66-2	CAR31581.1	No hit	
Nudivirus-specific genes							
Unknown	*HzNVorf9*	HzNV-1	22788719	CcHzNVORF9-like1	CAR31582.1	0.094	41
				CcHzNVORF9-like2	CAR31583.1	0.001	42
	HzNVorf64	HzNV-1	22788771	CcHzNVORF64-like	CAR31584.1	$1e^{-15}$	45
	HzNVorf89	HzNV-1	22788796	CcHzNVORF89-like	CAR31585.1	No hit	
	HzNVorf106	HzNV-1	22788813	CcHzNVORF106-like	CAR31586.1	No hit	
	HzNVorf128	HzNV-1	22788834	CcHzNVORF128-like	CAR31587.1	$1e^{-05}$	43
	HzNVorf140	HzNV-1	22788845	CcHzNVORF140-like	CAR31588.1	$8e^{-09}$	45
	HzNVorf144	PmV	160432004	–	–	–	–
	PmV hypothetical protein	PmV	134260388	CcPmV-like	CAR31589.1	0.003	44

B

Gene type and function	Genes expressed in wasp ovaries producing bracovirus particles	*C. inanitus* protein: Name	Acc. no.	e value	Similarity (%)	Cc/Ci protein similarity (%)
Nudivirus/baculovirus core genes						
Transcription of viral genes	*RNA polymerase subunit lef-4*	CiLEF-4	CAR40187.1	$3e^{-04}$	39	–
	RNA polymerase subunit lef-8	–	–	–	–	–
	RNA polymerase subunit p47	–	–	–	–	–
	Transcript. initiation factor lef-5	–	–	–	–	–
Packaging and assembly	*38K (Ac98)*	Ci38K	CAR40188.1	$6e^{-18}$	58	66
	Very late factor-1 Vlf-1	CiVLF-1	CAR40190.1	$2e^{-08}$	50	–
	vp91	CiVP91	CAR40191.1	$8e^{-04}$	38	–
ODV envelope components	*Per os infectivity factor p74*	Cip74	CAR40192.1	$5e^{-65}$	46	51
	Per os infectivity factor pif-1	CiPIF-1	CAR40193.1	$3e^{-31}$	45	–
	Per os infectivity factor pif-2	CiPIF-2	CAR40194.1	$1e^{-39}$	48	–
	Per os infectivity factor pif-3	–	–	–	–	–
	odv-e56	CiODV-E56	CAR40195.1	$7e^{-05}$	39	44
						41
Unknown	*19 kDa (Ac96)*	Ci19kDa	CAR40196.1	$2e^{-12}$	53	55
Nudivirus/lepidopteran baculovirus core gene						
ODV envelope component	*odv-e66*	CiODV-E66	CAR40197.1	$8e^{-12}$	41	54
						42
Nudivirus-specific genes						
Unknown	*HzNVorf9*	CiHzNVORF9-like1	CAR40198.1	0.009	41	82
		CiHzNVORF9-like2	CAR40199.1	No hit		74
	HzNVorf64	–	–	–	–	–
	HzNVorf89	CiHzNVORF89-like	CAR40200.1	0.094	40	60
	HzNVorf106	CiHzNVORF106-like	CAR40201.1	1.1	46	69
	HzNVorf128	CiHzNVORF128-like	CAR40202.1	$1e^{-04}$	48	81
	HzNVorf140	CiHzNVORF140-like	CAR40203.1	$9e^{-10}$	48	56
	HzNVorf144	CiHzNVORF144-like	CAR40240.1	$6e^{-20}$	51	–
	PmV hypothetical protein	CiPmV-like	CAR40204.1	0.064	42	69

*CcLEF-5 is slightly closer to AdorNPV than to HzNV-1 LEF-5.

of ODV-E66, a conserved envelope protein encoded by lepidopteran baculoviruses and a nudivirus (Table 1). We found no viral genes involved in DNA replication, which suggests that these mRNAs are rare or that host genes are involved.

The nudivirus *Heliothis zea* nudivirus-1 (HzNV-1) was originally identified as a persistent viral infection in a cell line from adult ovarian tissue of the moth *Heliothis zea* (*21*) and is closely related to HzNV-2, a sexually transmitted virus replicating in insect reproductive tissues (*22*). The genomes of HzNV-1; of *Gryllus bimaculatus* nudivirus (GbNV), a virus infecting the fat body of crickets; and of *Oryctes rhinoceros* nudivirus (OrNV), a biological control agent used to protect palm trees against the rhinoceros beetle have been sequenced (*15*, *23*, *24*). Partial data are also available for a shrimp nudivirus, designated here as PmV (GenBank accession number: 160432003). In braconid wasp ovaries, we identified eight ESTs similar to nudivirus-specific genes; two are shared between several nudiviruses (*HzNVorf64-like* and *HzNVorf144-like*); five have been found only in HzNV-1 (*HzNVorf9-like*, *HzNVorf89-like*, *HzNVorf106-like*, *HzNVorf128-like,* and *HzNVorf140-like*); and one is specific to the shrimp nudivirus (*PmV-like*). All braconid viral products have less than 60% similarity with nudiviral proteins (Table 1, A and B, e values $6e^{-20}$ to 1.1), which is not surprising, considering that the ancestral nudivirus is hypothesized to have integrated into the wasp genome 100 million years ago (*25*).

The prevalence of nudiviruses in hymenoptera is unknown, but the fact that nudivirus transcripts were not detected in the ichneumonid species indicates that parasitic wasps are not universally infected. The isolation of the nudivirus-related *p74* genes from eight *Cotesia* species using wasps collected in different parts of the world (fig. S1) indicated that this gene is stably associated with these wasps and does not belong to a pathogenic virus present in our laboratory strain. We could also show by sequencing *Cotesia congregata* large genomic DNA regions that at least 10 nudivirus-related genes are chromosomally integrated. Five genes were found to be clustered (*38K*, *HzNVorf9-like1*, *HzNVorf89-like*, *pif-3*, and *PmV-like*); others were dispersed (*HzNVorf128-like*, *lef-8, odv-e56-1*, *p74*, and *HzNVorf140-like*) and flanked by wasp genomic DNA (Fig. 1). Quantitative real-time reverse transcriptase polymerase chain reaction (RT-PCR) showed that the expression of nudivirus-related genes increases in female wasp pupae (Fig. 2), coinciding with the onset of bracovirus replication (*26*). Moreover, expression in the ovaries is confined to the calyx region where particles are produced (Fig. 2).

To confirm that the nudivirus-related genes produce bracoviruses, we analyzed proteins from purified particles by tandem mass spectrometry (MS/MS). We identified the products of six genes as components of the particles in both CcBV and CiBV (*Chelonus inanitus* bracovirus); the protein ODV-E66 in CcBV particles; and the proteins ODV-E56, p74, PIF-1, and PIF-2 in CiBV (indicated in red in Table 1, see table S2 for detailed results). Altogether, 20 different nudivirus-related gene products were identified as components of bracovirus particles, which demonstrated a functional link between the nudiviral machinery and bracoviruses.

Chelonus inanitus and *Cotesia congregata* belong to the most distantly related subfamilies of bracovirus-associated wasps (*25*), which suggests that the nudiviral machinery is present in all these species. Accordingly, we amplified the most conserved nudivirus-related genes (*HzNVorf9-like1* and *HzNVorf128-like*) from most wasp subfamilies of the microgastroid complex (Fig. 3). We conclude that these genes were present before the radiation of the complex and originated from a virus integrated in a chromosome of the ancestral wasp. The cluster

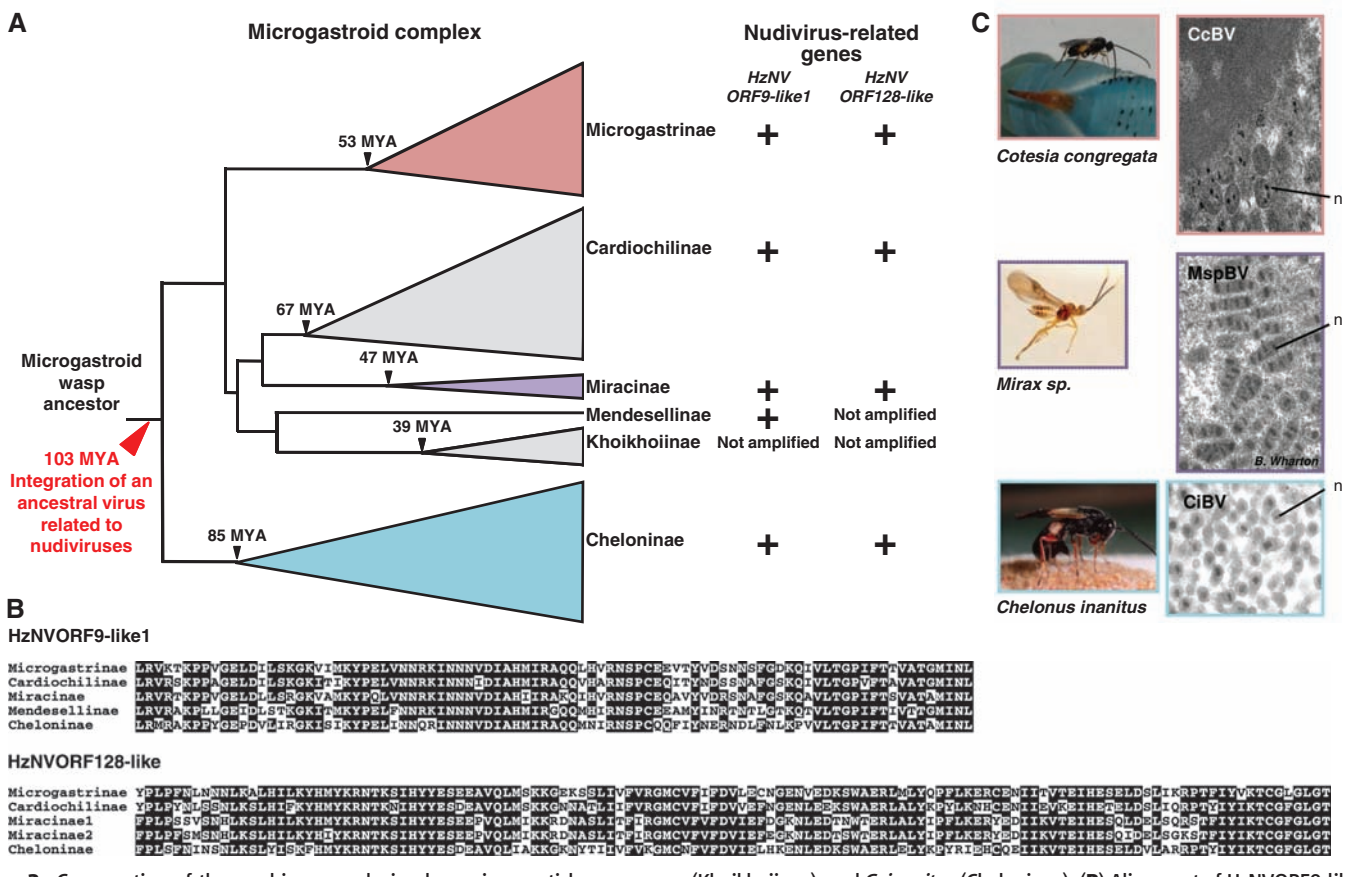


Fig. 3. Conservation of the machinery producing bracovirus particles among microgastroid wasps. The tree and dates of radiations are taken from (*25*). (**A**) Nudivirus-related genes amplified with DNA from species belonging to different subfamilies: *C. congregata* (Microgastrinae), *Toxoneuron nigriceps* (Cardiochilinae), *Mirax* sp. (Miracinae), *Epsilogaster* sp. (Mendesellinae), *Sania* sp. (Khoikhoiinae), and *C. inanitus* (Cheloninae). (**B**) Alignment of HzNVORF9-like1 and HzNVORF128-like proteins deduced from the amplified sequences. (**C**) Bracovirus particles visualized by TEM from *C. congregata*, *Mirax* sp., and *C. inanitus*. n, Nucleocapsid. The particles contain one (CiBV) or several nucleocapsids (CcBV and MspBV), dispersed (CcBV) or organized (MspBV).

of genes present in *Cotesia congregata* could constitute the remnants of the genome of this ancestral virus. The level of similarity between *HzNVorf9-like1* and *HzNVorf128-like* products from *Chelonus inanitus* and *Cotesia congregata* reaches 80% and thus indicates a strong conservation of their functions (Table 1). Some proteins show a lower conservation (for example for p74, 46% similarity), which suggests these sequences may be involved in more specific interactions with the host and have had to evolve more rapidly due to selective pressures associated with infection.

The overall conservation of the nudiviral machinery encoded by the wasps contrasts sharply with the lack of similarity found between the DNA enclosed in CiBV and CcBV particles. No common genes were found between CiBV and other bracovirus genomes (*27*); only sequences involved in the production of the circular dsDNAs of the particles (excision sequences) were conserved (*28*, *29*). Furthermore, packaged bracovirus genomes do not contain any nudivirus-related genes. We hypothesize that shortly after initial integration of the nudivirus ancestor, viral DNA might have been replaced by wasp DNA in the particles (possibly by translocation of sequences allowing excision and encapsidation) and that most genes promoting parasitism were acquired later and independently in bracovirus-associated wasps.

It is well documented that genes of viral origin are used by eukaryotes to ensure physiological functions, such as the syncytins involved in trophoblast differentiation, which originated from retroviral envelope proteins independently acquired by primates and mice (*30*). However, the bracovirus-wasp associations represent the only example, so far, of the incorporation of genes encoding a complex viral machinery that allows its eukaryotic host to transfer and express heterologous genes in target organisms. In this regard, unraveling bracovirus particle assembly could contribute to the design of new vectors for gene therapy.

References and Notes

1. N. A. Moran, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 8627 (2007).
2. B. A. Webb, M. R. Strand, in *Comprehensive Molecular Insect Science*, L. I. Gilbert, K. Iatrou, S. S. Gill, Eds. (Elsevier, Amsterdam, 2005), vol. 6, pp. 323–360.
3. S. Wyder, F. Blank, B. Lanzrein, *J. Insect Physiol.* **49**, 491 (2003).
4. E. Espagne *et al.*, *Science* **306**, 286 (2004).
5. J. A. Kroemer, B. A. Webb, *Annu. Rev. Entomol.* **49**, 431 (2004).
6. J. B. Whitfield, S. Asgari, *J. Insect Physiol.* **49**, 397 (2003).
7. B. A. Federici, Y. Bigot, *J. Insect Physiol.* **49**, 419 (2003).
8. J. B. Whitfield, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 7508 (2002).
9. H. Thoetkiattikul, M. H. Beck, M. R. Strand, *Proc. Natl. Acad. Sci. U.S.A.* **102**, 11426 (2005).
10. B. Provost *et al.*, *J. Virol.* **78**, 13090 (2004).
11. E. Espagne *et al.*, *J. Virol.* **79**, 9765 (2005).
12. L. Deng, D. B. Stoltz, B. A. Webb, *Virology* **269**, 440 (2000).
13. Materials and methods are available as supporting material on *Science* Online.
14. A. M. Abd-Alla *et al.*, *J. Virol.* **82**, 4595 (2008).
15. Y. Wang, R. G. Kleespies, A. M. Huger, J. A. Jehle, *J. Virol.* **81**, 5395 (2007).
16. E. A. Herniou, J. A. Olszewski, J. S. Cory, D. R. O'Reilly, *Annu. Rev. Entomol.* **48**, 211 (2003).
17. A. L. Vanarsdall, V. S. Mikhailov, G. F. Rohrmann, *Curr. Drug Targets* **8**, 1096 (2007).
18. M. M. van Oers, J. M. Vlak, *Curr. Drug Targets* **8**, 1051 (2007).
19. A. L. Passarelli, L. A. Guarino, *Curr. Drug Targets* **8**, 1103 (2007).
20. S. C. Braunagel, M. D. Summers, *Curr. Drug Targets* **8**, 1084 (2007).
21. R. R. Granados, T. Nguyen, B. Cato, *Intervirology* **10**, 309 (1978).
22. A. K. Raina *et al.*, *J. Invertebr. Pathol.* **76**, 6 (2000).
23. C. H. Cheng *et al.*, *J. Virol.* **76**, 9024 (2002).
24. Y. Wang, R. G. Kleespies, M. B. Ramle, J. A. Jehle, *J. Virol. Methods* **152**, 106 (2008).
25. N. Murphy, J. C. Banks, J. B. Whitfield, A. D. Austin, *Mol. Phylogenet. Evol.* **47**, 378 (2008).
26. F. Pasquier-Barre *et al.*, *J. Gen. Virol.* **83**, 2035 (2002).
27. B. Weber, M. Annaheim, B. Lanzrein, *Arch. Insect Biochem. Physiol.* **66**, 9 (2007).
28. M. Annaheim, B. Lanzrein, *J. Gen. Virol.* **88**, 450 (2007).
29. C. A. Desjardins *et al.*, *BMC Microbiol.* **7**, 61 (2007).
30. F. Mallet *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **101**, 1731 (2004).

31. Investigations of *C. congregata* and *H. didymator* were performed as part of the project EVPARASITOID supported by the agencies Agence Nationale de la Recherche (France) and CNRS (GDR 2153, GDR 2157, and IFR 136). *C. inanitus* study was supported by the Swiss National Science Foundation (grants 3100A0-103964 and 116067 to B.L.). We thank E. Herniou for helpful discussions and C. Ménoret and C. Labrousse for insect-rearing. Sequences have been deposited in the EMBL Nucleotide Sequence Database, accession numbers FM201559-80, FM201582-97, and FM877774 for the viral cDNAs and FM212911-15 for the chromosomally integrated form of nudivirus-related genes.

Supporting Online Material

www.sciencemag.org/cgi/content/full/323/5916/926/DC1
Material and Methods
Fig. S1
Tables S1 to S5
References

3 October 2008; accepted 18 December 2008
10.1126/science.1166788

Effects of Genetic Perturbation on Seasonal Life History Plasticity

Amity M. Wilczek,[1] Judith L. Roe,[2] Mary C. Knapp,[2] Martha D. Cooper,[1] Cristina Lopez-Gallego,[1]* Laura J. Martin,[1]† Christopher D. Muir,[1]‡ Sheina Sim,[2]§ Alexis Walker,[1] Jillian Anderson,[1] J. Franklin Egan,[1]‖ Brook T. Moyers,[1]¶ Renee Petipas,[1]# Antonis Giakountis,[3] Erika Charbit,[2] George Coupland,[3] Stephen M. Welch,[2] Johanna Schmitt[1]**

Like many species, the model plant *Arabidopsis thaliana* exhibits multiple different life histories in natural environments. We grew mutants impaired in different signaling pathways in field experiments across the species' native European range in order to dissect the mechanisms underlying this variation. Unexpectedly, mutational loss at loci implicated in the cold requirement for flowering had little effect on life history except in late-summer cohorts. A genetically informed photothermal model of progression toward flowering explained most of the observed variation and predicted an abrupt transition from autumn flowering to spring flowering in late-summer germinants. Environmental signals control the timing of this transition, creating a critical window of acute sensitivity to genetic and climatic change that may be common for seasonally regulated life history traits.

The seasonal timing of critical life history events is often under strong natural selection, requiring plants to integrate and respond appropriately to multiple environmental signals. In the case of flowering time, these include day length, ambient temperature, and winter chilling (vernalization) cues. Genetic mechanisms of flowering response to environmental cues have been well characterized in the model plant *Arabidopsis thaliana* in the laboratory. However, the response of *Arabidopsis* to complex, natural environmental cues is unknown. The native distribution of *Arabidopsis* encompasses climates from the Mediterranean, where it is a winter annual that germinates in late autumn and flowers in late winter or early spring, to northern Scandinavia, where the species germinates in early autumn, overwinters, and flowers in late spring and early summer. In England and northwestern Europe, *Arabidopsis* either is a winter annual or undergoes rapid cycling by germinating in early autumn, spring, or summer, thus flowering without vernalization (*1*, *2*). Thus, in nature the species experiences marked variation in photothermal environments across sites and seasons.

The timing and environmental sensitivity of flowering in *Arabidopsis* are regulated by a network of genes in several converging pathways (*3*–*5*) (fig. S1). Inductive long days are perceived by the photoperiod pathway, which accelerates flowering acting through *GIGANTEA* (*GI*) and *CONSTANS* (*CO*) (*6*, *7*). The gibberellin pathway (*8*) and high temperatures (*9*–*12*) also promote flowering. These signals activate floral integrator genes such as *FLOWERING LOCUS T* (*FT*) (*3*). The ability of these integrators to respond is controlled by repressor genes, notably *FLOWERING*

LOCUS C (*FLC*), which in turn is activated by genes such as *FRIGIDA* (*FRI*) and is repressed by vernalization and by genes in the autonomous pathway such as *LUMINIDEPENDENS* (*LD*) and *FVE* (*13*). Prolonged exposure to cold induces *VERNALIZATION-INSENSITIVE-3* (*VIN3*), which initiates stable epigenetic repression of *FLC* (*13*–*15*). Deficiencies in *FLC* activators such as *FRI* remove the vernalization requirement for flowering and may cause a rapid-cycling life history (*16*, *17*).

To assess the relative contributions of different signaling pathways across environments, we grew mutants impaired in response to specific environmental signals and documented environmental conditions at five field sites spanning the native European climatic range: Oulu, Finland; Norwich, UK; Cologne and Halle, Germany; and Valencia, Spain (figs. S2 and S3) (*18*). Across these sites, we established a total of nine seasonal cohorts timed to coincide with the germination of local natural populations (fig. S2). To compare the relative contribution of the photoperiod pathway to flowering across sites and seasons, we examined the flowering delay in mutants of *CO*, which is activated by long-day signals, and in mutants of *GI*, which promotes flowering both by activating *CO* (*19*) and through a *CO*-independent derepression of *FT* (*20*). Both photoperiod pathway mutants (*co-2* in Ler, *gi-2* in Col) showed initiation of flowering significantly later than in the wild type in all plantings except for those planted at Halle in the fall (Fig. 1 and table S1). These mutants flowered especially late in the summer plantings and the Norwich fall cohort, planted under declining long days (Fig. 1).

To examine the effects of different *FLC* repression levels, and thus vernalization requirements, we compared the *fri* null Col-0 wild-type to Col-*FRI*-Sf2, which contains an introgressed strong allele of the *FLC* activator *FRI* that naturally occurs in the Sf-2 ecotype (*21*). We also examined the effects of the autonomous pathway mutants *ld-1*, *fve-3*, and *fve-4*, which impair repression of *FLC* expression and delay flowering in the absence of vernalization, even in a null *fri* background (*22*). Col-*FRI*-Sf2 and the autonomous pathway mutants flowered at similar times within each of the plantings. Although *FRI* introgression and autonomous pathway mutation were expected to confer constitutive winter-annual life histories on the basis of greenhouse studies (*6*, *21*), these lines displayed delays of only ~10 days when planted in summer cohorts, which lacked vernalization (table S2). Thus, *FLC*-induced repression of flowering can be overridden, possibly as a result of natural temperature fluctuations (fig. S4) (*23*) or light conditions (*21*). Col-*FRI*-Sf2 and the autonomous pathway mutants both demonstrated small but significant delays in bolting in the spring cohorts and displayed an expected winter-annual life history in all autumn plantings (Fig. 1 and table S1). However, the Col wild type behaved as a rapid cycler (its laboratory phenotype) only in Norwich, where it flowered in fall more than 100 days earlier than did Col-*FRI*-Sf2 and Col plants containing autonomous pathway mutations. Unexpectedly, in the remaining autumn cohorts, Col flowered much later than in Norwich, displaying a clear winter-annual life history in Halle. Thus, loss of *FRI* function converts winter annuals to rapid cyclers only under a narrow range of natural environmental conditions.

The *vin3-1* deletion, which is deficient in its ability to respond to vernalization cues, had no significant effect on flowering time in the summer cohorts relative to the vernalization-sensitive background wild type (Col-*FRI*-Sf2) (Fig. 1 and table S1), as expected. In contrast, *vin3-1* mutants flowered later than the control in autumn plantings in a manner that was correlated with exposure to vernalizing temperatures (fig. S3C). Additionally, *vin3-1* mutation significantly delayed flowering in spring cohorts (Fig. 1), where plants were exposed to vernalizing temperatures early in development (fig. S3C). Thus, vernalization signals predominated in the later autumn cohorts, photoperiod predominated in summer plantings, and both substantially contributed to the flowering response of spring-germinating plants.

We created a genetically informed photothermal model of *Arabidopsis* development in which plants transition from nonflowering rosettes to bolting (the first visible sign of floral initiation) once they have accumulated a threshold number of environmentally determined photothermal developmental units—temporal measures that incorporate climatic conditions and are interconvertible with real time [e.g., (*24*)]. Phenology models empirically scale photothermal units so that developmental transitions occur at a common threshold of accumulated units independent of particular environmental conditions. The accumulation of appropriately scaled photothermal units over time will therefore show the developmental progression of plants toward the initiation of reproductive growth (Fig. 2). We extended classical phenological modeling approaches by linking individual scaling factors to the activities of specific genes and their regulators. In our gene network–based model, each genotype accumu-

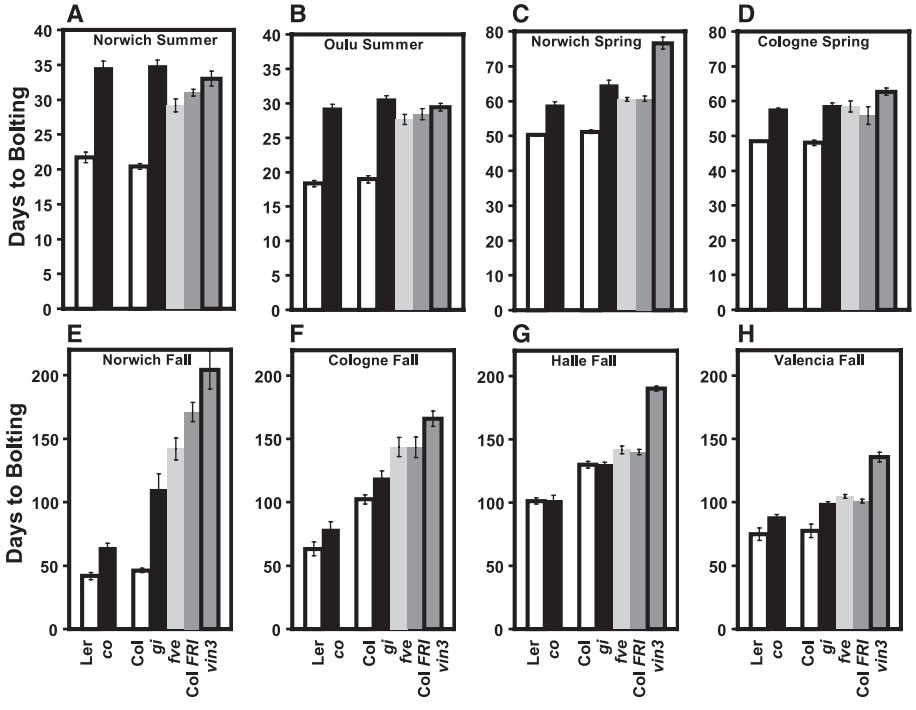


Fig. 1. Sensitivity of flowering-time pathways across sites and seasons. Bars show mean bolting times (±SE) of mutants in different flowering-time pathways along with their wild-type controls. (**A**) Norwich summer 2007 planting; (**B**) Oulu summer 2006; (**C**) Norwich spring 2007; (**D**) Cologne spring 2007; (**E**) Norwich fall 2006; (**F**) Cologne fall 2006; (**G**) Halle fall 2006; (**H**) Valencia fall 2006. Note the difference in scale among locations and seasons.

[1]Department of Ecology and Evolutionary Biology, Brown University, Providence, RI 02912, USA. [2]Department of Agronomy, Kansas State University, Manhattan, KS 66506, USA. [3]Max Planck Institute for Plant Breeding Research, Cologne D-50829, Germany.

*Present address: Biology Institute, Universidad de Antioquia, Medellin AA 1226, Colombia.
†Present address: Department of Natural Resources, Cornell University, Ithaca, NY 14853, USA.
‡Present address: Department of Biology, Indiana University, Bloomington, IN 47405, USA.
§Present address: Department of Biological Sciences, University of Notre Dame, Notre Dame, IN 46556, USA.
||Present address: Department of Crop and Soil Sciences, Pennsylvania State University, University Park, PA 16802, USA.
¶Present address: Department of Botany, University of British Columbia, Vancouver, BC V6T 1Z4, Canada.
#Present address: Department of Biology, University of Vermont, Burlington, VT 05405, USA.
**To whom correspondence should be addressed. E-mail: johanna_schmitt@brown.edu

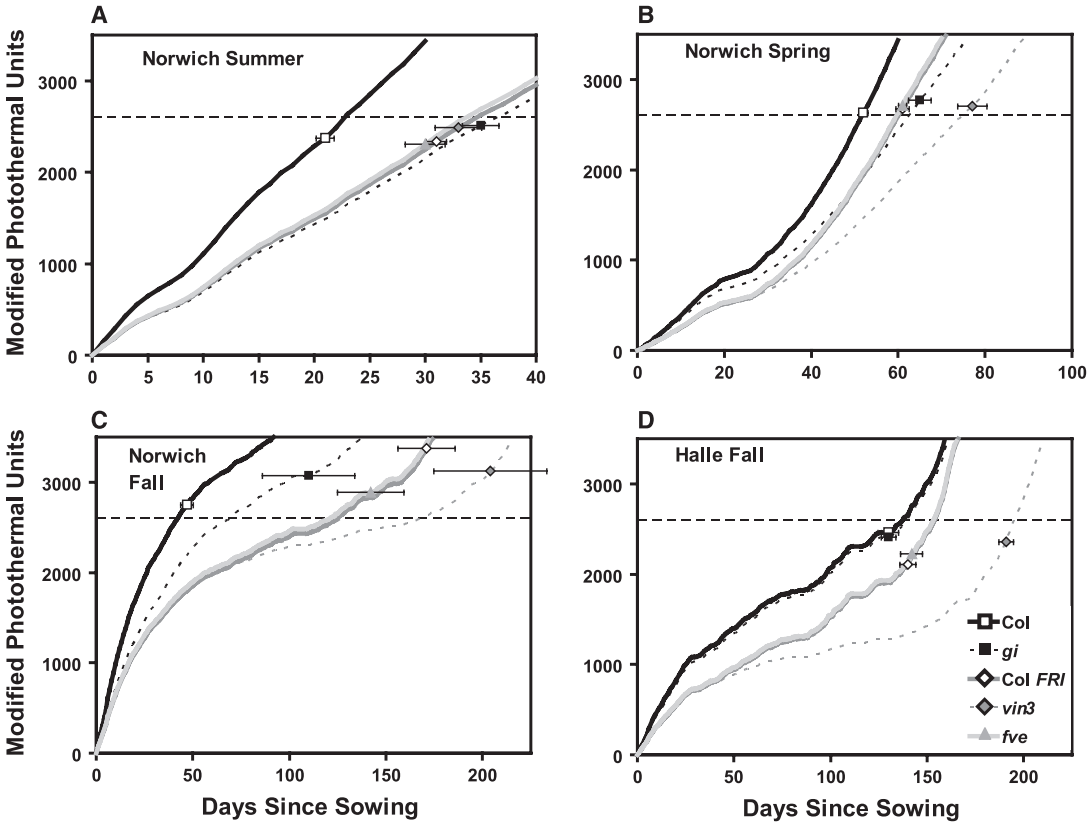


Fig. 2. The accumulation of modified photothermal units from the time of sowing for genotypes in the Col background differing in environmental sensitivity. The model estimated the threshold for bolting at 2604 modified photothermal units (horizontal dashed line), a constant for all environments and genotypes. Each genotype is depicted by a single trace, which shows the accumulation of modified photothermal units in that particular mutant or wild-type background. Observed bolting times (days on the *x* axis, modified photothermal units on the *y* axis) with confidence intervals are shown on each of the accumulation traces. Note differences in scale of *x* axis among seasons. (**A**) Norwich summer 2007; (**B**) Norwich spring 2007; (**C**) Norwich fall 2006; (**D**) Halle fall 2006.

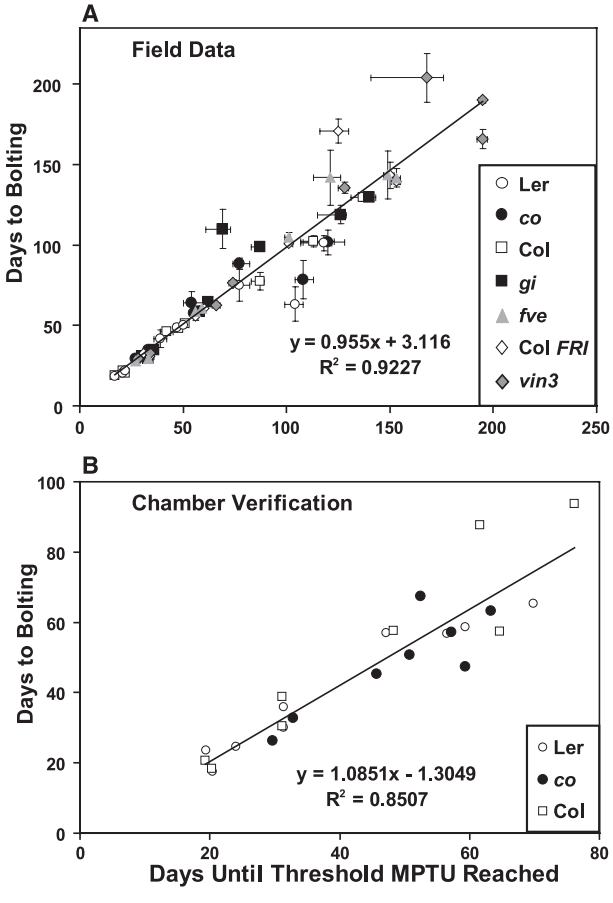


Fig. 3. Observed versus predicted bolting times. (**A**) Field-observed bolting phenotypes of Ler-1, *co-2*, Col-0, *gi-2*, Col-*FRI*-Sf2, *vin3-1* (in Col-*FRI*-Sf2), and *fve-3* across the nine experimental plantings and the predicted phenotypes from the fitted model. Line means with confidence intervals are shown. Confidence intervals for the modified photothermal units (MPTU) were back-calculated to find the day at which the lower and upper 95% confidence limit for the bolting threshold was reached. (**B**) Verification against growth chamber data not used in model construction, showing results from Ler, *co* (both *co*-2 and *co*-3), and Col across nine treatments in two experiments.

lated photothermal units on the basis of day lengths and hourly temperature averages in each experimental planting, modified according to the environmental sensitivity of the line. For example, in our model, both photoperiod mutants and wild-type lines (Col versus *gi-2*, Ler versus *co-2*) were assumed to develop at similar rates under short days.

Photoperiod mutants did not respond to increasing day length, but the modeled developmental rate of wild-type plants accelerated above a critical short day length (Fig. 2) until a maximum rate was reached at a saturating long day length (*25*). Chamber experiments confirmed this piecewise linear response to day length in the Ler accession and identified a critical short day length of ~10 hours. However, flowering in the *co-2* mutant was insensitive to day length (fig. S5). Development was also assumed to accelerate during vernalization because of the repression of *FLC*, with the modeled degree of acceleration varying according to genetic background (Ler versus Col), the allelic state of *FRI* (Col *fri* versus Col *FRI*), and the presence or absence of functional alleles in the autonomous pathway (Col versus *ld* and *fve*); however, *vin3-1* plants did not experience any *FLC* repression from exposure to cold temperatures (Fig. 2). A novel feature of the model, resulting from the scaling of developmental rate with pathway activities, was that all mutants within a given genetic background were predicted to flower at a common threshold of modified photothermal units. Parameter values for the critical day length window, vernalization effective-

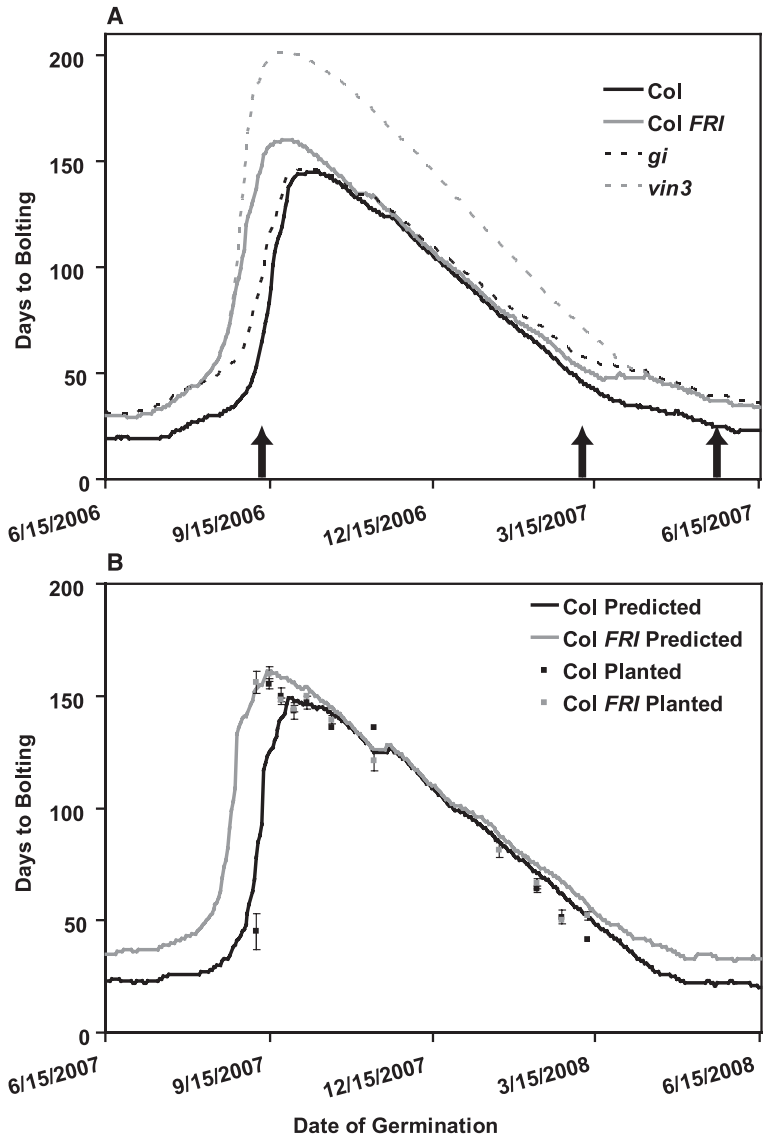


Fig. 4. Timing of bolting as a function of time of germination. (**A**) In Norwich, a narrow window of extreme sensitivity to germination timing is predicted when sown in early autumn. Timing of experimental sowings is shown by arrows. (**B**) Model validation in Cologne with measured bolting data from repeated plantings of *FRI* functional and nonfunctional ecotypes into the field from September 2007 through March 2008. Our model predicted 94% of the variation in bolting time, confirming the expectation that *FRI* results in a life history conversion in early-autumn cohorts.

ness temperature function, time until vernalization saturation, and the degree-hour base temperature were drawn from empirical studies (table S3). Short-day developmental rate and the initial levels of *FLC* repression of development were optimized to the lowest coefficient of variation of modified photothermal units accumulated at observed bolting times across all plantings (*18*).

Our best-fit model explained more than 92% of the observed overall variance in days to bolting (Fig. 3A) and was in agreement with observed reversals in flowering-time order of different genotypes across plantings (e.g., *gi-2* versus Col-*FRI*-Sf2, Cologne and Norwich spring versus fall; Fig. 2 and table S2). The model, with parameter values derived from our field experiment, also accurately predicted the behavior of plants not used in the model's construction and accounted for most of the flowering-time variation of Col, Ler, and *co* mutants under a wide range of controlled conditions [(*11*) and table S4; $R^2 = 0.85$ for the data in both studies, Fig. 3B]. Thus, a modified photothermal model, informed by understanding of pathway function and empirical measurements of gene effects in controlled environments, can be used to quantify the shifting balance and sensitivity of flowering-time pathways under a broad range of climate conditions.

We examined the relative sensitivity of flowering time to parameters representing vernalization sensing, day length sensing, short-day development rate, and baseline *FLC* levels (table S3) and found that flowering time was most sensitive to the initial *FLC* level in both Col and Ler. The short-day development rate, which represents the baseline at which plants proceed toward flowering in the absence of long-day or vernalization cues, had a large effect on flowering time. Indeed, development was slowed by 60 to 74% before vernalization depending on background, and short-day development proceeded only 63 to 69% as quickly as in long days. Similar genetically based sensitivity to photoperiod and/or vernalization cues has been observed in other species and traits (*24*, *26*), in some cases with similar integration of converging pathways (*27*–*29*). The linking of scaling factors to pathway activity depends on network architecture rather than gene identity, which suggests that similar models will prove useful in other species, even where the genes involved do not share a common evolutionary history.

Our model also was able to predict the effects of novel environments on flowering time across genotypes. For example, we used the model to generate predicted reaction norms of flowering date to germination timing in Norwich and Cologne, using actual daily on-site photothermal data to simulate the flowering dates of seeds germinating on successive days over a yearly cycle (Fig. 4). The model predicted a striking, abrupt transition from rapid-cycling to winter-annual life histories for all genotypes germinating in late summer. During this narrow temporal window of sensitivity, all genotypes switched from autumn to spring flowering; the high-*FLC* lines Col-*FRI*-Sf2 and *vin3-1* transitioned between life histories earliest and showed the largest flowering-time response (Fig. 4). A series of repeated plantings into the field in Cologne of Col and Col-*FRI*-Sf2 seedlings throughout the autumn and spring confirmed the predicted abrupt life history transition (Fig. 4B, $R^2 = 0.94$). This rapid rise in bolting time with later germination in autumn will be general across many climates, as it is driven by common patterns of seasonal periodicity of light and temperature cues (fig. S6). Thus, the impact of genetic perturbation on flowering time and life history expression under natural conditions is acutely sensitive to germination timing.

Our results suggest that *A. thaliana* ecotypes cannot simply be divided into two discrete classes of winter-annual and rapid-cycling genotypes. Rather, most ecotypes may be capable of both life histories but vary in the sensitivity and timing of the rapid transition between them. Consequently, natural variation in flowering pathways may have the greatest phenotypic expression and exposure to natural selection in climates that permit late-summer germination. Although the current model was developed for *A. thaliana*, the underlying principle of linking genetic architecture to environmental sensitivity is broadly applicable across plant species and seasonal traits. Our ability to predict the consequences of genetic variation under current and future climates has

implications for optimizing plant breeding and cultivation strategies.

References and Notes

1. F. Laibach, *Arabidopsis Information Service 01S* (1965).
2. L. Thompson, *J. Ecol.* **82**, 63 (1994).
3. Y. Kobayashi, D. Weigel, *Genes Dev.* **21**, 2371 (2007).
4. I. Ausin, C. Alonso-Blanco, J. M. Martinez-Zapater, *Int. J. Dev. Biol.* **49**, 689 (2005).
5. I. Baurle, C. Dean, *Cell* **125**, 655 (2006).
6. M. Koornneef, C. J. Hanhart, J. H. van der Veen, *Mol. Gen. Genet.* **229**, 57 (1991).
7. J. Putterill, F. Robson, K. Lee, R. Simon, G. Coupland, *Cell* **80**, 847 (1995).
8. R. N. Wilson, J. W. Heckman, C. R. Somerville, *Plant Physiol.* **100**, 403 (1992).
9. M. A. Blazquez, J. H. Ahn, D. Weigel, *Nat. Genet.* **33**, 168 (2003).
10. S. M. Welch, J. L. Roe, Z. Dong, *Agron. J.* **95**, 71 (2003).
11. J. Lempe *et al.*, *PLoS Genet.* **1**, e6 (2005).
12. S. Balasubramanian, S. Sureshkumar, J. Lempe, D. Weigel, *PLoS Genet.* **2**, e106 (2006).
13. R. J. Schmitz, R. M. Amasino, *Biochim. Biophys. Acta* **1769**, 269 (2007).
14. S. Sung, R. M. Amasino, *Nature* **427**, 159 (2004).
15. E. S. Dennis, W. J. Peacock, *Curr. Opin. Plant Biol.* **10**, 520 (2007).
16. G. G. Simpson, C. Dean, *Science* **296**, 285 (2002).
17. S. D. Michaels, Y. H. He, K. C. Scortecci, R. M. Amasino, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 10102 (2003).
18. See supporting material on *Science* Online.
19. T. Mizoguchi *et al.*, *Plant Cell* **17**, 2255 (2005).
20. J. H. Jung *et al.*, *Plant Cell* **19**, 2736 (2007).
21. I. Lee, R. M. Amasino, *Plant Physiol.* **108**, 157 (1995).
22. S. D. Michaels, R. M. Amasino, *Plant Cell* **13**, 935 (2001).
23. L. T. Burghardt *et al.*, in *19th International Conference on Arabidopsis Research* (Montreal, 23–27 July 2008), abstract ICAR802.
24. G. S. McMaster *et al.*, *Ann. Bot.* **102**, 561 (2008).
25. S. M. Welch, Z. Dong, J. L. Roe, *Aust. J. Agric. Res.* **56**, 919 (2005).
26. J. P. De Melo-Abreu *et al.*, *Agric. For. Meteorol.* **125**, 117 (2004).
27. G. T. Howe *et al.*, *Can. J. Bot.* **81**, 1247 (2003).
28. H. Böhlenius *et al.*, *Science* **312**, 1040 (2006); published online 3 May 2006 (10.1126/science.1126038).
29. J. Cockram *et al.*, *J. Exp. Bot.* **58**, 1231 (2007).

30. We thank M. Blazquez, C. Dean, M. Hoffmann, M. Koornneef, H. Kuittinen, and O. Savolainen for hosting these experiments at the European field sites; L. Albertson, L. Burghardt, C. Cooper, M. F. Cooper, E. Josephs, A. Lockwood, S. Myllylä, C. Oakley, R. Palmer, S. Rudder, and E. vonWettberg for assistance with planting and censusing; B. Robertson, M. Gosling, C. Lister, F. Eikelmann, G. Leuffen, W. Schuchert, T. Kauppila, and H. Eissner for technical and field assistance; R. Griffin, J. Kreikemeier, and G. Ragusa for assistance with the chamber experiment; A. Heim for contributions to the processing of temperature data; R. Amasino and J. Martinez-Zapater for sharing lines; S. Ruzsa of the Purugganan lab, who bulked lines for the experiments; and R. Amasino and C. Dean for manuscript comments. Supported by NSF Frontiers in Integrative Biological Research program grant EF-0425759 and an Alexander von Humboldt Research Award.

Supporting Online Material

www.sciencemag.org/cgi/content/full/1165826/DC1
Materials and Methods
Figs. S1 to S11
Tables S1 to S6
References

11 September 2008; accepted 19 November 2008
Published online 15 January 2009;
10.1126/science.1165826
Include this information when citing this paper.

The Hallucinogen *N,N*-Dimethyltryptamine (DMT) Is an Endogenous Sigma-1 Receptor Regulator

Dominique Fontanilla,[1] Molly Johannessen,[2] Abdol R. Hajipour,[3] Nicholas V. Cozzi,[1] Meyer B. Jackson,[2] Arnold E. Ruoho[1]*

The sigma-1 receptor is widely distributed in the central nervous system and periphery. Originally mischaracterized as an opioid receptor, the sigma-1 receptor binds a vast number of synthetic compounds but does not bind opioid peptides; it is currently considered an orphan receptor. The sigma-1 receptor pharmacophore includes an alkylamine core, also found in the endogenous compound *N,N*-dimethyltryptamine (DMT). DMT acts as a hallucinogen, but its receptor target has been unclear. DMT bound to sigma-1 receptors and inhibited voltage-gated sodium ion (Na^+) channels in both native cardiac myocytes and heterologous cells that express sigma-1 receptors. DMT induced hypermobility in wild-type mice but not in sigma-1 receptor knockout mice. These biochemical, physiological, and behavioral experiments indicate that DMT is an endogenous agonist for the sigma-1 receptor.

The sigma-1 receptor binds a broad range of synthetic compounds (*1*). It has long been suspected that the sigma-1 receptor is targeted by endogenous ligands, and several candidates have been proposed (*2*, *3*). Although progesterone and other neuroactive steroids are known to bind sigma-1 receptors and regulate some of their functions (*1*, *4*), they do not exhibit agonist properties on sigma-1–regulated ion channels in electrophysiological experiments (*5*).

Our search for a sigma receptor endogenous ligand (or ligands) was based on a variant of the canonical sigma-1 receptor ligand pharmacophore (*6*), but with a more basic structure (Fig. 1A). Otherwise dissimilar sigma-1 receptor ligands possess a common *N*-substituted pharmacophore (Fig. 1A): an *N,N*-dialkyl or *N*-alkyl-*N*-aralkyl product, most easily recognized in the high-affinity sigma-1 receptor ligand, fenpropimorph (*7*). Similar chemical backbones can be derived from other sigma-1 receptor ligands such as haloperidol and cocaine (Fig. 1A). *N*-substituted trace amines harbor this sigma-1 receptor ligand pharmacophore, but their interactions with sigma receptors have not been determined. Of particular interest is the only known endogenous mammalian *N,N*-dimethylated trace amine, *N,N*-dimethyltryptamine (DMT) (*8*–*10*). In addition to being one of the active compounds in psychoactive snuffs (*yopo, epená*) and sacramental teas (*ayahuasca, yagé*) used in native shamanic rituals in South America, DMT can be produced by enzymes in mammalian lung (*11*) and in rodent brain (*12*). DMT has been found in human urine, blood, and cerebrospinal fluid (*9*, *13*). Although there are no conclusive quantitative studies measuring the abundance of endogenous DMT because of its rapid metabolism (*14*), DMT concentrations can be localized and elevated in certain instances. Evidence suggests that DMT can be locally sequestered into brain neurotransmitter storage vesicles and that DMT production increases in rodent brain under environmental stress (*8*). Although a family of heterotrimeric GTP-binding protein (G protein)–coupled receptors (GPCRs) known as the trace amine receptors (TARs) was discovered in 2001 (*15*), only two members of this family respond to trace amines and have been renamed trace amine-associated receptors (TAARs) (*16*). Because other binding targets for trace amines and DMT are likely (*8*), we first examined the sigma-1 receptor binding affinities of the trace amines and their *N*-methylated and *N,N*-dimethylated counterparts.

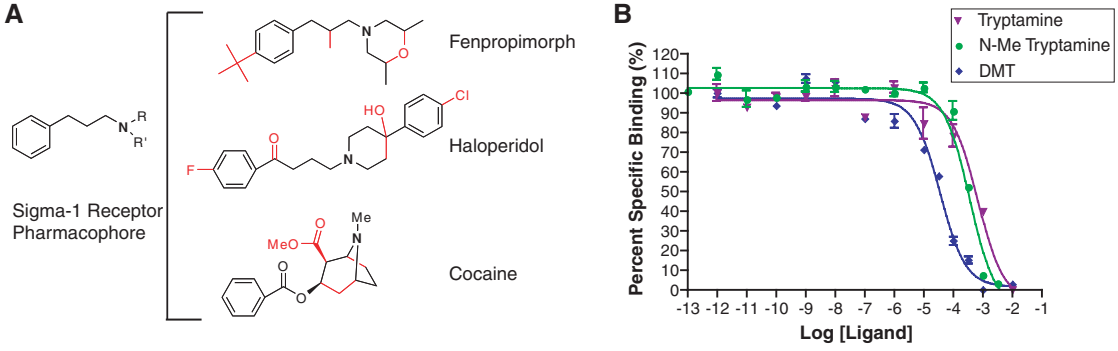
Competition assays against the sigma-1 receptor–specific ligand, (+)-[^{3}H]-pentazocine (10 nM), determined that the nonmethylated trace amines tryptamine, phenethylamine, and tyramine bound the sigma-1 receptor poorly (Fig. 1C), with dissociation constant (K_d) values of 431, 97.4, and >30,000 μM, respectively. By contrast, the *N*-methylated and *N,N*-dimethylated derivatives of these compounds bound sigma-1 receptors more tightly, with a clear increase in affinity as the ligands approached the sigma-1 receptor ligand pharmacophore (Fig. 1, A and B). With the exception of the *N*-methylated tyramines, this trend did not apply to the sigma-2 receptor, which differs pharmacologically and functionally from the sigma-1 receptor (Fig. 1C). Tryptamine, phenethylamine, and *N*-methyltyramine had the highest sigma-2 receptor affinities, with K_d values of 4.91, 7.31, and 6.61 μM, respectively. In contrast to sigma-1 receptors, *N*-methylation and *N,N*-

[1]Department of Pharmacology, University of Wisconsin School of Medicine and Public Health, Madison, WI 53706, USA. [2]Department of Physiology, University of Wisconsin School of Medicine and Public Health, Madison, WI 53706, USA. [3]Pharmaceutical Research Laboratory, Department of Chemistry, Isfahan University of Technology, Isfahan 84156, IR Iran.

*To whom correspondence should be addressed. E-mail: aeruoho@wisc.edu

dimethylation of tryptamine and phenethylamine decreased sigma-2 receptor affinity (Fig. 1C).

We tested the ability of tryptamine, *N*-methyltryptamine, and DMT to block sigma receptor photolabeling in rat liver homogenates by two radioactive photoaffinity labels, the sigma-1 receptor–specific cocaine derivative 3-[^{125}I]iodo-4-azidococaine ([^{125}I]-IACoc) (*17*) (Fig. 2A) and the sigma-1 and sigma-2 receptor fenpropimorph derivative 1-*N*-(2′,6′-dimethyl-morpholino)-3-(4-azido-3-[^{125}I]iodo-phenyl) propane ([^{125}I]IAF) (*18*) (Fig. 2B). Both of these compounds have been used to identify the drug binding region of the sigma-1 receptor (*18*, *19*). As anticipated, [^{125}I]-IACoc [sigma-1 K_d = 0.126 nM (*17*)] photolabeling of the 26-kD sigma-1 receptor (Fig. 2A) was protected best by DMT, with 61% protection by 50 μM DMT and almost 100% protection by 100 μM DMT. By contrast, tryptamine



C

Ligand	R_1	R_2	Sigma-1 K_D (μM) (± SEM, n=3), R^2	Sigma-2 K_D (μM) (± SEM, n=3), R^2	Ratio Sigma2/Sigma1
Tryptamine	H—	H—	431.55 (± 342.5, 0.87)	4.91 (± 5.2, 0.92)	0.011
N-methyltryptamine	H—	H_3C—	150 (± 61.6, 0.97)	12.82 (± 7.0, 0.94)	0.085
N,N'-dimethyltryptamine	H_3C—	H_3C—	14.75 (± 7.0, 0.96)	21.71 (± 10.8, 0.94)	1.472
Phenethylamine (PEA)	H—	H—	97.4 (± 23.3, 0.99)	7.31 (±5.2, 0.97)	0.075
N-methylPEA	H—	H_3C—	59.85 (± 22.3, 0.96)	21.93 (± 12.5, 0.94)	0.366
N,N'-dimethylPEA	H_3C—	H_3C—	12.89 (± 6.9, 0.92)	21.16 (± 11.8, 0.93)	1.642
Tyramine	H—	H—	>30,000	607 (± 422, 0.89)	>0.020
N-methyltyramine	H—	H_3C—	12.43 (± 2.7, 0.98)	6.61 (± 10.6, 0.91)	0.532
N,N'-dimethyltyramine	H_3C—	H_3C—	38.27 (± 10.3, 0.98)	64.29 (± 43.0, 0.94)	1.680
N-ethyltyramine	H—		1.62 (± 0.5 0.97)	N/A	N/A

Fig. 1. Sigma-1 receptor ligand pharmacophore and binding affinities. (**A**) A basic sigma-1 receptor ligand pharmacophore variant of Glennon *et al*. (*6*) was derived by removal of the red bonds from the sigma-1 receptor ligands fenpropimorph, haloperidol, and cocaine. (**B**) Competitive binding curves of tryptamine, *N*-methyltryptamine, and DMT, against the radioactive sigma-1 receptor ligand [^{3}H]-(+)-pentazocine. Curves are shown as percent specific binding (5 μM haloperidol). (**C**) Sigma-1 and sigma-2 receptor K_d values of trace amines and their *N*-methylated and *N*,*N*-dimethylated derivatives (scheme S2). Included are SEM values (n = 3 binding experiments) and R^2 values for a nonlinear regression curve fit. Solid arrows denote the direction of increasing affinity.

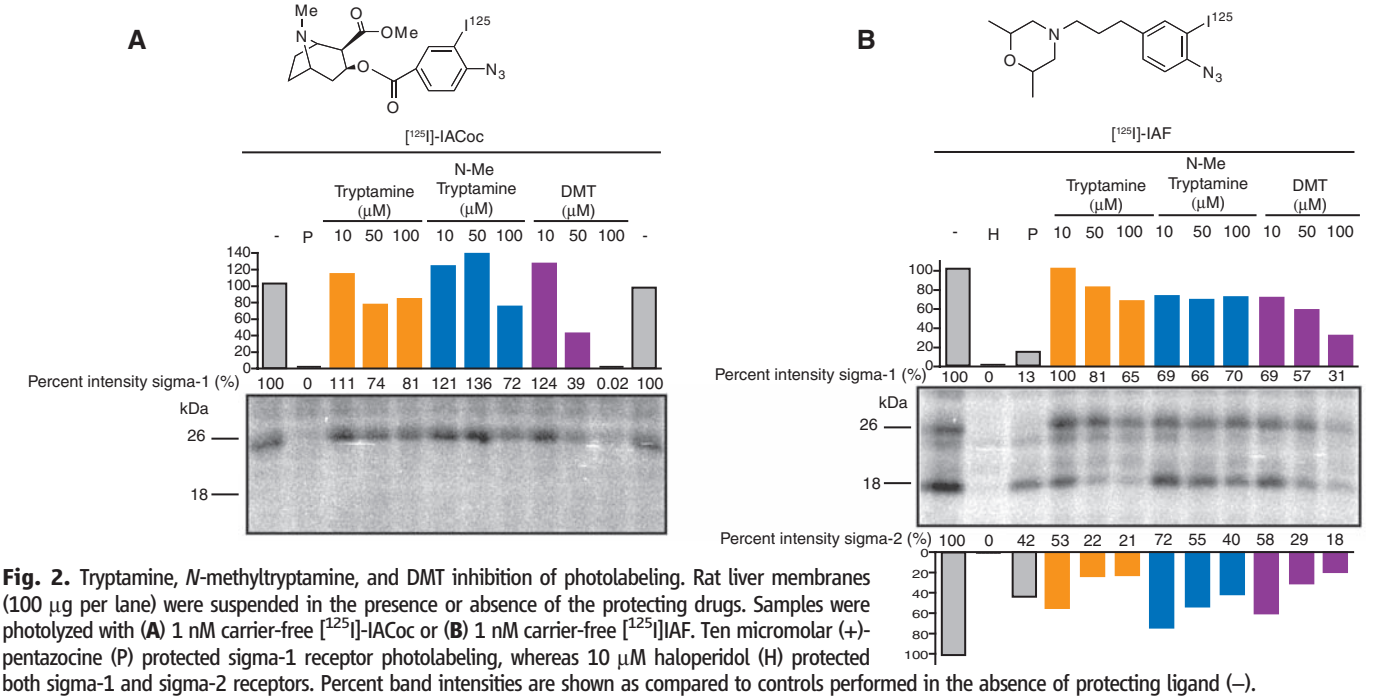


Fig. 2. Tryptamine, *N*-methyltryptamine, and DMT inhibition of photolabeling. Rat liver membranes (100 μg per lane) were suspended in the presence or absence of the protecting drugs. Samples were photolyzed with (**A**) 1 nM carrier-free [^{125}I]-IACoc or (**B**) 1 nM carrier-free [^{125}I]IAF. Ten micromolar (+)-pentazocine (P) protected sigma-1 receptor photolabeling, whereas 10 μM haloperidol (H) protected both sigma-1 and sigma-2 receptors. Percent band intensities are shown as compared to controls performed in the absence of protecting ligand (–).

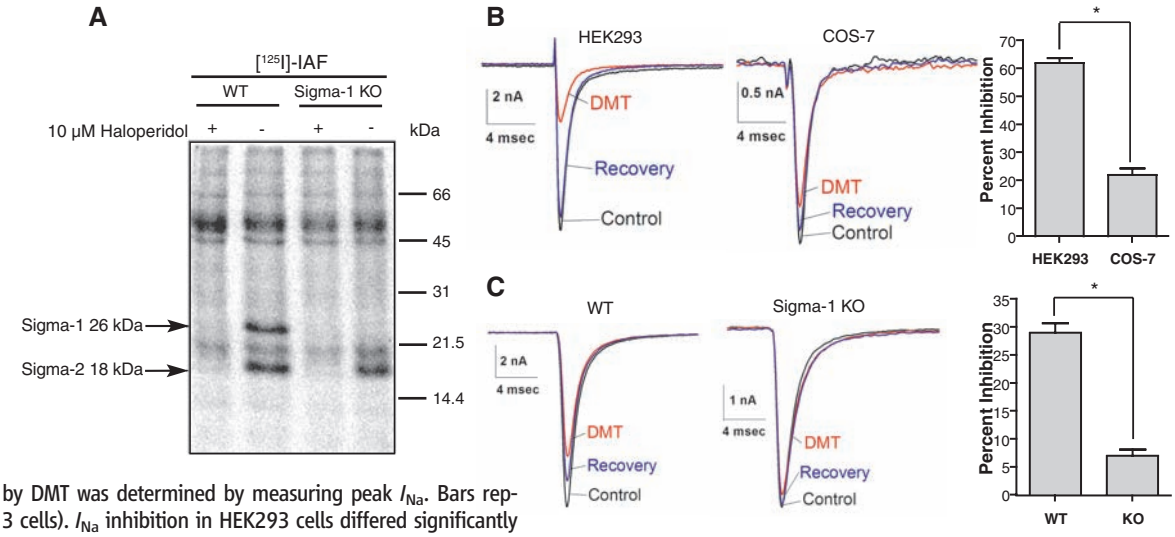


Fig. 3. Sodium channel inhibition by DMT. (**A**) In the presence or absence of 10 μM haloperidol, wild type (WT) or sigma-1 receptor knockout (KO) mouse liver homogenates (200 μg/lane) were photolabeled with 1 nM [^{125}I]IAF. (**B**) Examples of I_{Na} evoked by steps from –80 to –10 mV in HEK293 or COS-7 cells expressing hNav1.5 channel in the absence (control, black), presence (DMT, red), and after wash out (recovery, blue) of 100 μM DMT. Average inhibition by DMT was determined by measuring peak I_{Na}. Bars represent mean ± SEM (n = 3 cells). I_{Na} inhibition in HEK293 cells differed significantly from that in COS-7 cells (*P < 0.03). (**C**) Examples of I_{Na} evoked as described in (B) in neonatal cardiac myocytes from WT and KO mice in the absence (control, black), presence (DMT, red), and after wash out (recovery, blue) of 100 μM DMT. Current inhibition in WT was significantly different from that in KO (*P < 0.002, n = 7 neonatal cardiac myocytes).

and *N*-methyltryptamine protected minimally against sigma-1 receptor [^{125}I]-IACoc photolabeling, even at these high concentrations (Fig. 2A). Similarly, [^{125}I]IAF photolabeling of the sigma-1 [K_d = 194 nM (*18*)] receptor showed that DMT was the most potent protector. Ten micromolar DMT provided 31% protection, whereas 50 and 100 μM DMT provided 43 and 69% protection, respectively (Fig. 2B). With the exception of *N*-methyltryptamine, protection of [^{125}I]IAF sigma-2 [K_d = 2780 nM (*18*)] receptor photolabeling paralleled the sigma-2 binding data. Tryptamine afforded the greatest protection of sigma-2 receptor photolabeling, with values of 47, 78, and 79% for 10, 50, and 100 μM, respectively (Fig. 2B).

An important biological activity of sigma receptor activation is the inhibition of ion channels, which operates through protein-protein interactions without mediation by G proteins and protein kinases (*20*–*22*). In addition to modulating various types of voltage-activated K^+ channels (*21*, *23*, *24*), the sigma-1 receptor associates with the Kv1.4 K^+ channel in posterior pituitary nerve terminals, as well as in *Xenopus* oocytes (*22*). Sigma receptor ligands also modulate N-, L-, P/Q-, and R-type Ca^{2+} channels in rat sympathetic and parasympathetic neurons (*25*). Sigma receptor ligands modulate cardiac voltage-gated Na^+ channels (hNav1.5) in human embryonic kidney 293 (HEK293) cells, COS-7 cells, and neonatal mouse cardiac myocytes (*26*). To evaluate the capacity of DMT to induce physiological responses by binding to sigma receptors, we examined the action of DMT on voltage-activated Na^+ current. Patch-clamp recordings from HEK293 cells stably expressing the human cardiac Na^+ channel hNav1.5 revealed voltage-activated Na^+ currents (I_{Na}) in response to voltage steps from –80 to –10 mV (Fig. 3B). Application of 100 μM DMT inhibited I_{Na} by 62 ± 3% (n = 3 HEK293 cells), which reversed upon DMT removal. With hNav1.5 transiently transfected into COS-7 cells, 100 μM DMT inhibited I_{Na} by only 22 ± 4% (n = 3 COS-7 cells), but photolabeling has shown that these cells have much lower concentrations of endogenous sigma-1 receptors compared to HEK293 cells (fig. S1 and Fig. 3B). The difference between DMT inhibition of I_{Na} in HEK293 and COS-7 cells (Fig. 3B, P < 0.03) thus demonstrates the dependence of I_{Na} inhibition on sigma-1 receptors. Experiments in cardiac myocytes demonstrated the same DMT action in a native preparation (Fig. 3C) and enabled further demonstration of sigma-1 receptor dependence by using a sigma-1 receptor

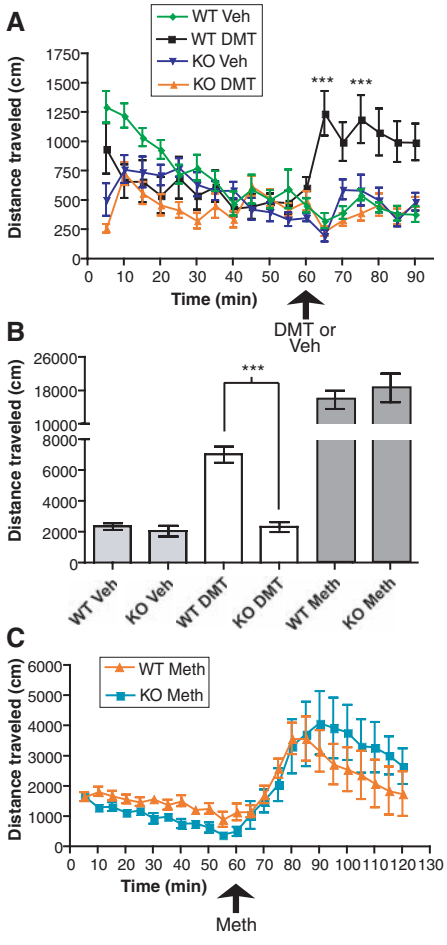


Fig. 4. DMT-induced hypermobility abrogated in the sigma-1 KO mouse. (**A**) Distances traveled by WT and KO mice were measured in an open-field assay in 5-min increments. Pargyline was injected 2 hours before DMT or vehicle (Veh) ip injection. Bars represent mean ± SEM (n = 8 to 14 mice). WT mice showed a significant (***P < 0.0001) increase in mobility in response to DMT as compared to KO mice. (**B**) Total distance traveled over 30 min after DMT, vehicle (Veh), or methamphetamine (Meth, n = 6 mice) injection in WT and KO mice. (**C**) Methamphetamine serves as a positive control for hypermobility in KO mice.

knockout mouse (*27*). [^{125}I]IAF photolabeling of liver homogenates from wild-type (WT) and sigma-1 receptor knockout (KO) mice indeed showed the absence of sigma-1 receptor (26 kD) in the KO samples (Fig. 3A). In WT neonatal cardiac myocytes, 100 μM DMT reversibly inhibited I_{Na} by 29 ± 3% (n = 7 WT myocytes), whereas I_{Na} was reduced by only 7 ± 2% (n = 7 KO myocytes) in KO myocytes (Fig. 3C, $P < 0.002$).

Both DMT and sigma receptor ligands influence animal behavior. DMT injection induces hypermobility in rodents concurrently treated with the monoamine oxidase inhibitor pargyline (*28*), and this action is not antagonized by blockers of dopamine or serotonin receptors, but is potently inhibited by haloperidol (*28*). Although haloperidol is thought to act in part through the dopamine D_2 receptor system, it is also a potent sigma-1 receptor agonist [sigma-1 inhibition constant (K_i) = 3 nM (*29*); sigma-2 K_i = 54 nM (*29*)] when inhibiting voltage-gated ion channels (*5*, *25*). Haloperidol reduces brain concentrations of DMT (*8*) and DMT inhibits haloperidol binding in brain tissue more robustly than the dopamine agonist apomorphine (*8*). On the basis of these findings, which were discovered before sigma receptor identification, DMT has been hypothesized to act through an unknown "hallucinogen" receptor (*8*). We confirmed results (*28*) that intraperitoneal (ip) administration of DMT (2 mg per kilogram of body weight) 2 hours after pargyline (75 mg/kg, ip) injection induced hypermobility in WT mice (7025 ± 524.1 cm, n = 12 WT mice) in an open-field assay. Identical drug treatments in sigma-1 receptor KO mice had no hypermobility action (2328 ± 322.9 cm, n = 12 KO mice, $P < 0.0001$; Fig. 4, A and B). This result is particularly important to our understanding of sigma-1 receptor biological function because the KO mice are viable and fertile (*27*). The sigma-1 receptor dependence of DMT-induced hypermobility parallels that induced by the sigma-1 receptor ligand (+)-SKF10047 in WT but not in KO mice (*27*). As a positive control, methamphetamine, which is thought to act through catecholaminergic systems, induced hypermobility in both WT and KO mice (3 mg/kg, ip, n = 6 mice; Fig. 4, B and C) with a reduced onset rate compared with that seen for DMT (Fig. 4, A and C). This indicates that behavioral actions of DMT depend on the sigma-1 receptor, which may provide an alternative research area for psychiatric disorders that have not been linked to dopamine or *N*-methyl-D-aspartate systems.

The binding, biochemical, physiological, and behavioral studies reported here all support the hypothesis that DMT acts as a ligand for the sigma-1 receptor. On the basis of our binding results and the sigma-1 receptor pharmacophore, endogenous trace amines and their *N*-methyl and *N,N*-dimethyl derivatives are likely to serve as endogenous sigma receptor regulators. Moreover, DMT, the only known mammalian *N,N*-dimethylated trace amine, can activate the sigma-1 receptor to modulate Na^+ channels. The recent discovery that the sigma-1 receptor functions as a molecular chaperone (*30*) may be relevant, because sigma-1 receptors, which are observed in the endoplasmic reticulum, associate with plasma membrane Kv 1.4 channels (*22*) and may serve as a molecular chaperone for ion channels. Furthermore, the behavioral effect of DMT may be due to activation or inhibition of sigma-1 receptor chaperone activity instead of, or in addition to, DMT/sigma-1 receptor modulation of ion channels. These studies thus suggest that this natural hallucinogen could exert its action by binding to sigma-1 receptors, which are abundant in the brain (*1*, *27*). This discovery may also extend to *N,N*-dimethylated neurotransmitters such as the psychoactive serotonin derivative *N,N*-dimethylserotonin (bufotenine), which has been found at elevated concentrations in the urine of schizophrenic patients (*10*). The finding that DMT and sigma-1 receptors act as a ligand-receptor pair provides a long-awaited connection that will enable researchers to elucidate the biological functions of both of these molecules.

References and Notes

1. T. Hayashi, T. P. Su, *CNS Drugs* **18**, 269 (2004).
2. P. Bouchard *et al.*, *Eur. J. Neurosci.* **7**, 1952 (1995).
3. T. P. Su, A. D. Weissman, S. Y. Yeh, *Life Sci.* **38**, 2199 (1986).
4. T. P. Su, E. D. London, J. H. Jaffe, *Science* **240**, 219 (1988).
5. R. A. Wilke *et al.*, *J. Physiol.* **517**, 391 (1999).
6. R. A. Glennon *et al.*, *J. Med. Chem.* **37**, 1214 (1994).
7. F. F. Moebius, R. J. Reiter, M. Hanner, H. Glossmann, *Br. J. Pharmacol.* **121**, 1 (1997).
8. S. A. Barker, J. A. Monti, S. T. Christian, *Int. Rev. Neurobiol.* **22**, 83 (1981).
9. F. Franzen, H. Gross, *Nature* **206**, 1052 (1965).
10. M. S. Jacob, D. E. Presti, *Med. Hypotheses* **64**, 930 (2005).
11. J. Axelrod, *Science* **134**, 343 (1961).
12. J. M. Saavedra, J. Axelrod, *Science* **175**, 1365 (1972).
13. J. M. Beaton, P. E. Morris, *Mech. Ageing Dev.* **25**, 343 (1984).
14. S. A. Burchett, T. P. Hicks, *Prog. Neurobiol.* **79**, 223 (2006).
15. B. Borowsky *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 8966 (2001).
16. L. Lindemann *et al.*, *Genomics* **85**, 372 (2005).
17. J. R. Kahoun, A. E. Ruoho, *Proc. Natl. Acad. Sci. U.S.A.* **89**, 1393 (1992).
18. A. Pal *et al.*, *Mol. Pharmacol.* **72**, 921 (2007).
19. Y. Chen, A. R. Hajipour, M. K. Sievert, M. Arbabian, A. E. Ruoho, *Biochemistry* **46**, 3532 (2007).
20. P. J. Lupardus *et al.*, *J. Physiol.* **526**, 527 (2000).
21. H. Zhang, J. Cuevas, *J. Pharmacol. Exp. Ther.* **313**, 1387 (2005).
22. E. Aydar, C. P. Palmer, V. A. Klyachko, M. B. Jackson, *Neuron* **34**, 399 (2002).
23. R. A. Wilke *et al.*, *J. Biol. Chem.* **274**, 18387 (1999).
24. C. Kennedy, G. Henderson, *Neuroscience* **35**, 725 (1990).
25. H. Zhang, J. Cuevas, *J. Neurophysiol.* **87**, 2867 (2002).
26. M. A. Johannessen, A. Ramos-Serrano, S. Ramachandran, A. E. Ruoho, M. B. Jackson, "Sigma receptor modulation of voltage-dependent sodium channels," Program No. 466.22, Annual Neuroscience Meeting, San Diego, CA, 5 November 2007.
27. F. Langa *et al.*, *Eur. J. Neurosci.* **18**, 2188 (2003).
28. P. Jenner, C. D. Marsden, C. M. Thanki, *Br. J. Pharmacol.* **69**, 69 (1980).
29. R. R. Matsumoto, B. Pouw, *Eur. J. Pharmacol.* **401**, 155 (2000).
30. T. Hayashi, T. P. Su, *Cell* **131**, 596 (2007).

31. We thank the Corinna Burger laboratory for use of their mouse behavior equipment, and A. Paul and T. Mavlyutov for providing [^{125}I]IAF and [^{125}I]-IACoc, respectively. Supported by the Molecular and Cellular Pharmacology (MCP) Graduate Program training grant from NIH T32 GM08688 and by the NIH Ruth L. Kirschstein National Research Service Award (NRSA) (F31 DA022932) from the National Institute on Drug Abuse (to D.F.). This work was funded by NIH grants R01 MH065503 (to A.E.R.) and NS30016 (to M.B.J.).

Supporting Online Material

www.sciencemag.org/cgi/content/full/323/5916/934/DC1
Materials and Methods
Fig. S1 and scheme S2
References

18 September 2008; accepted 10 December 2008
10.1126/science.1166127

When Your Gain Is My Pain and Your Pain Is My Gain: Neural Correlates of Envy and Schadenfreude

Hidehiko Takahashi,[1,2,3]* Motoichiro Kato,[4] Masato Matsuura,[2] Dean Mobbs,[5] Tetsuya Suhara,[1] Yoshiro Okubo[6]

We often evaluate the self and others from social comparisons. We feel envy when the target person has superior and self-relevant characteristics. Schadenfreude occurs when envied persons fall from grace. To elucidate the neurocognitive mechanisms of envy and schadenfreude, we conducted two functional magnetic resonance imaging studies. In study one, the participants read information concerning target persons characterized by levels of possession and self-relevance of comparison domains. When the target person's possession was superior and self-relevant, stronger envy and stronger anterior cingulate cortex (ACC) activation were induced. In study two, stronger schadenfreude and stronger striatum activation were induced when misfortunes happened to envied persons. ACC activation in study one predicted ventral striatum activation in study two. Our findings document mechanisms of painful emotion, envy, and a rewarding reaction, schadenfreude.

Envy is one of the seven biblical sins, the Shakespearian "green-eyed monster," and what Bertrand Russell (*1*) called an unfortunate facet of human nature. It is an irrational, unpleasant feeling and a "painful emotion" (*2*) characterized by feelings of inferiority and resentment produced by an awareness of another's superior quality, achievement, or possessions (*3*). Understanding envy is important because of its broad implications, ranging from individual mat-

ters to social problems. It concerns personal life satisfaction (*4*), self-evaluation/maintenance (*5*), and economic and political issues (*6*–*8*). We judge objects more by comparison than by their intrinsic worth and value (*9*), and self-evaluations are often derived from social comparisons with people who are self-relevant, sharing similar attributes, characteristics, group memberships, and interests (for example, gender, age, and social class) (*10*).

When envy is evoked, we often have a desire to possess the same advantage or may wish that the other lacks it (*3*). When misfortune occurs to others, emotions can manifest themselves in several ways. We can sympathize and have feelings of concern and sorrow for the other person (*11*, *12*), but we can also experience schadenfreude, a rewarding feeling derived from another's misfortune (*13*). Schadenfreude is closely related to envy, and it is more likely to arise when misfortune happens to a person who is advantaged and self-relevant than to someone who is neither advantaged nor self-relevant (*13*–*15*).

We investigated the brain activation associated with envy and schadenfreude. We conducted two functional magnetic resonance imaging (fMRI) studies to test two complementary hypotheses. In the first study, we hypothesized that, not only the level of possession of the person we compare ourselves with, but also the self-relevance of the comparison domain affects brain activation associated with envy through social comparison. We usually have a positive self-concept, and we experience a feeling of discomfort when we perform in a way that violates this self-concept (*16*). The anterior cingulate cortex (ACC) is activated when this positive self-concept conflicts with external information (*17*, *18*). Bearing in mind that envy is a painful emotion, we hypothesized that envy activates the dorsal ACC (dACC), where cognitive conflicts (*19*) or social pain (*12*, *20*) are processed. We predicted that dACC activation is stronger when an envied person has superior and more self-relevant possessions. In the second study, we hypothesized that a misfortune happening to an envied person produces greater brain activation associated with schadenfreude than misfortune happening to a person who is not envied. Schadenfreude should activate the ventral striatum, a central node of reward processing.

Nineteen healthy volunteers [10 men and 9 women, mean age = 22.1 ± 1.4 (SD) years] participated in the two fMRI studies. We used a scenario method as in previous social affective neuroimaging studies (*21*, *22*). Each participant was presented with a scenario in which the protagonist (oneself) and three other target persons appeared. Materials were employed from an initial survey to validate our expected results (*23*). Before the fMRI scans, we asked the participants to read and understand the scenario thoroughly and to imagine the protagonist of the scenario as themselves. In study one, we aimed to determine the level of envy in terms of whether possessions of the target person were superior or not and whether domains of comparison were self-relevant or not. In short, for male participants, the protagonist of the scenario was male and average in terms of possessions such as ability, quality, and social status. Male student A shared similar attributes with the protagonist. He possessed superior quality and ability, and the domains of comparison were important and relevant to the protagonist [superior and high relevance (SpHi)]. Female student B had different attributes and background from the protagonist. She also possessed superior quality and ability, but the domains of comparison were neither important nor relevant to the protagonist [superior and low relevance (SpLo)]. Female student C had different attributes and background from the protagonist. She possessed mediocre quality and ability, and the domains of comparison were neither important nor relevant to the protagonist [average and low relevance (AvLo)]. The scenario for male participants and profiles of the persons are shown in the appendix in (*23*). The profiles of the three target persons and comparison domains are summarized in table S1, and a schematic depiction of the stimuli and design is shown in fig. S1. We performed event-related fMRI analysis with statistical parametric mapping 2 to examine activations in response to SpHi, SpLo, and AvLo. In study two, successive misfortunes happened to student A (SpHi) and student C (AvLo) in the scenario examining reaction in response to misfortunes happening to others. A list of misfortunes is provided in table S1, and a schematic depiction of the stimuli and design is shown in fig. S2. We analyzed neural responses to misfortunes on SpHi (MisSpHi) and AvLo (MisAvLo). After the scans, the participants rated each event presented in study one in terms of how much envy they felt for the three students (i.e., 1 = no envy, 6 = extremely envious). Similarly, the participants also reported the intensity of their pleasure (schadenfreude) (1 = no pleasure, 6 = extremely pleasant) in response to misfortunes happening to students A and C in study two. That is, they gave one envy score per domain per student in study one and one schadenfreude score per misfortune per student in study two.

The self-rating results of the participants in the fMRI study were comparable to the results obtained in the initial survey. The mean values of the ratings of envy for students A, B, and C were 4.0 ± 1.0, 2.1 ± 0.8, and 1.0 ± 0.0, respectively. The mean values of schadenfreude for students A and C were

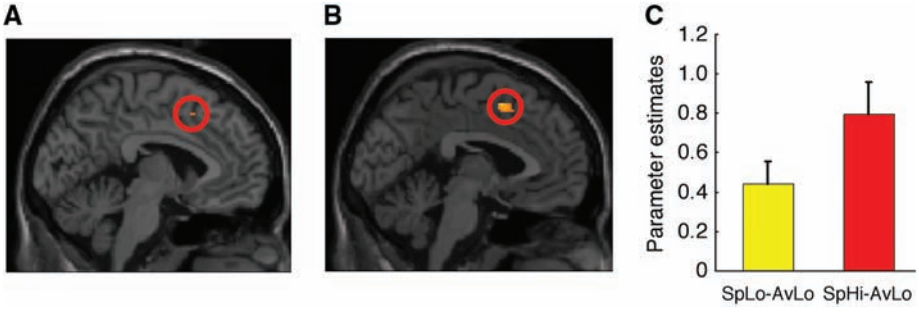


Fig. 1. Brain activation in dACC was modulated by relevance of comparison domain. Brain activations in response to (**A**) the SpLo minus AvLo condition and (**B**) the SpHi minus AvLo condition. (**C**) Mean for parameter estimates at the peak of dACC activation for SpHi-AvLo contrast (red) was greater than that for SpLo-AvLo contrast (yellow) (t = 2.56, P = 0.02). Error bars represent SE.

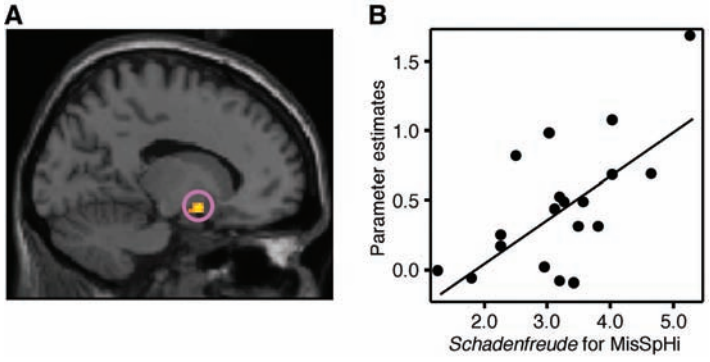


Fig. 2. Correlation between self-rating of schadenfreude and ventral striatum activation across participants. (**A**) Image showing correlation between mean rating of schadenfreude for MisSpHi and the ventral striatum in MisSpHi-MisAvLo contrast across participants. (**B**) Plots and regression line of correlation (r = 0.65, P = 0.002) between schadenfreude and parameter estimates of the ventral striatum activation for MisSpHi-MisAvLo contrast at a peak voxel (–14, 2, –12).

[1]Department of Molecular Neuroimaging, National Institute of Radiological Sciences, 9-1, 4-chome, Anagawa, Inage-ku, Chiba, 263-8555, Japan. [2]Department of Life Sciences and Bioinformatics, Graduate School of Health Sciences, Tokyo Medical and Dental University, 1-5-45 Yushima, Bunkyo-ku Tokyo, 113-8549, Japan. [3]Precursory Research for Embryonic Science and Technology, Japan Science and Technology Agency, 4-1-8 Honcho, Kawaguchi, Saitama, 332-0012, Japan. [4]Department of Neuropsychiatry, Keio University School of Medicine, 35 Shinanomachi, Shinjuku-ku, Tokyo, 160-8582, Japan. [5]Medical Research Council (MRC)–Cognition and Brain Sciences Unit, University of Cambridge, Cambridge, CB2 7EF, UK. [6]Department of Neuropsychiatry, Nippon Medical School, 1-1-5, Sendagi, Bunkyo-ku, Tokyo, 113-8603, Japan.

*To whom correspondence should be addressed. E-mail: hidehiko@nirs.go.jp

3.3 ± 1.0 and 1.0 ± 0.0, respectively. Self-rating scores of envy for student A were positively correlated with the magnitude of schadenfreude for student A (correlation coefficient $r = 0.50$, $P = 0.03$). Both SpHi-AvLo and SpLo-AvLo conditions produced activations in dACC, a region implicated in the processing of conflict or pain, but dACC activation was greater in the SpHi-AvLo condition ($x = -4$, $y = 8$, $z = 54$, z score = 4.07) than in the SpLo-AvLo condition ($x = -4$, $y = 16$, $z = 46$, Z score = 3.65) (Fig. 1, A to C). Regression analysis revealed positive linear correlation between self-rating scores of envy and the degree of activation in the dACC ($x = -2$, $y = 10$, $z = 52$, z score = 4.36) in SpHi-AvLo contrast (fig. S3, A and B). The MisSpHi-MisAvLo condition produced activations in the reward-related regions: the dorsal striatum (caudate, putamen) ($x = -16$, $y = -2$, $z = 16$, z score = 4.44), the ventral striatum including the nucleus accumbens ($x = -12$, $y = 6$, $z = -10$, z score = 4.41), and the medial orbitofrontal cortex ($x = -8$, $y = 54$, $z = -10$, z score = 3.46) (fig. S4, A and B). There was correlation between the intensity of schadenfreude and the degree of activation in the ventral striatum ($x = -14$, $y = 2$, $z = -12$, z score = 3.98) in MisSpHi-MisAvLo contrast (Fig. 2, A and B). dACC ($x = -2$, $y = 10$, $z = 52$) activation in SpHi-AvLo contrast was positively correlated with ventral striatum ($x = -14$, $y = 2$, $z = -12$) activation in MisSpHi-MisAvLo contrast (Fig. 3).

This study investigated the neurocognitive mechanisms of envy and schadenfreude and the role of social comparison in the central processing of these emotions. At the behavioral level in study one, the intensity of envy is modulated by the quality of the possession of the person we compare with and the self-relevance of the comparison domain. That is, if the possession of the target person is superior and the comparison domain is self-relevant, we feel intense envy.

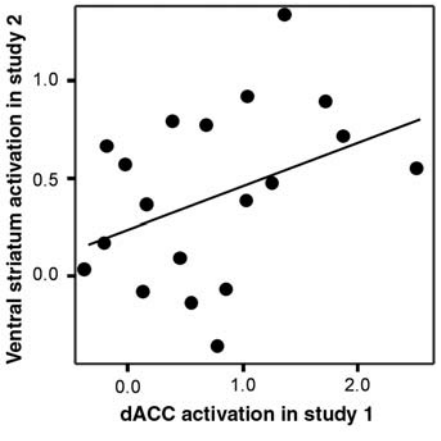


Fig. 3. Relation between dACC activation associated with envy and ventral striatum activation associated with schadenfreude. The *x* axis indicates the parameter estimates of dACC activation for SpHi-AvLo contrast at a peak voxel (–2, 10, 52). The *y* axis indicates the parameter estimates of the ventral striatum activation for MisSpHi-MisAvLo contrast at a peak voxel (–14, 2, –12). Positive correlation between dACC activation in study one and ventral striatum activation in study two across participants is shown ($r = 0.39$, $P = 0.01$).

When the comparison domain is not self-relevant, we do not feel strong envy, even if the possession is superior. When the comparison target is neither superior nor self-relevant, we are indifferent to the target. Activation of dACC was also modulated by possession quality and self-relevance. Stronger dACC activation was observed when one felt stronger envy. Moreover, between-participant correlation analysis demonstrated that people with stronger envy showed greater activation in dACC. At the behavioral level in study two, stronger schadenfreude was related to stronger envy, and schadenfreude arose when misfortune occurred to a person who was advantaged and self-relevant. Striatal activation was observed when misfortune happened to an envied person but not when it happened to a nonenvied person. Between-participant analysis revealed that people with stronger schadenfreude showed greater activation in the ventral striatum.

ACC activation in response to envy stimuli might reflect a painful feature of this emotion. It was comparable to caudal ACC activation in response to pain in the self but not to pain in others (empathic pain) (*12*), suggesting that the participants experienced a painful feeling. Activation in this region has been reported in response to social pain (distress of social exclusion) (*20*). Taken together, envy might be a social pain in the self, with feelings of being excluded from the field that one is concerned with.

We are usually motivated to maintain a positive self-concept (*16*), and we feel discomfort when our self-concept is threatened by others who outperform ourselves in a self-relevant domain. Considering the role of dACC in conflict-monitoring (*19*), the association between envy and dACC activation suggests that envy is a condition in which information recognized by social comparison conflicts with positive self-concept. Experiencing discomfort motivates us to reduce it. Discomfort arising from others outperforming us in our cherished domains can be resolved by reducing the relevance of the domain to us or changing relative performance (*16*). Students in our scenario might change their major or club at the university and, ultimately, their goals in life. Alternatively, they might make an effort to improve their own performance or possession. On the contrary, they might wish that the other lacks advantages, or they may even obstruct the advantaged student (with malice). Similarly, from an economic perspective, envy has productive and destructive effects on economic growth. It motivates the members in organizations to enhance their own performances or to sabotage their opponents' performances (*24*). When misfortune occurs to an advantaged person and contributes to narrowing the gap of relative performance in an important domain, discomfort or pain is reduced, and a pleasant feeling is induced. This pleasure at another's misfortune is correspondent to the activation of the ventral striatum and the medial orbitofrontal cortex (*25*, *26*). The striatum has also been implicated in altruistic punishment (*27*) and observing an unfair person receiving pain (*28*). Stronger dACC activation induced by the most envied student in study one predicted stronger ventral striatum activation when misfortunes occurred to the student in study two. This means that people who tend to have higher pain or conflict are more likely to have a strong pleasant feeling once they are relieved from this pain. Thus, our findings propose a neurocognitive mechanism of a psychologically rewarding reaction, schadenfreude, and its relation to envy. At the same time, ventral striatum activation without receiving an actual reward indicates that we did not evaluate objects solely by their absolute value but that social comparison plays a substantial role in evaluation (*29*).

References and Notes

1. B. Russell, *The Conquest of Happiness* (W.W. Norton, New York, 1930).
2. Aristotle, *The Art of Rhetoric* (Penguin Books, London, 1981).
3. W. G. Parrott, R. H. Smith, *J. Pers. Soc. Psychol.* **64**, 906 (1993).
4. R. H. Smith, W. G. Parrott, E. F. Diener, R. H. Hoyle, S. H. Kim, *Pers. Soc. Psychol. Bull.* **25**, 1007 (1999).
5. A. Tesser, *Adv. Exp. Soc. Psychol.* **21**, 181 (1988).
6. J. Rawls, *A Theory of Justice* (Harvard Univ. Press, Cambridge, MA, 1971).
7. R. Nozick, *Anarchy, State, and Utopia* (Basic Books, New York, 1974).
8. G. F. de la Mora, *Egalitarian Envy: The Political Foundations of Social Justice* (Paragon House, New York, 1987).
9. D. Hume, *A Treatise of Human Nature* (Oxford Univ. Press, Oxford, 1978).
10. L. Festinger, *Hum. Relat.* **7**, 117 (1954).
11. N. Eisenberg, *Novartis Found. Symp.* **278**, 71 (2007).
12. T. Singer *et al*., *Science* **303**, 1157 (2004).
13. F. Heider, *The Psychology of Interpersonal Relations* (Wiley & Sons, New York, 1958).
14. B. de Spinoza, *The Ethics* (Biblio Bazaar, Charleston, SC, 2006).
15. W. W. van Dijk, J. W. Ouwerkerk, S. Goslinga, M. Nieweg, M. Gallucci, *Emotion* **6**, 156 (2006).
16. A. Tesser, D. Cornell, *J. Exp. Soc. Psychol.* **27**, 501 (1991).
17. D. M. Amodio *et al*., *Psychol. Sci.* **15**, 88 (2004).
18. E. Harmon-Jones, *Biol. Psychol.* **67**, 51 (2004).
19. J. G. Kerns *et al*., *Science* **303**, 1023 (2004).
20. N. I. Eisenberger, M. D. Lieberman, K. D. Williams, *Science* **302**, 290 (2003).
21. J. D. Greene, R. B. Sommerville, L. E. Nystrom, J. M. Darley, J. D. Cohen, *Science* **293**, 2105 (2001).
22. T. Sharot, A. M. Riccardi, C. M. Raio, E. A. Phelps, *Nature* **450**, 102 (2007).
23. Materials and methods are available as supporting material on *Science* Online.
24. E. Lazear, *J. Polit. Econ.* **97**, 561 (1989).
25. S. M. McClure, M. K. York, P. R. Montague, *Neuroscientist* **10**, 260 (2004).
26. E. Fehr, C. F. Camerer, *Trends Cogn. Sci.* **11**, 419 (2007).
27. D. J.-F. de Quervain *et al*., *Science* **305**, 1254 (2004).
28. T. Singer *et al*., *Nature* **439**, 466 (2006).
29. K. Fliessbach *et al*., *Science* **318**, 1305 (2007).

30. We gratefully thank C. Frith for his valuable comments. This study was supported by a Health and Labor Sciences Research Grant for Comprehensive Research on Disability, Health, and Welfare (H20-SYOGAI-011) from the Japanese Ministry of Health, Labor, and Welfare, and a Health and Labor Sciences Research Grant for Research on Psychiatric and Neurological Diseases and Mental Health (H20-KOKORO-025) from the Japanese Ministry of Health, Labor, and Welfare. D.M. is support by MRC (UK).

Supporting Online Material

www.sciencemag.org/cgi/content/full/323/5916/937/DC1
Materials and Methods
SOM Text
Figs. S1 to S4
Table S1

8 September 2008; accepted 10 December 2008
10.1126/science.1165604

A Neural Mechanism for Microsaccade Generation in the Primate Superior Colliculus

Ziad M. Hafed,[1]* Laurent Goffart,[2] Richard J. Krauzlis[1]

During fixation, the eyes are not still but often exhibit microsaccadic movements. The function of microsaccades is controversial, largely because the neural mechanisms responsible for their generation are unknown. Here, we show that the superior colliculus (SC), a retinotopically organized structure involved in voluntary-saccade target selection, plays a causal role in microsaccade generation. Neurons in the foveal portion of the SC increase their activity before and during microsaccades with sizes of only a few minutes of arc and exhibit selectivity for the direction and amplitude of these movements. Reversible inactivation of these neurons significantly reduces microsaccade rate without otherwise compromising fixation. These results, coupled with computational modeling of SC activity, demonstrate that microsaccades are controlled by the SC and explain the link between microsaccades and visual attention.

Microsaccades are the very small (typically <12 min arc), involuntary, fast eye movements that occur during fixation (*1*–*3*). The behavioral properties and functional role of microsaccades have been extensively studied, and sometimes vigorously debated, for many years (*1*–*14*). However, the neural mechanisms responsible for their generation are unexplored. We show that the superior colliculus (SC), a retinotopically organized structure known to be important for selecting and initiating voluntary eye movements (*15*–*17*), is also part of the neural mechanism that controls microsaccades.

We analyzed SC activity associated with 15,205 microsaccades that occurred while monkeys fixated a small stationary spot (*18*). Each fixation trial lasted for 3500 ms, resulting in many microsaccades with a variety of directions and amplitudes (Fig. 1A and fig. S1). These movements had dynamics like those of larger saccades (*3*) (fig. S1A), consistent with evidence that premotor neurons (downstream from the SC) are active during movements as small as 12 to 15 min arc (*19*).

We targeted neurons in the rostral pole of the SC, which represents foveal goal locations (*18*, *20*). Figure 1A shows the spiking activity of a neuron in the left SC during a single trial containing nine microsaccades (highlighted in green). The neuron exhibited changes in activity that were correlated with microsaccades. For example, the microsaccades labeled 1 and 2 in Fig. 1A were predominantly downward and leftward, respectively, and both were associated with a reduction in the neuron's activity. In contrast, small, predominantly rightward microsaccades, such as the ones labeled 3 and 4, caused an increase in activity before microsaccade onset and

[1]Systems Neurobiology Laboratory, Salk Institute for Biological Studies, 10010 North Torrey Pines Road, La Jolla, CA 92037, USA. [2]Institut de Neurosciences Cognitives de la Méditerranée, Equipe Dynamique de la Perception Visuelle et de l'Action, UMR 6193, CNRS–Aix-Marseille Universités, 13402 Marseille, France.

*To whom correspondence should be addressed. E-mail: zhafed@salk.edu

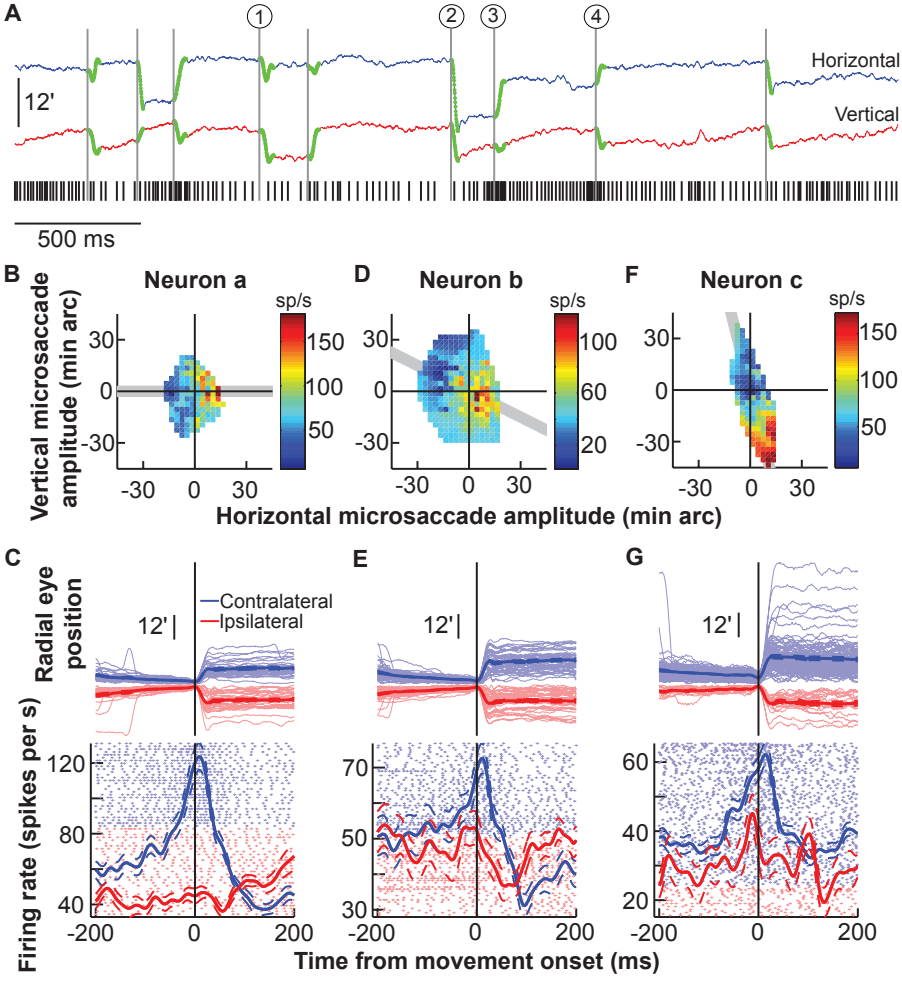


Fig. 1. Microsaccade-related activity in the SC. (**A**) Sample trial from one neuron. Microsaccades are highlighted in green on the eye position traces (*18*); the black tick marks indicate the neuron's spike times. The circled numbers highlight sample microsaccades that were associated with either a reduction in the neuron's activity (1 and 2) or a gradual build-up until movement onset (3 and 4). Upward deflections in the eye position traces denote right or up. (**B**) Peak discharge of the same neuron during all microsaccades plotted as a function of their horizontal and vertical amplitudes. The neuron preferred rightward movements sp, spikes. (**C**) Radial eye positions and neuronal activity for all microsaccades in the shaded gray region of (B) for the same neuron. Data are aligned on microsaccade onset and sorted by radial amplitude for contralateral (light blue) and ipsilateral (light red) movements. The vertical starting positions for successive microsaccade traces are also aligned but slightly offset from each other (by 0.06 min arc each) to facilitate visualization. Blue and red show average eye position or firing rate (with SEM envelopes). Light-colored dots show individual-movement spike rasters. (**D** and **E**) Similar to (B) and (C) but for another neuron, recorded from the same SC side, which preferred smaller movements directed to the lower right quadrant. (**F** and **G**) Similar to (B) and (C) but for a third neuron, from the same SC side, preferring larger movements (*22*).

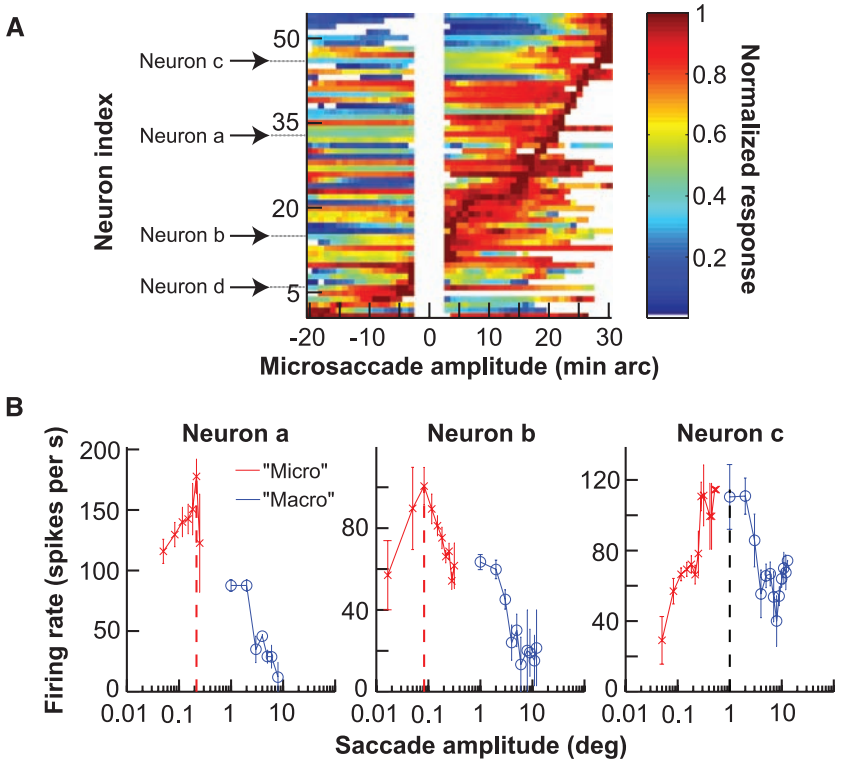


Fig. 2. Continuum of SC activity down to the smallest detectable saccades. (**A**) Peak discharge of each neuron during microsaccades plotted as a function of microsaccade amplitude along the neuron's preferred direction (positive is contralateral). Neurons are sorted on the basis of the amplitude that caused peak activity. Labeled neurons are the sample neurons of Fig. 1 and fig. S2 (for neuron d). White space in middle indicates unanalyzed bins (*18*). Each shown bin contains at least five microsaccades. (**B**) Comparison of neuronal activity during contralateral microsaccades along the neurons' preferred directions with activity during larger contralateral voluntary saccades. Some neurons were specifically tuned for microsaccades and paused or significantly reduced their activity for larger macrosaccades (neurons a and b). Other neurons continued to increase their discharge for slightly larger saccades (neuron c, fig. S3). Error bars denote SEM.

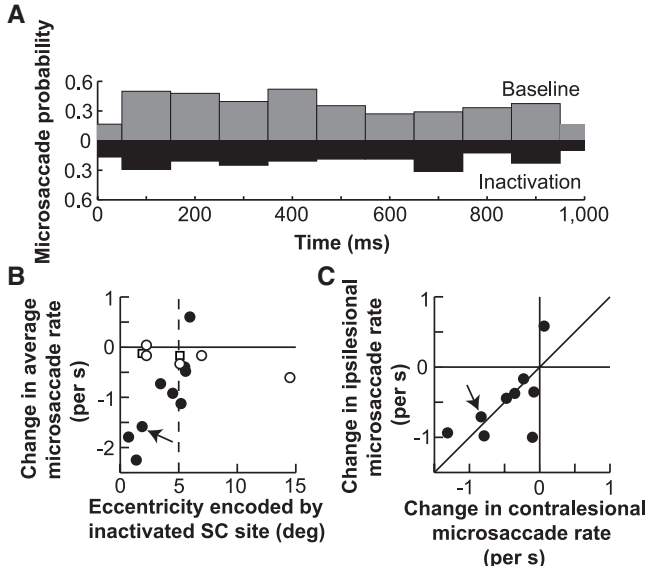


Fig. 3. A causal role for the SC in microsaccade generation. (**A**) Microsaccade probability during 1 s of fixation before (gray) and after (black) SC inactivation in one experiment [from monkey A (*18*)]. Inactivation reduced microsaccade rate. (**B**) Change in average microsaccade rate as a result of SC inactivation (average inactivation rate minus average baseline rate). Inactivation of the most central (i.e., rostral) SC sites caused the biggest reductions in microsaccades. Solid symbols indicate changes with $P < 0.05$ (rank-sum test, $N = 48$ trials for baseline, $N = 48$ trials for inactivation; average P value across solid symbols = 0.018). Squares indicate saline controls, which did not influence microsaccade rate. For a plot of absolute microsaccade rates before and after inactivation (also separated by monkey), see fig. S7. (**C**) Inactivation reduced both contralesional and ipsilesional microsaccades. For the experiments with a significant change in (B) ($N = 9/14$ experiments), data are plotted separately for contralesional and ipsilesional microsaccades. Eight of nine experiments fell in the lower left quadrant. Arrows in (B) and (C) point to the sample experiment of (A), performed at the same SC site as a subsequent saline control [square at the same eccentricity in (B)].

a burst during the movement itself. Thus, the neuron preferred particular microsaccade directions and amplitudes. We characterized these preferences across all microsaccades by plotting the neuron's activity as a function of the two-dimensional microsaccade amplitude, and we found consistent increases in activity for predominantly rightward microsaccades (Fig. 1B, neuron a). We also inspected the time course of the neuron's microsaccade-related activity. For this analysis, we included only microsaccades along the neuron's preferred axis (shaded region in Fig. 1B) to highlight activity associated with the preferred and nonpreferred directions, and we sorted the movements by radial amplitude to facilitate visualization (Fig. 1C). The neuron's activity was remarkably similar to the stereotypical saccade-related activity observed in caudal "build-up" neurons (*21*) during large "macro" saccades (see fig. S4, top, for samples), except that it happened for the smallest detectable eye movements.

Each neuron from the rostral SC exhibited individual preferences for a range of microsaccade directions and amplitudes. Figure 1, D to G, shows the activity of two additional sample neurons (neurons b and c), again recorded from the left SC, that preferred movements to the lower right quadrant. Neuron b showed a peak in discharge for ~6–min arc amplitudes and reduced its activity for smaller and larger movements. Thus, neuron b was tuned for saccade amplitudes smaller than those preferred by neuron a and also increased its activity for small leftward (i.e., ipsilateral) movements (Fig. 1D). Neuron c was tuned for amplitudes larger than those preferred by neuron a and increased its discharge for progressively larger microsaccades, up to ~44 min arc (Fig. 1F) (*22*).

Data from our population of neurons indicated that the SC contains a continuous representation of saccade amplitude and direction down to the smallest eye movements. We measured the activity of each neuron as a function of microsaccade amplitude along its preferred direction and sorted the neurons on the basis of their preferred microsaccade amplitudes (Fig. 2A). Neurons tiled the contralateral space of all measured microsaccade amplitudes and included some neurons whose activity was highest for ipsilateral movements as well. The raw activity of one of these neurons preferring ipsilateral microsaccades (labeled neuron d in Fig. 2A) is shown in fig. S2. The presence of such activity indicates that microsaccades involve activity distributed across the two SCs, functionally bridging the two visual hemi-fields to represent central targets.

Consistent with a continuous spatial representation throughout the SC (*21*), neuronal activity during microsaccades sometimes extended to small, voluntary saccades. Figure 2B illustrates this for the three sample neurons of Fig. 1 by plotting their activity as a function of amplitude during microsaccades, as well as during larger, visually guided saccades. This figure uses a logarithmic x-axis scale to magnify the representation of small microsaccades for easier visualization. Neurons a and b in Fig. 1 were specifically

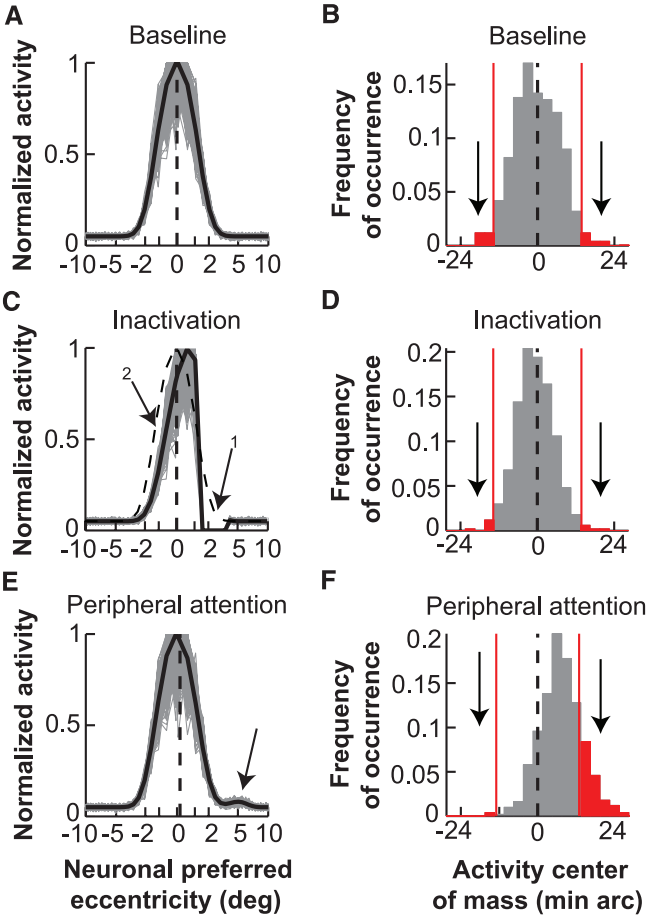


Fig. 4. A model of goal representation by the SC accounts for microsaccade generation. (**A**) Bilateral, one-dimensional SC map with activity representing the fixated goal. Activity is centered on neurons representing the central visual field (dashed black line). Gray traces show activity from 500 iterations of the model with each SC "neuron" exhibiting normally distributed activity (*18*). Negative values correspond to left, positive to right. (**B**) The instantaneous center of mass of SC activity in (A) was variable but had zero mean. Saccades are triggered when this center of mass deviates beyond a certain threshold (for example, red lines indicating 2 SD). (**C**) Simulating the effect of inactivating neurons representing ~2.5° (arrow 1). In the steady state, inactivation reduced the overall level of SC activity even on the intact side (arrow 2), but activity remained balanced bilaterally (vertical dashed line) (*24*). The dashed activity profile shows baseline. (**D**) The distribution of SC activity center of mass was narrower than in (B) but still had zero mean, explaining the reduction in both contralesional and ipsilesional microsaccades. For this simulation, there was a ~50% reduction in the frequency of deviations beyond the threshold of (B) (consistent with our experiments; fig. S7). (**E** and **F**) Simulating the momentary effects of a covert attention shift to a peripheral site at ~5° by introducing an increase in activity at this site (*18*). The average locus of activity was biased toward the peripheral site [dashed line in (E)], resulting in a higher probability of deviations from fixation toward the peripheral site than away from it (F).

tuned for movements with amplitudes classically associated with microsaccades, but neuron c was active for amplitudes that also included some voluntary movements larger than microsaccades. Similar observations were made for the remainder of our population. Neurons active during both microsaccades and voluntary saccades generally preferred voluntary saccades less than ~5° in amplitude (fig. S3). Neurons more caudal in the SC map (those preferring ~10° voluntary saccades in our data set) did not exhibit microsaccade-related activity (fig. S4).

SC activity during microsaccades was also distinct from modulations caused by small deviations of eye position from the fixated goal (*23*) (fig. S5), and it persisted in the absence of a visual stimulus (fig. S6).

We next investigated whether the SC is causally involved in microsaccade generation. Figure 3A shows the distribution of microsaccade occurrences during 1 s of steady-state fixation before and after reversibly inactivating a sample SC site. Inactivation caused a reduction in microsaccade rate (P = 0.0003; rank-sum test), and this was consistent across sessions, except when the inactivated SC neurons represented eccentricities larger than ~5° to 6° (Fig. 3B). This is consistent with our neuronal recordings that show a lack of microsaccade-related modulations for peripheral SC neurons (fig. S4). Inactivation reduced the probability of both contralesional and ipsilesional microsaccades (Fig. 3C), again consistent with our neuronal data showing activity for both contralateral and ipsilateral movements. Thus, rostral SC activity is directly involved in microsaccade generation.

A computational model, in which build-up neurons in the rostral SC encode foveal goal locations (*20*, *24*), provides a plausible mechanism for microsaccade generation (Fig. 4A). During steady-state fixation, the selected goal is foveal and results in bilateral activity centered on neurons representing the central visual field. The instantaneous locus of this activity may be viewed as a stochastic process with zero mean (Fig. 4B). According to this model, microsaccades are triggered when the center of mass of SC activity deviates sufficiently from zero.

Our model explains the inactivation-induced reduction in microsaccades. We simulated the effects of inactivation by eliminating the activity of a subset of model neurons (Fig. 4C, arrow 1). In steady state, inactivation reduces activity even on the intact side of the SC, without creating an imbalance across the right and left SC (*24*) (Fig. 4C). Thus, with the same stochastic processes influencing SC activity, the locus of the center of mass of this activity is less variable, which reduces the probability of deviation beyond the threshold for triggering saccades (Fig. 4D). Because the activity remains balanced bilaterally (*24*) (Fig. 4, C and D), inactivation reduces the frequency of both contralesional and ipsilesional microsaccades.

Our model can also explain why microsaccades are influenced by covert attention shifts (*8*, *9*). Cognitive factors such as attention and prior expectations alter build-up neuron activity (*25*, *26*); these changes could bias SC activity sufficiently to cause asymmetries in microsaccade generation. We therefore simulated the momentary effects on SC activity of attending to a peripheral location (Fig. 4E). These effects caused the average locus of SC activity to shift slightly toward the peripheral site, resulting in a higher probability of supra-threshold deviations toward the attended location (Fig. 4F).

Our results provide a description of the neural mechanisms for generating saccades smaller than 12 min arc. Coupled with the report that premotor neurons in the brainstem reticular formation are active during saccades as small as 12 to 15 min arc (*19*), our results demonstrate that microsaccades share the same neural mechanisms as voluntary saccades. Our proposal that microsaccade occurrence depends on the variability of SC activity representing salient goal locations reconciles seemingly disparate observations about microsaccades. Because SC activity can be manipulated by task set as well as by shifts of attention (*25*, *26*), this interpretation can explain why microsaccades might be voluntarily suppressed in some conditions (*6*) but asymmetrically increased in others (*8*, *9*) (Fig. 4F). Our interpretation also suggests that microsaccades may be part of a visual feedback loop. If the variability in target-related activity increases because of perceptual fading (*12*), the resulting microsaccades would correct for this fading through their influence on visual cortical activity (*27*–*30*). This in turn reduces the variability of SC signals defining the fixated target location. This mechanism can explain the increases in microsaccade rate and variability during fixation with no visual stimuli (*6*, *10*), as well as the behavioral correlations between increases or decreases in microsaccade rates and intensifying or fading visual percepts (*12*).

References and Notes

1. H. B. Barlow, *J. Physiol.* **116**, 290 (1952).
2. R. M. Steinman, G. M. Haddad, A. A. Skavenski, D. Wyman, *Science* **181**, 810 (1973).
3. B. L. Zuber, L. Stark, G. Cook, *Science* **150**, 1459 (1965).

4. B. Bridgeman, J. Palca, *Vision Res.* **20**, 813 (1980).
5. A. Kingstone, R. Fendrich, C. M. Wessinger, P. A. Reuter-Lorenz, *Percept. Psychophys.* **57**, 796 (1995).
6. R. M. Steinman, R. J. Cunitz, G. T. Timberlake, M. Herman, *Science* **155**, 1577 (1967).
7. B. J. Winterson, H. Collewijn, *Vision Res.* **16**, 1387 (1976).
8. Z. M. Hafed, J. J. Clark, *Vision Res.* **42**, 2533 (2002).
9. R. Engbert, R. Kliegl, *Vision Res.* **43**, 1035 (2003).
10. T. N. Cornsweet, *J. Opt. Soc. Am.* **46**, 987 (1956).
11. E. Kowler, R. M. Steinman, *Vision Res.* **20**, 273 (1980).
12. S. Martinez-Conde, S. L. Macknik, X. G. Troncoso, T. A. Dyar, *Neuron* **49**, 297 (2006).
13. F. Ratliff, L. A. Riggs, *J. Exp. Psychol.* **40**, 687 (1950).
14. S. Martinez-Conde, S. L. Macknik, D. H. Hubel, *Nat. Rev. Neurosci.* **5**, 229 (2004).
15. R. M. McPeek, E. L. Keller, *Nat. Neurosci.* **7**, 757 (2004).
16. C. D. Carello, R. J. Krauzlis, *Neuron* **43**, 575 (2004).
17. R. J. Krauzlis, *J. Neurosci.* **23**, 4333 (2003).
18. Materials and methods, including behavioral tasks, single-neuron recordings, and inactivations, are available as supporting material on *Science* Online.
19. J. A. M. Van Gisbergen, D. A. Robinson, S. A. Gielen, *J. Neurophysiol.* **45**, 417 (1981).
20. Z. M. Hafed, R. J. Krauzlis, *J. Neurosci.* **28**, 9426 (2008).
21. D. P. Munoz, R. H. Wurtz, *J. Neurophysiol.* **73**, 2313 (1995).
22. Because microsaccades are not under full experimental control, they posed special challenges, described in more detail in (*18*).
23. R. J. Krauzlis, M. A. Basso, R. H. Wurtz, *J. Neurophysiol.* **84**, 876 (2000).
24. Z. M. Hafed, L. Goffart, R. J. Krauzlis, *J. Neurosci.* **28**, 8124 (2008).
25. A. Ignashchenkova, P. W. Dicke, T. Haarmeier, P. Thier, *Nat. Neurosci.* **7**, 56 (2004).
26. M. A. Basso, R. H. Wurtz, *J. Neurosci.* **18**, 7519 (1998).
27. S. Martinez-Conde, S. L. Macknik, D. H. Hubel, *Nat. Neurosci.* **3**, 251 (2000).
28. D. A. Leopold, N. K. Logothetis, *Exp. Brain Res.* **123**, 341 (1998).
29. D. M. Snodderly, I. Kagan, M. Gur, *Vis. Neurosci.* **18**, 259 (2001).
30. W. Bair, L. P. O'Keefe, *Vis. Neurosci.* **15**, 779 (1998).
31. Funding provided by NIH grant EY12212 and Agence Nationale de la Recherche grant RETINAE. We thank K. Nielsen and V. Ciaramitaro for helpful comments.

Supporting Online Material
www.sciencemag.org/cgi/content/full/323/5916/940/DC1
Materials and Methods
Figs. S1 to S7

18 September 2008; accepted 10 December 2008
10.1126/science.1166112

An ABC Transporter Controls Export of a *Drosophila* Germ Cell Attractant

Sara Ricardo and Ruth Lehmann*

Directed cell migration, which is critical for embryonic development, leukocyte trafficking, and cell metastasis, depends on chemoattraction. 3-hydroxy-3-methylglutaryl coenzyme A reductase regulates the production of an attractant for *Drosophila* germ cells that may itself be geranylated. Chemoattractants are commonly secreted through a classical, signal peptide–dependent pathway, but a geranyl-modified attractant would require an alternative pathway. In budding yeast, pheromones produced by a-cells are farnesylated and secreted in a signal peptide–independent manner, requiring the adenosine triphosphate–binding cassette (ABC) transporter Ste6p. Here we show that *Drosophila* germ cell migration uses a similar pathway, demonstrating that invertebrate germ cells, like yeast cells, are attracted to lipid-modified peptides. Components of this unconventional export pathway are highly conserved, suggesting that this pathway may control the production of similarly modified chemoattractants in organisms ranging from yeast to humans.

In most organisms, primordial germ cells migrate from their site of origin to the somatic part of the gonad, where they develop into mature eggs and sperm. In *Drosophila*, germ cells migrate as single cells in a stereotyped manner and are guided by repellent and attractive cues toward the somatic gonad in the mesoderm (*1*). 3-hydroxyl-3-methyl-glutaryl-CoA reductase (HMG-CoAr or HMGCR) activity controls germ cell attraction to the mesoderm and the recruitment of germ cells to the somatic gonad. In *hmgcr* mutant embryos, germ cells fail to reach the somatic gonad; moreover, ectopic HMGCR expression is sufficient to attract germ cells to a new location (*2*). Embryos mutant for several enzymes in the *hmgcr* pathway, including the β subunit of type I geranygeranyl transferase (βGGTI), are similarly defective in germ cell migration, suggesting that geranylation is critical in attracting germ cells to the mesoderm (*3*).

Because *hmgcr* or *βggtI* mutant embryos show a rather specific germ cell migration defect, we favored the idea that the HMGCR pathway was required to geranylate a critical germ cell attractant rather than regulating a pathway that controlled synthesis or secretion of the attractant (*3*). This idea posits secretion of a geranyl-modified germ cell attractant, which would preclude secretion by a classic, signal peptide–dependent secretory pathway and require an alternative export mechanism. Such a mechanism has been described in yeast, where adenosine triphosphate (ATP)–binding cassette (ABC) transporters export farnesylated pheromones required for cell mating (*4*, *5*). We therefore asked whether a similar export mechanism exists in *Drosophila* and is required for germ cell attraction.

ABC transporters are conserved in organisms ranging from bacteria to humans and shuttle hydrophobic lipophilic compounds in an ATP-dependent manner (*6*, *7*). In *Saccharomyces cerevisiae* and *Schizosaccharomyces pombe*, the ABCB family members Ste6p and Mam1 play essential roles in the export of farnesyl-modified a-type and M-type mating factors, respectively. The human ABCB family member *mdr1* (*multidrug resistance*) gene is amplified in multidrug-resistant cells, and its homolog *mdr3* is a functional homolog of *STE6* (*8*, *9*), suggesting a close relationship between the ability of this class of transporters to export drugs and lipid-modified signaling molecules. To determine whether ABCB transporters have a role in exporting the putative *Drosophila* germ cell attractant, we analyzed expression patterns and germ cell migration in embryos mutant for ABCB transporters (*7*) (table S1). Among these, only *mdr49* showed an expression pattern and mutant phenotype consistent with a role in germ cell migration (Fig. 1A, fig. S1, and table S2). We generated a strong loss-of-function allele, $mdr49^{\Delta 3.16}$ (Fig. 1B, fig. S1A, and table S2), and $mdr49^{\Delta 3.16}$ mutant embryos showed defects in germ cell migration, like embryos in which the HMGCR pathway was mutant: Germ cells migrated through the posterior midgut but then failed to associate with the somatic gonad (Fig. 1C), which was properly specified (fig. S1E). This migration phenotype was observed only in *mdr49* mutants and not in mutants for other ABCB transporters, such as Mdr50, Mdr65, and CG7955 (table S1). To determine whether *mdr49* function is required in the mesoderm, we restored *mdr49* expression selectively in the mesoderm (fig. S1F) and found that it fully rescued the *mdr49* mutant phenotype (Fig. 1D). To test whether Mdr49 acts as an ABCB transporter, we asked whether Ste6p rescued the germ cell migration phenotype observed in *mdr49* embryos. Expressing *STE6* in the mesoderm rescued the migration defect (Fig. 1D and fig. S1F), suggesting that Mdr49 is functionally equivalent to Ste6p and acts in mesodermal cells to attract germ cells.

The germ cell migration phenotypes caused by mutations in Mdr49 and in components of the HMGCR pathway, such as geranylgeranyl-diphosphate synthase and ßGGTI, are strikingly similar (*2*, *3*). Thus, to determine whether Mdr49 acts as a transporter for a germ cell attractant that is geranylgeranyl-modified by the HMGCR pathway, we tested the genetic epistasis between *hmgcr* and *mdr49*. Previous experiments showed that overexpression of *hmgcr* in the central nervous system (CNS) is sufficient to attract germ cells to this tissue (*2*). We therefore reasoned that Mdr49 function should be necessary for the export of the ectopically produced attractant and that mutations in *mdr49* should suppress the *hmgcr* misexpression phenotype. Indeed, germ cell migration to the CNS was suppressed by reducing *mdr49* copy number by means of either the *mdr49* deficiency or P-element mutation. (Fig. 2 and table S3). Mutations in other ABCB transporters, such as *mdr50*, *mdr65*, and *CG7955*, did not significantly suppress the *hmgcr* overexpression pheno-

Howard Hughes Medical Institute (HHMI), Helen L. and Martin S. Kimmel Center for Biology and Medicine at the Skirball Institute, Department of Cell Biology, New York University School of Medicine, New York University, 540 First Avenue, New York, NY 10016, USA.

*To whom correspondence should be addressed. E-mail: lehmann@saturn.med.nyu.edu

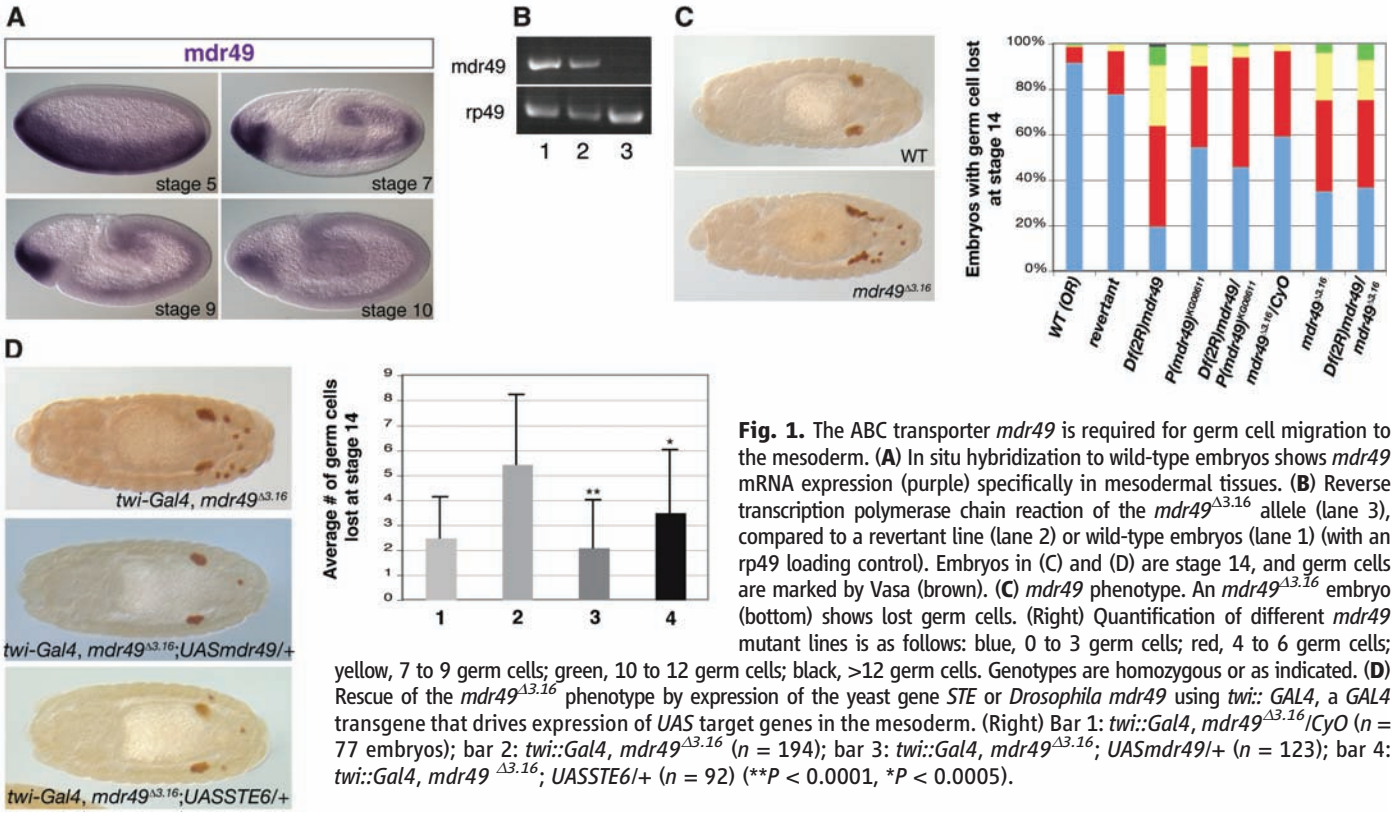


Fig. 1. The ABC transporter *mdr49* is required for germ cell migration to the mesoderm. (**A**) In situ hybridization to wild-type embryos shows *mdr49* mRNA expression (purple) specifically in mesodermal tissues. (**B**) Reverse transcription polymerase chain reaction of the $mdr49^{\Delta 3.16}$ allele (lane 3), compared to a revertant line (lane 2) or wild-type embryos (lane 1) (with an rp49 loading control). Embryos in (C) and (D) are stage 14, and germ cells are marked by Vasa (brown). (**C**) *mdr49* phenotype. An $mdr49^{\Delta 3.16}$ embryo (bottom) shows lost germ cells. (Right) Quantification of different *mdr49* mutant lines is as follows: blue, 0 to 3 germ cells; red, 4 to 6 germ cells; yellow, 7 to 9 germ cells; green, 10 to 12 germ cells; black, >12 germ cells. Genotypes are homozygous or as indicated. (**D**) Rescue of the $mdr49^{\Delta 3.16}$ phenotype by expression of the yeast gene *STE* or *Drosophila mdr49* using *twi:: GAL4*, a *GAL4* transgene that drives expression of *UAS* target genes in the mesoderm. (Right) Bar 1: *twi::Gal4*, $mdr49^{\Delta 3.16}$/*CyO* (n = 77 embryos); bar 2: *twi::Gal4*, $mdr49^{\Delta 3.16}$ (n = 194); bar 3: *twi::Gal4*, $mdr49^{\Delta 3.16}$; *UASmdr49*/+ (n = 123); bar 4: *twi::Gal4*, $mdr49^{\Delta 3.16}$; *UASSTE6*/+ (n = 92) (**P < 0.0001, *P < 0.0005).

type (table S1), which is consistent with a specific role for Mdr49 as an exporter for an HMGCR-dependent geranyl-modified germ cell attractant.

Prenylated proteins require additional modifications to be fully functional: cleavage of the AAX residues (A, aliphatic amino acid; X, any amino acid), prenyl proteolysis, and carboxymethylation (*10*) (Fig. 3A). We therefore asked whether the enzymes that catalyze these reactions are present in *Drosophila* and are required for germ cell migration. Prenyl proteolysis is catalyzed by Ste24p (prenyl protease type I) and Rce1p (prenyl protease type II), with the former being essential for a-factor modification (*11*, *12*). Carboxymethylation is achieved exclusively through the activity of Ste14p (*13*). Orthologs of these enzymes are also found in mammals, where they have similar substrates (table S4). A genomic search in *Drosophila* revealed predicted orthologs for the following: (i) prenyl protease type I as three proteins encoded by a cluster of genes (*CG9000/CG9001/CG9002*), which we term Dmel\Ste24a, -b, and -c, respectively; (ii) prenyl protease type II, termed Sras; and (iii) the isoprenylcysteine carboxylmethyltransferase as CG11268, termed Dmel\Ste14 (table S4 and fig. S2).

We next assessed the role of each enzyme in germ cell migration. Because specific mutations were not available for these enzymes, we analyzed deficiencies that deleted each gene [as well as other genes (*14*)]. Embryos with homozygous mutations in *Df(2R)Dmel\ste24*, which deletes the three *Drosophila ste24* genes, showed a specific germ cell migration phenotype similar to that of *mdr49* mutant embryos. On average, six germ cells

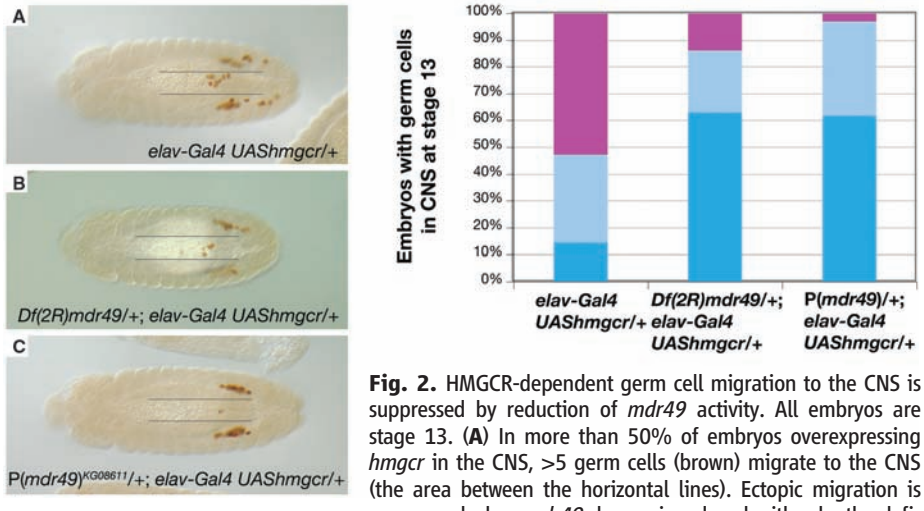


Fig. 2. HMGCR-dependent germ cell migration to the CNS is suppressed by reduction of *mdr49* activity. All embryos are stage 13. (**A**) In more than 50% of embryos overexpressing *hmgcr* in the CNS, >5 germ cells (brown) migrate to the CNS (the area between the horizontal lines). Ectopic migration is suppressed when *mdr49* dosage is reduced, either by the deficiency (**B**) or a P-element mutation (**C**). (Right) Quantification of germ cells in the CNS. Blue, 0 to 3 germ cells; light blue, 4 to 5 germ cells; magenta, >5 germ cells.

did not associate with the somatic gonad (Fig. 3B and table S2). Next, we asked whether the single *Drosophila* isoprenylcysteine carboxylmethyltransferase, *Dmel\ste14*, is required for germ cell migration. We were unable to analyze the germ cell migration phenotype of embryos homozygous for the deletion *Df(3L)ED4486* (*Df(3L)Dmel\ste14*) because of embryonic patterning defects, which resembled those observed in embryos with mutant *Drosophila* Rho GTPases and gave additional evidence that *Dmel\ste14* encodes a general postprenylation processing enzyme that modifies all prenylated proteins (*15*–*17*) (fig. S4). Instead of analyzing homozygous mutants, we asked whether reducing *Dmel\ste14* gene dosage was able to suppress germ cell mis-migration induced by ectopic expression of *hmgcr* in the CNS. As observed for the ABC transporter *mdr49*, reducing *Dmel\ste14* gene dosage significantly suppressed HMGCR-dependent ectopic germ cell migration, suggesting a requirement for this enzyme in attractant modification (Fig. 3C and table S5).

Our genetic analyses strongly suggest that a geranylgeranylated germ cell attractant is produced

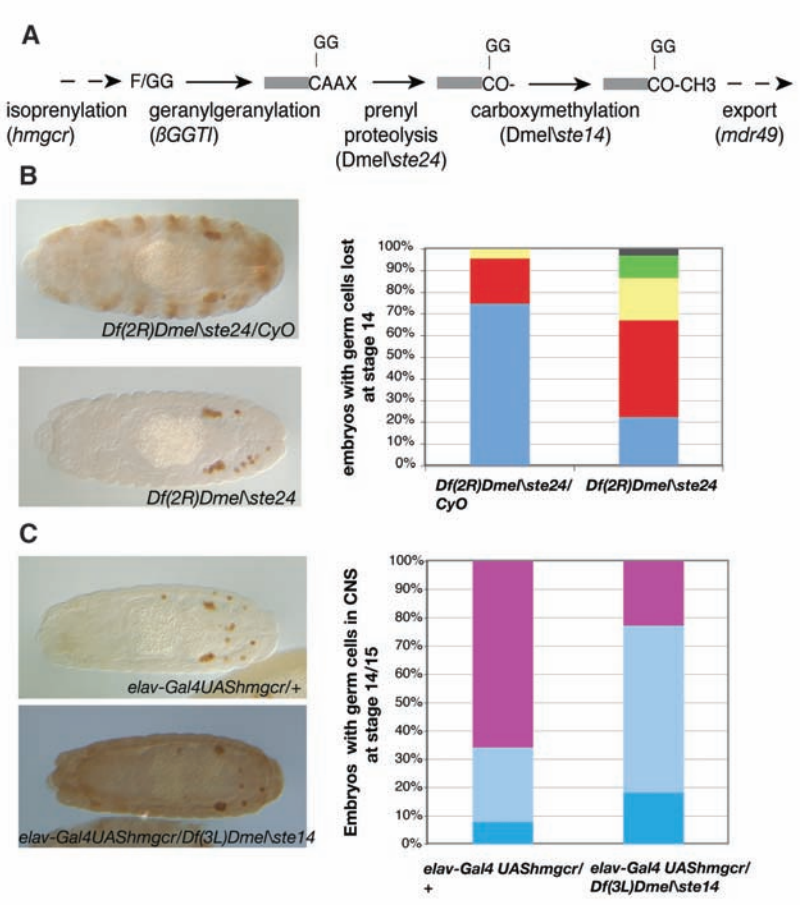


Fig. 3. Type I prenyl proteases and isoprenylcysteine carboxymethyltransferases are required for germ cell migration to the mesoderm. (**A**) A schematic of the prenyl processing pathway analyzed. CAAX (C, Cys) represents the prenylation motif to which a farnesyl (F) or geranylgeranyl (GG) group is added. (**B**) Embryos carrying a deficiency of *Dmel\ste24* show germ cells (brown) lost in the posterior mesoderm. Sibling control (top) and mutant (bottom) embryos are shown at left, and quantification is shown at right (blue, 0 to 3 germ cells; red, 4 to 6 germ cells; yellow, 7 to 9 germ cells; green, 10 to 12 germ cells; black, >12 germ cells). (**C**) *Df(3L)Dmel\ste14* suppresses *hmgcr*-dependent ectopic germ cell migration. Sibling control (top) and suppressed (bottom) embryos are shown at left and quantification at right (germ cells at ectopic sites). Blue, 0 to 5 germ cells; light blue, 6 to 10 germ cells; magenta, >10 germ cells. All embryos are at stage 14/15.

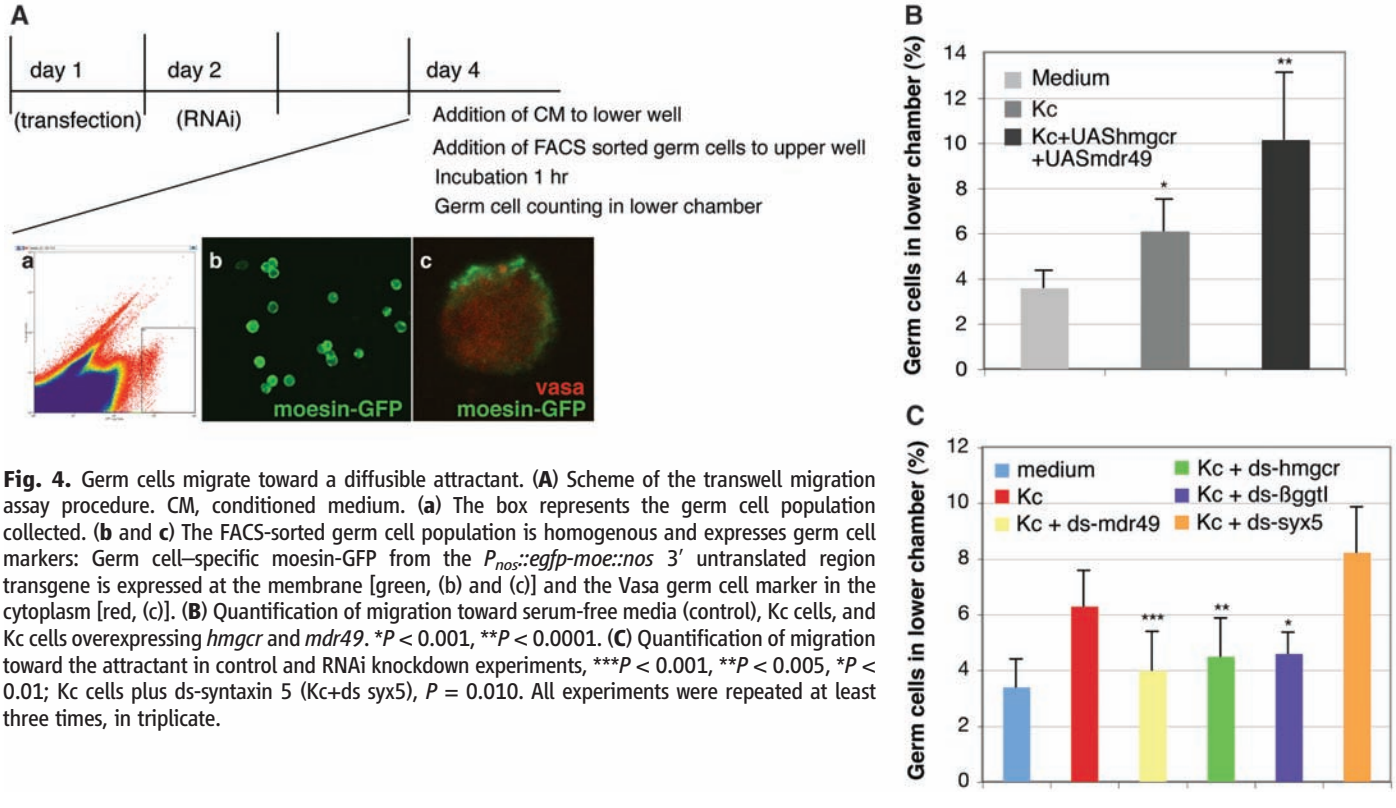


Fig. 4. Germ cells migrate toward a diffusible attractant. (**A**) Scheme of the transwell migration assay procedure. CM, conditioned medium. (**a**) The box represents the germ cell population collected. (**b** and **c**) The FACS-sorted germ cell population is homogenous and expresses germ cell markers: Germ cell–specific moesin-GFP from the P_{nos}*::egfp-moe::nos* 3′ untranslated region transgene is expressed at the membrane [green, (b) and (c)] and the Vasa germ cell marker in the cytoplasm [red, (c)]. (**B**) Quantification of migration toward serum-free media (control), Kc cells, and Kc cells overexpressing *hmgcr* and *mdr49*. *$P < 0.001$, **$P < 0.0001$. (**C**) Quantification of migration toward the attractant in control and RNAi knockdown experiments, ***$P < 0.001$, **$P < 0.005$, *$P < 0.01$; Kc cells plus ds-syntaxin 5 (Kc+ds syx5), $P = 0.010$. All experiments were repeated at least three times, in triplicate.

and modified in the somatic gonad and exported via an ABCB transporter. In order to more directly test for the production of a molecule that can act as a diffusible attractant, we devised an in vitro germ cell migration assay. We used germ cells sorted by fluorescence-activated cell sorting (FACS) from embryos 2 to 10 hours old that expressed moesin–green fluorescent protein (GFP) specifically in germ cells (*18*). Germ cells were placed in the upper well of a transwell chemotaxis chamber and scored for their migration toward medium conditioned with secreted proteins (conditioned medium) from Kc cells, an insect cell line (*19*), in the lower well (Fig. 4A). Germ cells migrated to the bottom well containing conditioned medium from cells overexpressing *hmgcr* and *mdr49* (Fig. 4B). Fewer cells moved toward unconditioned control medium or medium conditioned from parental Kc cells, which express low levels of *hmgcr* and *mdr49* (*19*) (fig. S5A). To determine whether migration toward parental Kc cells is dependent on *hmgcr* and *mdr49*, we used RNA interference (RNAi) to reduce the expression of HMGCR pathway members in Kc cells. We found that reducing *hmgcr*, β*GGTI*, and ABC transporter (*mdr49*) expression fully blocked germ cell migration toward Kc cell–conditioned medium (Fig. 4C and fig. S5B). These results are consistent with our genetic data and support the notion that the prenylated *Drosophila* germ cell attractant is active in a secreted form.

To test whether the classic, signal peptide–dependent pathway is also required for secretion of the germ cell attractant, we used RNAi to reduce the levels of *syntaxin 5*, which encodes a SNARE (SNAP and NSF attachment receptor) protein essential for constitutive secretion (*20*). Germ cells were similarly attracted to conditioned medium from parental cells and *syntaxin 5*–deficient Kc cells, indicating that production of the attractant does not depend upon constitutive secretion (Figs. 4C and S5A). Our results demonstrate that the *Drosophila* germ cell attractant is geranylgeranylated and secreted by mesodermal cells in a signal peptide-independent manner through an ABCB transporter. The modifications, processing, and export pathway used to generate and secrete the germ cell attractant strikingly resemble that of a-factor in mating yeast.

ABC transporters have been studied mostly in the context of cancer drug resistance or their role in toxin protection. Our findings reveal a new function for this conserved, non-conventional secretory pathway in a multicellular organism and suggest that this pathway may be used to export signals required for cell-cell communication in organisms other than *Drosophila* and yeast (*21*–*23*). Yeast lacks many of the secreted cell signaling molecules, such as Hedgehog, Wnt, and bone morphogenetic protein, used by multicellular organisms to communicate between cells. The use of a prenylated signal may thus be an ancient mechanism of cell communication. It is striking that this pathway is used in yeast and flies to facilitate the migration and adhesion of germ cells, the essential cells for reproduction.

References and Notes

1. A. C. Santos, R. Lehmann, *Curr. Biol.* **14**, R578 (2004).
2. M. Van Doren, H. T. Broihier, L. A. Moore, R. Lehmann, *Nature* **396**, 466 (1998).
3. A. C. Santos, R. Lehmann, *Dev. Cell* **6**, 283 (2004).
4. P. U. Christensen, J. Davey, O. Nielsen, *Mol. Gen. Genet.* **255**, 226 (1997).
5. J. P. McGrath, A. Varshavsky, *Nature* **340**, 400 (1989).
6. C. F. Higgins, *Annu. Rev. Cell Biol.* **8**, 67 (1992).
7. M. Dean, Y. Hamon, G. Chimini, *J. Lipid Res.* **42**, 1007 (2001).
8. M. Raymond, P. Gros, M. Whiteway, D. Y. Thomas, *Science* **256**, 232 (1992).
9. I. B. Roninson *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **83**, 4538 (1986).
10. A. M. Winter-Vann, P. J. Casey, *Nat. Rev. Cancer* **5**, 405 (2005).
11. A. Tam *et al.*, *J. Cell Biol.* **142**, 635 (1998).
12. V. L. Boyartchuk, M. N. Ashby, J. Rine, *Science* **275**, 1796 (1997).
13. C. A. Hrycyna, S. K. Sapperstein, S. Clarke, S. Michaelis, *EMBO J.* **10**, 1699 (1991).
14. Materials and methods are available as supporting material on *Science* Online.
15. J. E. Johndrow, C. R. Magie, S. M. Parkhurst, *Biochem. Cell Biol.* **82**, 643 (2004).
16. J. Settleman, *Dev. Cell* **1**, 321 (2001).
17. C. R. Magie, M. R. Meyer, M. S. Gorsuch, S. M. Parkhurst, *Development* **126**, 5353 (1999).
18. H. Sano, A. D. Renault, R. Lehmann, *J. Cell Biol.* **171**, 675 (2005).
19. G. Echalier, A. Ohanessian, *C. R. Acad. Sci. Hebd. Seances Acad. Sci. D* **268**, 1771 (1969).
20. F. Bard *et al.*, *Nature* **439**, 604 (2006).
21. A. H. Schinkel, *Semin. Cancer Biol.* **8**, 161 (1997).
22. J. J. Smit *et al.*, *Cell* **75**, 451 (1993).
23. G. K. Leung *et al.*, *J. Biol. Chem.* **276**, 29051 (2001).

24. We thank P. Lopez and G. de la Cruz for FACS sorting; S. Michaelis for reagents and discussions; J. Treisman, Blomington stock center, and the Drosophila Genomics Resource Center for fly lines and reagents; the Lehmann Lab; and D. Siekhaus and S. Burden for comments on the manuscript. This work was supported by NIH grant HD49100. S.R. is a HHMI research associate; R.L. is an HHMI investigator and a member of the Kimmel Stem Cell Center.

Supporting Online Material

www.sciencemag.org/cgi/content/full/323/5916/943/DC1
Materials and Methods
SOM Text
Figs. S1 to S6
Tables S1 to S5
References

22 September 2008; accepted 5 December 2008
10.1126/science.1166239

The Orphan G Protein–Coupled Receptor 3 Modulates Amyloid-Beta Peptide Generation in Neurons

Amantha Thathiah,[1,2] Kurt Spittaels,[3]* Marcel Hoffmann,[3]† Mik Staes,[3]‡ Adrian Cohen,[3]§ Katrién Horré,[1,2] Mieke Vanbrabant,[1,2] Frea Coun,[4] Veerle Baekelandt,[4] André Delacourte,[6] David F. Fischer,[5]|| Dirk Pollet,[3] Bart De Strooper,[1,2]|| Pascal Merchiers[3]¶

Deposition of the amyloid-β peptide is a pathological hallmark of Alzheimer's disease. A high-throughput functional genomics screen identified G protein–coupled receptor 3 (GPR3), a constitutively active orphan G protein–coupled receptor, as a modulator of amyloid-β production. Overexpression of GPR3 stimulated amyloid-β production, whereas genetic ablation of GPR3 prevented accumulation of the amyloid-β peptide in vitro and in an Alzheimer's disease mouse model. GPR3 expression led to increased formation and cell-surface localization of the mature γ-secretase complex in the absence of an effect on Notch processing. GPR3 is highly expressed in areas of the normal human brain implicated in Alzheimer's disease and is elevated in the sporadic Alzheimer's disease brain. Thus, GPR3 represents a potential therapeutic target for the treatment of Alzheimer's disease.

Accumulation of the amyloid-β peptide (Aβ) is a central pathological feature in the brain of Alzheimer's disease (AD) patients (*1*). Aβ is generated after sequential cleavage of the β-amyloid precursor protein (APP) by the β- and γ-secretases (*2*–*4*), whereas initial cleavage by the α-secretase within the Aβ sequence precludes Aβ generation. All three secretases are considered to be relevant therapeutic targets for AD. Here, we sought to identify modulators of Aβ production using a high-throughput functional genomics screen, the full-length human FLeXSelect cDNA library (table S1), which is composed of 4217 individual adenoviruses, representing the transcripts of 1905 unique genes encoding potential drug targets (*5*). Several primary hits affecting Aβ secretion were identified (Fig. 1A) and confirmed using independent viral stocks. Genes previously shown to modulate Aβ levels or APP processing were identified: APP, the serotonin receptor HTR2C (*6*), and the prostaglandin E2 receptor PTGER2 (*7*) led to an increase in $A\beta_{1\text{-}42}$ production, whereas BACE2 repressed Aβ production (*8*, *9*). Several other Aβ-modulating targets were identified and confirmed in the human SH-SY5Y neuroblastoma cell line. Secondary assays, including a secreted alkaline phosphatase (SEAP) assay to exclude proteins that affect general secretion or transport mechanisms (fig. S3A), RNA interference (RNAi)–mediated knockdown experiments (Fig. 1D and fig. S2), real-time reverse transcription polymerase chain reaction (RT-PCR) analysis of human brain RNA (tables S4 and S5), and mass spectrometric analysis of immunoprecipi-

tated Aβ peptides (fig. S4), were used to prioritize the targets. Among the candidate targets, G protein–coupled receptor 3 (GPR3) was of particular interest because the GPR3 gene has been mapped to the candidate AD linkage region on chromosome 1p36.1-p34.3 (*10*, *11*), and GPR3 is predominantly expressed in the central nervous system (*10*, *12*). Given that Ad-GPR3 transduction of SH-SY5Y or human embryonic kidney (HEK) 293 APP_{770} cells led to a robust increase in $Aβ_{1-40}$ and $Aβ_{1-42}$ release in a dose-responsive manner (Fig. 1, B and C), we assessed the functional relevance of reduced GPR3 expression on Aβ production using an RNAi-mediated strategy. Specific siRNA duplexes efficiently suppressed $Aβ_{1-42}$ secretion to 50% below control levels (Fig. 1D). Thus, endogenous GPR3 plays a role in the regulation of secreted Aβ without affecting SEAP release (fig. S3B).

To investigate whether GPR3 regulates Aβ secretion through modulation of β-secretase activity, we determined the expression and activity of the β-secretase after Ad-GPR3 transduction. Immunoprecipitation–mass spectrometry analysis did not indicate increased β-secretase activity (fig. S4), and β-secretase expression levels were unchanged after Ad-GPR3 transduction (see below), which indicates that GPR3 modulation of Aβ generation occurs downstream of β-secretase activity. Expression of APP-C99, a direct γ-secretase substrate and immediate precursor of Aβ, together with Ad-GPR3 in primary hippocampal neurons yielded substantial increases in $Aβ_{1-40}$ and $Aβ_{1-42}$ release (Fig. 2, A and B), which suggests that GPR3-mediated Aβ generation occurs through modulation of the γ-secretase cleavage of APP without affecting expression of the γ-secretase subunits (fig. S5A). Furthermore, L-685,458, a highly selective γ-secretase inhibitor (*13*), effectively abolished the increase in secreted $Aβ_{1-40}$ and $Aβ_{1-42}$ upon GPR3 overexpression (Fig. 2C and fig. S5B).

GPR3 is an orphan G protein–coupled receptor (GPCR), although a putative ligand has been identified (*14*), and GPR3 constitutively elevates cyclic adenosine monophosphate (cAMP) levels through adenylate cyclase activation (*15*, *16*), which implies that it intrinsically activates the G protein G_s. Elevation of intracellular cAMP levels after expression of a constitutively active mutant of the thyroid stimulating hormone receptor (TSHr), another GPCR, failed to stimulate Aβ release in HEK293 APP_{695} cells (fig. S13C).

[1]Department of Molecular and Developmental Genetics, Vlaams Institute for Biotechnology. [2]Center for Human Genetics, Catholic University of Leuven, Herestraat 49, 3000 Leuven, Belgium. [3]Galapagos, Generaal De Wittelaan L11 A3, 2800 Mechelen, Belgium. [4]Laboratory for Neurobiology and Gene Therapy, Division of Molecular Medicine, Catholic University of Leuven, Kapucijnenvoer 33, 3000 Leuven. [5]BioFocus DPI, a Galapagos company, Darwinweg 24, 2333 CR Leiden, Netherlands. [6]INSERM Unit 837, Jean-Pierre Aubert Research Center, 1 Place de Verdun, 59045 Lille Cedex, France.

*Present address: Tibotec BVBA, Generaal De Wittelaan L11B3, 2800 Mechelen, Belgium.
†Present address: Avantium Technologies B.V., Zekeringstraat 29, 1014 BV Amsterdam, Netherlands.
‡Present address: Terumo Europe NV, Interleuvenlaan 40, 3001 Leuven, Belgium.
§Present address: Nijmegen Centre for Molecular Life Sciences, Radboud University Nijmegen Medical Centre, Geert Grooteplein 28, 6500HB Nijmegen, Netherlands.
‖To whom correspondence should be addressed. E-mail: david.fischer@glpg.com (D.F.F.); bart.destrooper@med.kuleuven.be (B.D.S.)
¶Present address: Ablynx NV, Technologie Park 4, 9052 Zwijnaarde, Belgium.

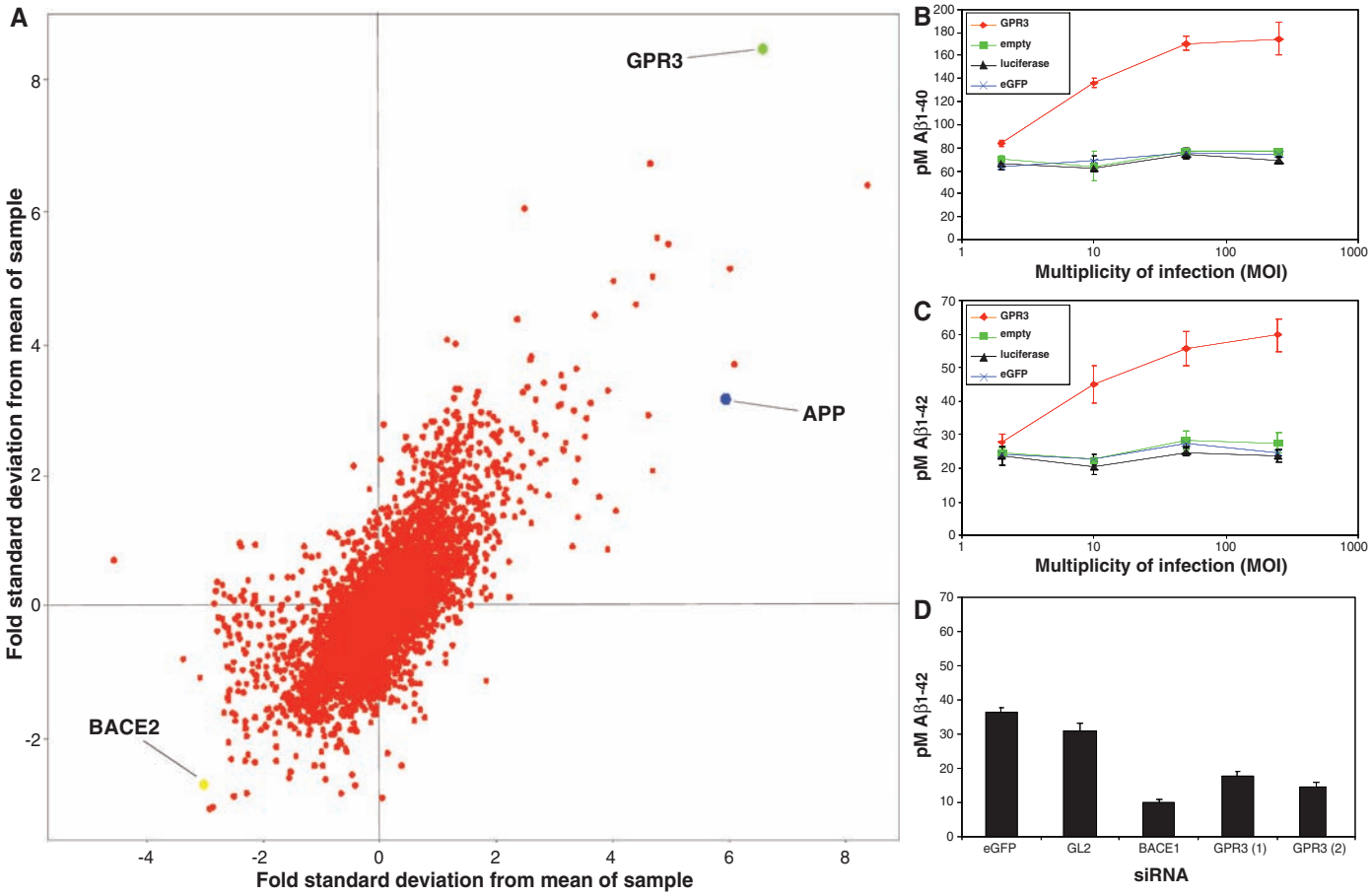


Fig. 1. GPR3 modulates Aβ generation. (**A**) Duplicate screening data were plotted against each other and expressed as relative (fold) standard deviation based on the mean of the sample. (**B** and **C**) Overexpression of GPR3 increased secretion of $Aβ_{1-40}$ and $Aβ_{1-42}$ peptides in a dose-responsive manner in HEK293 APP_{770} cells. The mean ± SD (n = 4) is shown. (**D**) Two independent siRNAs (D-003951-01 and D-003951-02) directed against GPR3 suppressed $Aβ_{1-42}$ secretion to below enhanced GFP (eGFP) and GL2 (luciferase) control levels in HEK293 APP_{770} cells. siRNAs directed against BACE1 were used as a positive control. The mean ± SD (n = 5) is shown.

Similarly, expression of a G_sDN mutant or inactivation of $G_{s\alpha}$ failed to reduce the GPR3-mediated increase in Aβ generation (figs. S13 and S14). We then examined the potential effect of GPR3 expression on the assembly and/or cell surface expression of the γ-secretase complex subunits NCT, PS1-NTF, PS1-CTF, APH-1a, and PEN-2 (*17*). NCT immunostaining typically revealed three bands using blue native polyacrylamide gel electrophoresis (BN-PAGE) (Fig. 2D). Two bands between 150- and 440-kD represent subcomplexes of NCT–APH-1 and NCT–APH-1–PS1-CTF, respectively. The ~440-kD band represents the mature γ-secretase complex (*18*). GPR3 overexpression led to an increase in the relative amount of mature γ-secretase complex (Fig. 2D). Furthermore, we observed increased cell surface expression of the γ-secretase subunits after Ad-GPR3 transduction (Fig. 2E).

GPCRs and the γ-secretase complex have been shown to be localized in detergent-resistant membranes (DRMs) (*19*, *20*). Thus, we determined the localization of the γ-secretase subunits and GPR3 by differential flotation after sucrose density gradient centrifugation (Fig. 2F). Notably, the γ-secretase subunits and GPR3 partially co-distributed in low-density fractions 3 and 4. Moreover, GPR3 overexpression resulted in an enrichment of the γ-secretase subunits in low-buoyancy fractions, which suggests that expression of GPR3 leads to increased localization of the γ-secretase complex in DRMs.

The most severe side effects due to absence of γ-secretase activity are caused by deficient Notch signaling (*21*). Consequently, intramembranous cleavage of Notch was evaluated in HEK293 wild-type (WT) cells using a truncated form of the Notch receptor, Myc-tagged Ad-NΔE, that undergoes cleavage by the γ-secretase, releasing the Notch intracellular cytoplasmic domain (NICD) (*22*). Using an antibody against cleaved Notch (Val 1744), which recognizes a neoepitope in the NICD after γ-secretase cleavage, we determined that similar levels of the NICD were generated in green fluorescent protein (GFP)– and GPR3-expressing cells (Fig. 3A). The selectivity of GPR3 modulation of γ-secretase activity with regard to APP relative to Notch was

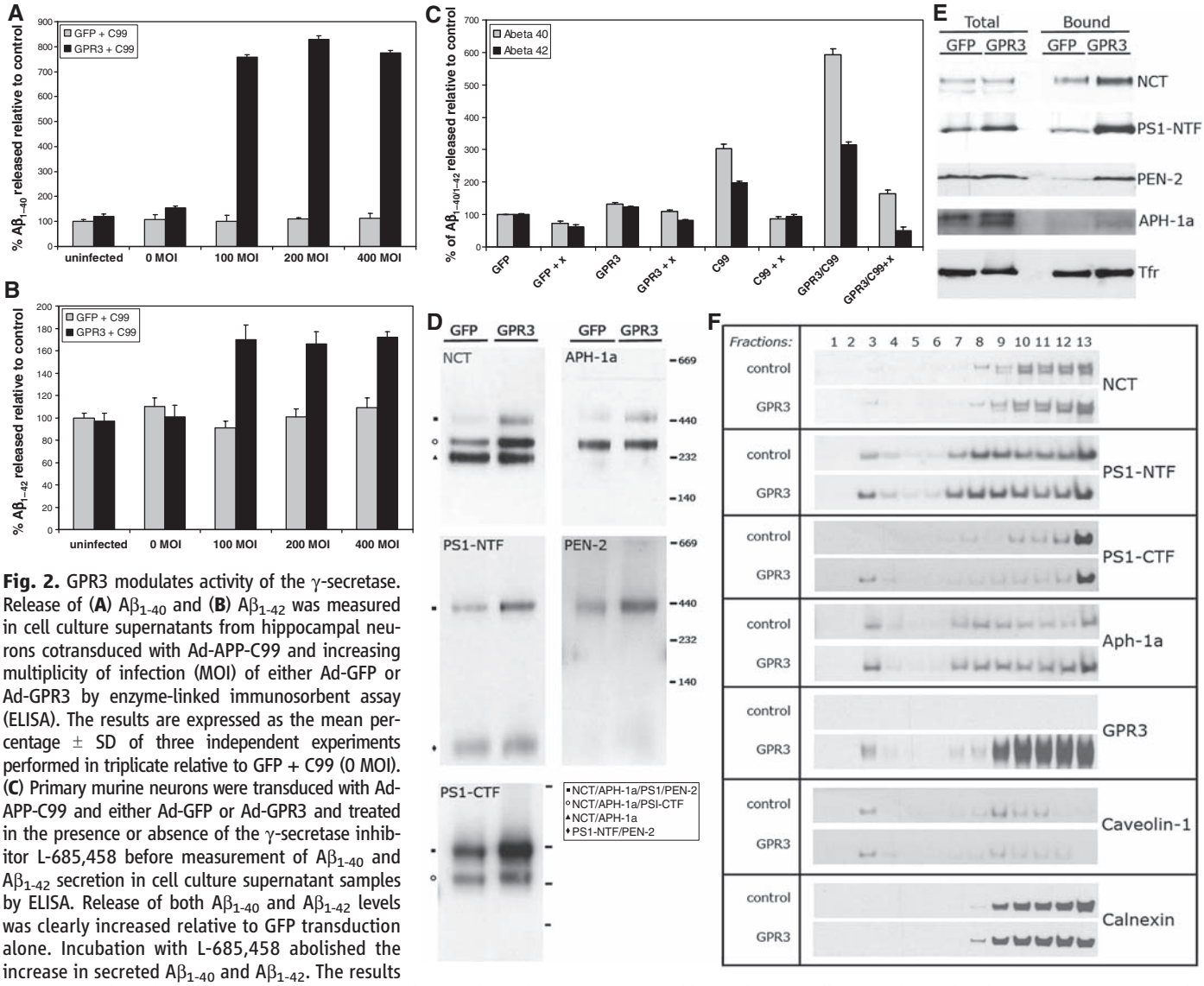


Fig. 2. GPR3 modulates activity of the γ-secretase. Release of (**A**) $A\beta_{1-40}$ and (**B**) $A\beta_{1-42}$ was measured in cell culture supernatants from hippocampal neurons cotransduced with Ad-APP-C99 and increasing multiplicity of infection (MOI) of either Ad-GFP or Ad-GPR3 by enzyme-linked immunosorbent assay (ELISA). The results are expressed as the mean percentage ± SD of three independent experiments performed in triplicate relative to GFP + C99 (0 MOI). (**C**) Primary murine neurons were transduced with Ad-APP-C99 and either Ad-GFP or Ad-GPR3 and treated in the presence or absence of the γ-secretase inhibitor L-685,458 before measurement of $A\beta_{1-40}$ and $A\beta_{1-42}$ secretion in cell culture supernatant samples by ELISA. Release of both $A\beta_{1-40}$ and $A\beta_{1-42}$ levels was clearly increased relative to GFP transduction alone. Incubation with L-685,458 abolished the increase in secreted $A\beta_{1-40}$ and $A\beta_{1-42}$. The results are expressed as the mean percentage ± SD of three independent experiments performed in triplicate relative to GFP (control). (**D**) HEK293 WT cell membrane extracts, after transduction with Ad-GFP or Ad-GPR3, were separated under native conditions by BN-PAGE and analyzed by immunoblot using antibodies that recognize the γ-secretase complex subunits (indicated in the upper left corner of each blot). Ad-GPR3 transduction enhances expression of the mature γ-secretase complex. (**E**) Immunoblot analysis of cell surface biotinylated proteins. Expression of the transferrin receptor (Tfr) and total cell extracts (Total) were used as controls. (**F**) HEK293 WT cells, transfected with empty vector (control) or GPR3 cDNA, were subject to sucrose gradient fractionation. Equivalent volumes were assessed by immunoblot using antibodies that recognize the γ-secretase subunits, RT-GPR3, and the organelle-specific markers: caveolin-1 (caveolae) and calnexin (ER).

confirmed by assessment of the APP intracellular domain (AICD) generation (*23*). Similar to the NICD, the AICD is generated by γ-secretase cleavage. Cotransduction with GPR3 led to an enhancement of AICD formation (Fig. 3B), which was prevented by the γ-secretase inhibitor L-685,458, demonstrating that release of the AICD, but not the NICD, is modulated by GPR3.

To establish the physiological relevance of Aβ modulation by GPR3, we expressed GPR3 in vivo in an AD mouse model expressing KM670/671NL (Swedish) mutated APP and the L166P mutated PS1 (*APP/PS1*) (*24*). We performed stereotactic injection of the purified GPR3 adenoviral vector or a control GFP vector into the hippocampus of 3-month-old *APP/PS1* transgenic mice. Mice were killed after 1 week, and the hippocampus was removed and prepared for analysis. The contralateral cerebral hemisphere was not injected and was used as an internal control. Hippocampal expression of GPR3 enhanced $A\beta_{1\text{-}40}$ and $A\beta_{1\text{-}42}$ generation in vivo (Fig. 4A) without affecting γ-secretase expression (fig. S6). We then evaluated the endogenous physiological relevance of GPR3 with regard to APP proteolytic processing using $Gpr3^{+/+}$, $^{+/-}$, and $^{-/-}$ primary hippocampal neuronal cultures. Release of $A\beta_{1\text{-}40}$ and $A\beta_{1\text{-}42}$ was substantially reduced in $Gpr3^{-/-}$ relative to $Gpr3^{+/+}$ and $Gpr3^{+/-}$ neuronal cultures (Fig. 4B). Expression of the γ-secretase subunits and Ad-APP-C99 was indistinguishable among the genotypes (fig. S7); however, mature γ-secretase complex formation was reduced in $Gpr3^{-/-}$ neurons (fig. S8). Interestingly, reconstitution of GPR3 expression was sufficient to restore Aβ release in $Gpr3^{-/-}$ neurons relative to control (Fig. 4C and fig. S9), establishing the contribution of endogenous GPR3 toward Aβ generation.

We then investigated the in vivo consequence of the absence of GPR3 on Aβ generation in an AD mouse model. *APP/PS1* transgenic mice were mated with $Gpr3^{-/-}$ mice. Both hemizygosity and complete genetic ablation of GPR3 expression resulted in a dramatic reduction in $A\beta_{1\text{-}40}$ and $A\beta_{1\text{-}42}$ levels (Fig. 4D) in the absence of an effect on γ-secretase expression (fig. S10). Thus, endogenous GPR3 is involved in Aβ generation.

In normal human brain tissue sections, GPR3 is strongly expressed in neurons in the hippocampus, amygdala, cortex, entorhinal cortex, and thalamus, regions of the brain that strongly correlate with the pathogenesis of AD (Fig. 4E) (*11*, *25*). GPR3 expression is also elevated in the brains of sporadic AD patients relative to age-matched controls (Fig. 4F), which provides support for the involvement of GPR3 in AD.

We have defined a role for GPR3 in γ-secretase modulation of Aβ generation in vitro and in vivo. Given that mice deficient in Notch signaling display an embryonic lethal phenotype and die on about embryonic day 9.5 (*21*) and in view of the apparent absence of a Notch loss-of-function phenotype in *Gpr3*-deficient mice, which survive to adulthood and have only been reported to exhibit a reproductive defect thus far (*26*), this correlative in vivo evidence supports the in vitro studies, which suggests that GPR3 does not modulate NICD generation. GPR3 appears to promote complex assembly of the γ-secretase, resulting in increased trafficking of the γ-secretase components and the mature γ-secretase complex to the cell surface and increased localization in DRMs, which eventually leads to an increase in Aβ generation. Activation-induced receptor-mediated endocytosis of the β_2-adrenergic receptor, together with the γ-secretase complex, also results in an increase in Aβ production (*27*), although a role for the β_2-adrenergic receptor in Notch processing is not clear.

Thus, the level of expression of GPR3 regulates localization of the γ-secretase complex, thereby affecting the amyloidogenic processing of APP, which suggests that GPR3 is an interesting AD therapeutic target. The finding that GPR3 is involved in assembly of the γ-secretase complex also suggests an intriguing regulatory mechanism for proteolytic activity of the γ-secretase, whereby a pool of inactive intracellular subcomplexes can be mobilized to assemble into mature complexes and redistribute to specific microdomains in the cell membrane to differentially affect the processing of APP and Notch. If similar regulation mechanisms are involved in the cleavage of other γ-secretase substrates, this might provide a level of specificity to the promiscuous proteolytic activities of this intriguing complex.

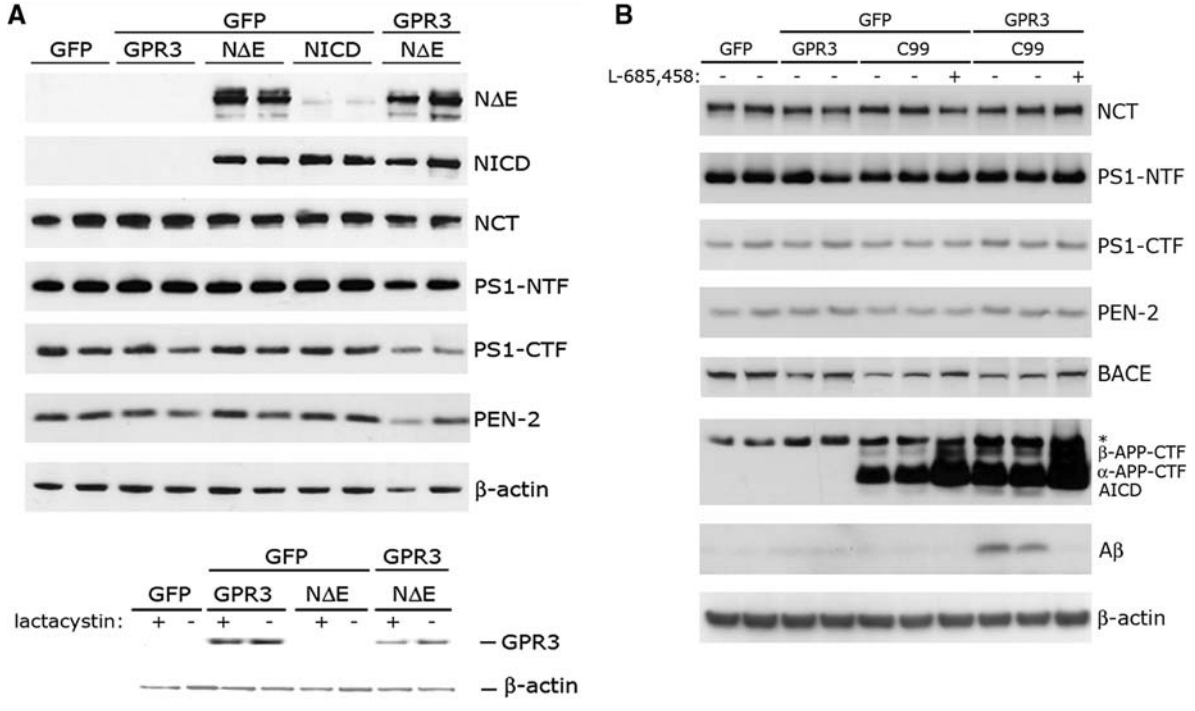


Fig. 3. GPR3 stimulates an increase in AICD generation but does not have an effect on the S3 cleavage of Notch. (**A**) Notch processing is unaffected in HEK293 WT cells cotransduced with Ad-NΔE and either Ad-GFP or Ad-GPR3. Ad-GFP and/or Ad-NICD transduction were used as controls. Lactacystin (10 μM) was used to prevent degradation of the NICD, which was visualized with a cleavage-specific antibody (Notch1 Val-1744). GPR3 expression levels were not affected by lactacystin. (**B**) Cotransduction of HEK293 WT cells with Ad-APP-C99 and Ad-GPR3 led to increased AICD generation. Ad-GFP alone was used as a negative control. The γ-secretase inhibitor L-685,458 abrogated AICD and Aβ generation.

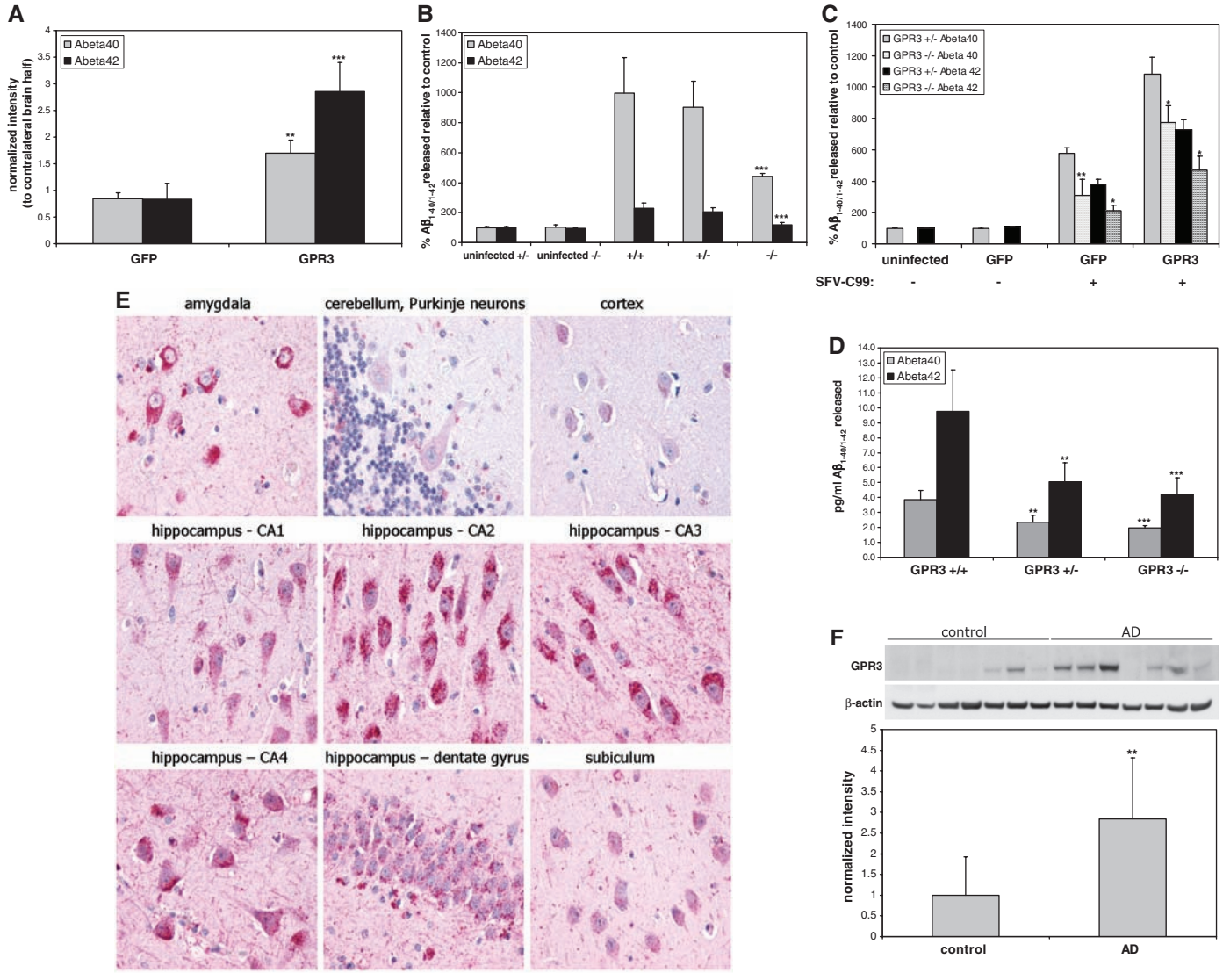


Fig. 4. In vivo modulation of GPR3 expression affects Aβ generation in an AD mouse model. (**A**) Hippocampal concentrations of soluble $A\beta_{1\text{-}40}$ and $A\beta_{1\text{-}42}$ levels after stereotactic injection of either Ad-GFP or Ad-GPR3 were determined by ELISA (**, $P = 0.0038$; ***, $P = 0.0033$ by the unpaired Student's t test: $n = 12$ independent animals per treatment group). (**B**) Release of $A\beta_{1\text{-}40}$ and $A\beta_{1\text{-}42}$ was considerably reduced in hippocampal neuronal $Gpr3^{-/-}$ cultures relative to the $Gpr3^{+/+}$ and $Gpr3^{+/-}$ cultures (***, $P < 0.001$ relative to +/+ or +/–). (**C**) Release of $A\beta_{1\text{-}40}$ and $A\beta_{1\text{-}42}$ was rescued after Ad-GPR3, but not Ad-GFP, transduction in E17 $Gpr3^{-/-}$ hippocampal neurons cotransduced with Ad-APP-C99 (**, $P < 0.01$ relative to GFP (+/–) Abeta 40; *, $P < 0.05$ relative to GFP (–/–) Abeta 40, GFP (+/–) Abeta 42, or GFP (–/–) Abeta 42. (**D**) Hippocampal concentrations of soluble $A\beta_{1\text{-}40}$ and $A\beta_{1\text{-}42}$ levels in *APP/PS1* transgenic mice crossed with $Gpr3^{+/+}$, $Gpr3^{+/-}$, or $Gpr3^{-/-}$ mice were determined by ELISA (**, $P = 0.003$ relative to $Gpr3^{+/+}$; ***, $P = 0.0003$ relative to $Gpr3^{+/+}$ for $A\beta_{1\text{-}40}$ levels, **, $P = 0.002$ relative to $Gpr3^{+/+}$; ***, $P = 0.001$ relative to $Gpr3^{+/+}$ for $A\beta_{1\text{-}42}$ levels by the unpaired Student's t test: $n = 6$ independent mice per cross). (**E**) Immunolocalization of GPR3 in tissue sections from normal post-mortem human brains without neurological disease using an antibody specific for GPR3 (red). Counter nuclear staining (hematoxylin) is shown in blue. (**F**) Representative immunoblot analysis and densitometric quantification of GPR3 expression in control and AD patient brain samples suggest an elevation of GPR3 in AD patients (**, $P = 0.016$ by the unpaired Student's t test: $n = 14$ patients per group).

References and Notes

1. D. J. Selkoe, *Science* **298**, 789 (2002).
2. W. Annaert, B. De Strooper, *Annu. Rev. Cell Dev. Biol.* **18**, 25 (2002).
3. C. Haass, *EMBO J.* **23**, 483 (2004).
4. V. Wilquet, B. De Strooper, *Curr. Opin. Neurobiol.* **14**, 582 (2004).
5. A. L. Hopkins, C. R. Groom, *Nat. Rev. Drug Discov.* **1**, 727 (2002).
6. R. M. Nitsch, M. Deng, J. H. Growdon, R. J. Wurtman, *J. Biol. Chem.* **271**, 4188 (1996).
7. A. M. Pooler, A. A. Arjona, R. K. Lee, R. J. Wurtman, *Neurosci. Lett.* **362**, 127 (2004).
8. D. Dominguez *et al.*, *J. Biol. Chem.* **280**, 30797 (2005).
9. A. J. Saunders, T. W. Kim, R. E. Tanzi, *Science* **286**, 1255a (1999).
10. D. Blacker *et al.*, *Hum. Mol. Genet.* **12**, 23 (2003).
11. T. P. Iismaa *et al.*, *Genomics* **24**, 391 (1994).
12. S. Tanaka, K. Ishii, K. Kasai, S. O. Yoon, Y. Saeki, *J. Biol. Chem.* **282**, 10506 (2007).
13. Y. M. Li *et al.*, *Nature* **405**, 689 (2000).
14. O. Nyabi *et al.*, *J. Biol. Chem.* **278**, 43430 (2003).
15. A. Herreman *et al.*, *J. Cell Sci.* **116**, 1127 (2003).
16. S. Oddo *et al.*, *J. Biol. Chem.* **281**, 39413 (2006).
17. B. De Strooper, *Neuron* **38**, 9 (2003).
18. D. Spasic *et al.*, *J. Biol. Chem.* **281**, 26569 (2006).
19. B. Chini, M. Parenti, *J. Mol. Endocrinol.* **32**, 325 (2004).
20. K. S. Vetrivel *et al.*, *J. Biol. Chem.* **279**, 44945 (2004).
21. D. Selkoe, R. Kopan, *Annu. Rev. Neurosci.* **26**, 565 (2003).
22. E. H. Schroeter, J. A. Kisslinger, R. Kopan, *Nature* **393**, 382 (1998).
23. A. Weidemann *et al.*, *Biochemistry* **41**, 2825 (2002).
24. R. Radde *et al.*, *EMBO Rep.* **7**, 940 (2006).
25. M. Heiber *et al.*, *DNA Cell Biol.* **14**, 25 (1995).
26. C. Ledent *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **102**, 8922 (2005).
27. Y. Ni *et al.*, *Nat. Med.* **12**, 1390 (2006).

28. We acknowledge the contributions of W. Laenen, H. Klaassen, S. Hinnekint, S. de Vos, E. van Rossen,

A. Van de Poel, M. De Raeymaeker, K. Thys, J. Vandersmissen, J. van der Schueren, R. Brown, K. Albert, V. Coose, K. Pittois, S. Theunissen, V. Hendrix, E. Maes, K. Craessaerts, J. Prennen, A. Deelder, P. Scheltens, and M. Mercken. We are grateful to G. Vassart and C. Ledent (Université Libre de Bruxelles, Brussels, Belgium) for the RT-GPR3 and RT-TSHr cDNA and assistance with cAMP assays, C. H. Berlot (Geisinger Medical Center, Danville, PA) for the G_sDN and R280K cDNA, and Murat Bastepe (Harvard Medical School, Boston, MA) for the $G_{s\alpha}$-null mouse embryo fibroblasts. This project was supported by IWT990363, IWT010512, and IWT030623, a Methusalem grant from the Flemish government, the Fund for Scientific Research (Flanders, Belgium), and the Federal Office for Scientific Affairs (IUAP P6/43/), Belgium. B.D.S. has been a paid consultant for Galapagos, a drug discovery company. Galapagos holds U.S. patent 7429459 on the GPR3 findings and the technology and library used in this study.

Supporting Online Material
www.sciencemag.org/cgi/content/full/323/5916/946/DC1
Materials and Methods
SOM Text
Figs. S1 to S14
Tables S1 to S5
References

19 May 2008; accepted 19 December 2008
10.1126/science.1160649

Differences in Early Gesture Explain SES Disparities in Child Vocabulary Size at School Entry

Meredith L. Rowe* and Susan Goldin-Meadow

Children from low–socioeconomic status (SES) families, on average, arrive at school with smaller vocabularies than children from high-SES families. In an effort to identify precursors to, and possible remedies for, this inequality, we videotaped 50 children from families with a range of different SES interacting with parents at 14 months and assessed their vocabulary skills at 54 months. We found that children from high-SES families frequently used gesture to communicate at 14 months, a relation that was explained by parent gesture use (with speech controlled). In turn, the fact that children from high-SES families have large vocabularies at 54 months was explained by children's gesture use at 14 months. Thus, differences in early gesture help to explain the disparities in vocabulary that children bring with them to school.

It has long been recognized that children from high–socioeconomic status (SES) families have, on average, larger vocabularies than children from low-SES families (*1*). This SES gap in vocabulary size begins in the toddler years (*2*), widens until age four, and then remains relatively constant throughout the school years (*3*). Vocabulary is a key predictor of school success (*4*) and is a primary reason why low-SES children enter school at greater risk for failure than their high-SES peers (*5*). Early childhood is thus a critical educational period, as SES differences in language skills first emerge during these years (*3*, *6*).

What is it about a family's SES that leads to disparities in child vocabulary? Previous research suggests that the way parents talk to their children explains some of the relation between SES and child vocabulary (*1*, *7*–*9*). On average, parents from higher-SES groups talk more, use more diverse vocabulary, and use more complex syntax with their children than parents from lower-SES groups, and these differences relate to child vocabulary development (*2*, *7*, *8*, *10*–*13*).

Here we investigate another aspect of parent-child communication in relation to SES—parent and child gesture use. We know that children gesture to communicate before they use speech (*14*, *15*). Further, there is a positive relation between parent gesture and child gesture (*16*–*19*), and early child gesture predicts later child vocabulary, even controlling for early child speech (*16*, *20*). We build on this earlier work and ask: Are there SES differences in the way children and their parents use gesture? If so, might these differences help to explain the robust relation between SES and child vocabulary skill?

To address these questions, we videotaped 14-month-old children from 50 families (representing the demographic range of the Chicago area) engaging in their ordinary activities with their primary caregivers at home for 90 min. We transcribed all speech and gesture used by parent and child during the interaction to glean measures of spoken vocabulary and gesture use [see the supporting online material (SOM) for details on sample and coding].

The number of "gesture types," defined as the number of different meanings conveyed by gesture, served as our measure of child and parent gesture use (e.g., pointing at a dog = dog). Previous research has found child gesture types, rather than child gesture frequency, to be a better predictor of later-child spoken vocabulary size (*16*). At 14 months, children produced an average of 20.6 gesture types (SD = 11.9), and parents produced an average of 39.3 gesture types (SD = 25.6).

The number of "word types," defined as the number of different intelligible word roots produced by the speaker, served as our measure of spoken vocabulary. At 14 months, children used an average of 13 word types during the interaction (SD = 13.3), and parents used an average of 364 word types (SD = 132.0).

At child age 14 months, there was a positive relation between spoken word types and gesture types for both children (correlation coefficient $r = 0.61$, $P < 0.001$) and parents ($r = 0.67$, $P < 0.001$). Furthermore, parents who produced more gesture types had children who produced more gesture types ($r = 0.44$, $P < 0.001$). However, there was no relation between parent word types and child word types at this early stage of language production, nor was there a relation between parent word types and child gesture types.

On average, parents had 15.8 years of education (SD = 2.2) and an average family income level of $60,400 (SD = $31,365). Family income and education were positively related to one another ($r = 0.44$, $P = 0.001$) and were combined into one variable (SES) with the use of principal components analysis (see the SOM for more information on SES measures).

SES was positively related to child gesture at 14 months ($r = 0.30$, $P < 0.05$) and to parent gesture at child age 14 months ($r = 0.45$, $P = 0.001$). Thus, SES differences are reflected in early parent-child gesture use. However, there was no relation between SES and child word types, although there was a positive relation between SES and parent word types ($r = 0.44$, $P = 0.001$).

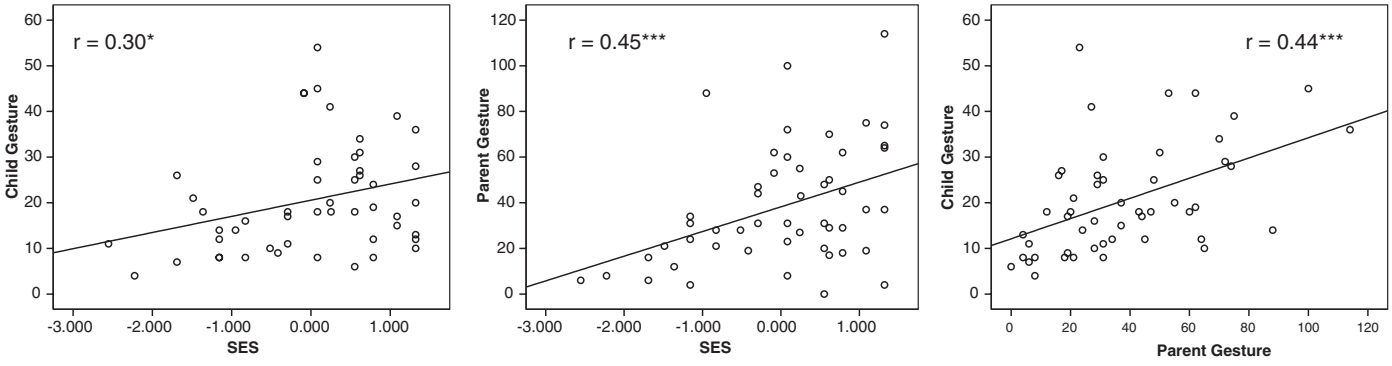
We used correlation and regression analyses to determine whether the positive relation between SES and children's early gesture use is mediated by parents' gesture use during interactions with their children. We followed guidelines for evaluating mediation models put forth by Baron and Kenny (*21*). Specifically, we used statistics to determine whether one variable explains a significant amount of the relation found between two other variables (see the SOM for more information on mediation analysis and assumptions).

The three scatter plots presented in the top of Fig. 1 show that the first three necessary conditions for mediation were met. The leftmost panel presents (i) the significant relation between the predictor variable (SES) and the outcome variable (child gesture). The middle panel displays (ii) the significant relation between the predictor variable (SES) and the potential mediator variable (parent gesture). The rightmost panel displays (iii) the significant relation between the mediator variable (parent gesture) and the outcome variable (child gesture). The

Department of Psychology, University of Chicago, 5848 South University Avenue, Chicago, IL 60637, USA.

*To whom correspondence should be addressed. E-mail: rowemer@uchicago.edu

fourth and final necessary condition for mediation is that the significant relation between the predictor variable (SES) and the outcome variable (child gesture) must be reduced when the mediator variable (parent gesture) is included in the model. This effect is shown in the bottom portion of Fig. 1: (iv) The relation between SES and child gesture (controlling for children's word types at 14 months) is no longer significant when parent gesture is included in the model—the parameter estimate for the SES effect reduces to 1.85 (from 3.45). Bootstrapping procedures to test the significance of this indirect effect (i.e., the product of the coefficients that make up the mediated effect) (*22*, *23*) gave a 95% confidence



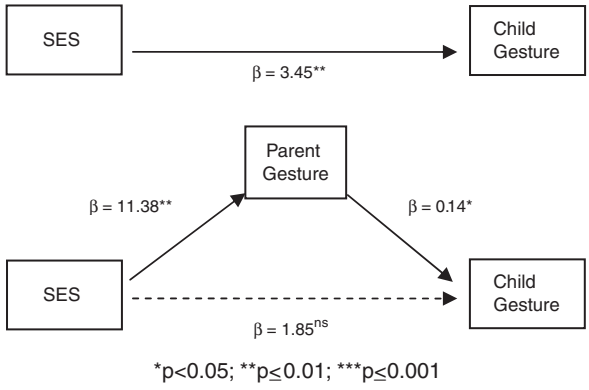
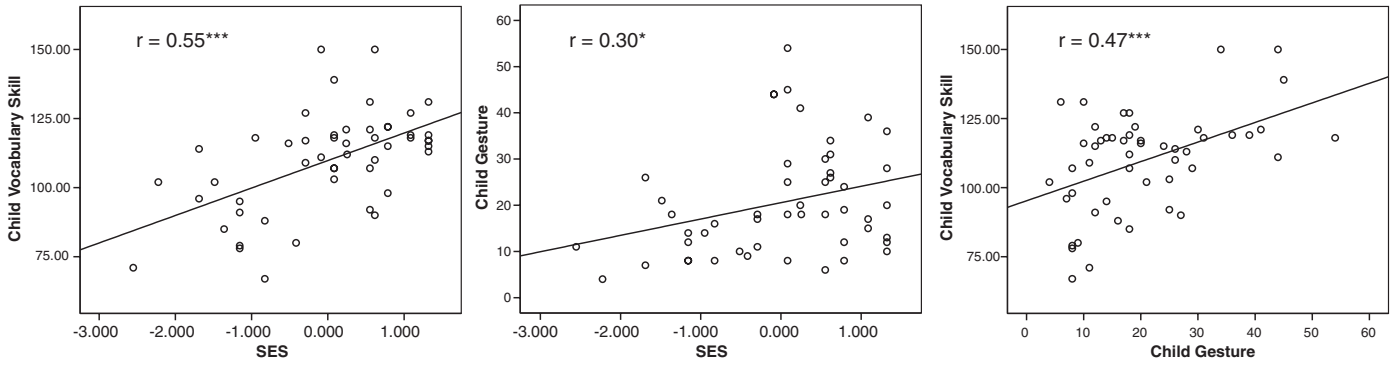


Fig. 1. Scatter plots showing relations between (i) SES and child gesture at 14 months (top left), (ii) SES and parent gesture at child age 14 months (top middle), and (iii) parent gesture and child gesture (top right). The bottom portion of the figure represents analysis (iv) showing that parent gesture mediates the relation between SES and child gesture, controlling for child speech at 14 months (n = 50 children). Taken together, SES and child speech explain 45% of the variation in child gesture; adding parent gesture explains 52%. Asterisks indicate the level of statistical significance as noted; β indicates the regression parameter estimate; ns, not significant (at the $P < 0.05$ level).



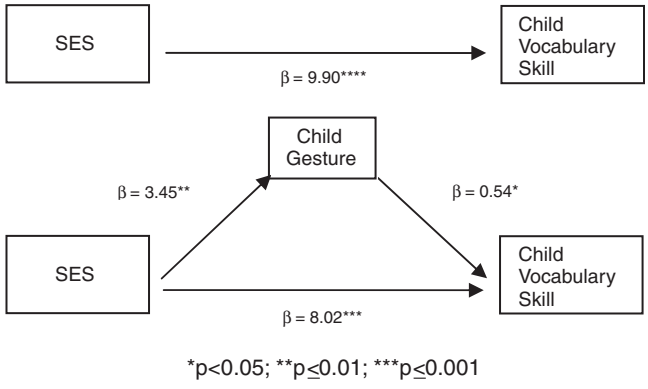


Fig. 2. Scatter plots showing relations between (i) SES and child vocabulary skill (PPVT) at 54 months (top left), (ii) SES and child gesture at 14 months (top middle), and (iii) child gesture and child vocabulary skill (top right). The bottom portion of the figure represents analysis (iv) showing that child gesture mediates the relation between SES and child vocabulary skill, controlling for child speech at 14 months (n = 50 children). Taken together, SES and child speech explain 33% of the variation in child vocabulary skill; adding child gesture explains 40%.

interval (corrected for bias) of 0.26 to 3.52, an interval that does not contain zero and thus indicates that the mediation effect is significant.

We ran an additional model including parent word types at child age 14 months. In this model, parent gesture and child word types remained significant predictors of child gesture, but neither SES nor parent word types were significant predictors. Thus, the relation between parent gesture and child gesture holds even with parent talk controlled.

We now ask whether the relation between SES and later child vocabulary skill can be explained by children's early gesture use. We assessed child vocabulary skill at 54 months with the use of the Peabody Picture Vocabulary Test (PPVT) (*24*). The average normed PPVT score for the sample was 109.8 (SD = 18.1). The scatter plots in Fig. 2 show that the first three conditions of mediation are met: (i) SES relates to children's PPVT scores, $r = 0.55$, $P < 0.001$ (leftmost panel); (ii) SES relates to child gesture, $r = 0.30$, $P < 0.05$ (middle panel); and (iii) child gesture relates to PPVT scores, $r = 0.47$, $P = 0.001$ (rightmost panel). The bottom portion of Fig. 2 shows the results of the mediation analysis. (iv) The relation between SES and PPVT (controlling for children's word types at 14 months) reduces in magnitude when child gesture is included in the model; the parameter estimate for the SES effect reduces to 8.02 (from 9.90). Bootstrapping procedures gave a 95% confidence interval (corrected for bias) of 0.44 to 4.51, an interval that does not contain zero and thus indicates that the mediation effect is significant. Thus, the effect of SES on child vocabulary at 54 months is mediated, in part, by children's gesture use at 14 months.

Overall, our findings are consistent with the following developmental history. When interacting with their children, parents from higher-SES groups use gesture to communicate a broader range of meanings than parents from lower-SES groups. By 14 months, children from these higher-SES families are using gesture to communicate more meanings than children from lower-SES families. Thus, as early as 14 months of age, children in different SES groups may be socialized to communicate more or fewer meanings via gesture. These early differences in gesture, in turn, help to explain the large disparities in vocabulary size that characterize children of different SES groups at school entry.

It is notable that, in the initial stages of language learning when SES differences in children's spoken vocabulary are not yet evident, we see SES differences in child gesture use. Children typically do not begin gesturing until around 10 months of age (*14*, *15*). Thus, SES differences are evident a mere 4 months (and possibly even sooner) after the onset of child gesture production.

Why do we see SES differences in children's early gestures? Although correlation does not imply causation or the direction of effects, our results implicate parent gesture as a plausible mechanism. Even before they produce their own gestures, children comprehend the gestures that others produce (*25*, *26*). Because children from high-SES backgrounds are exposed to a broader range of meanings in gesture than children from lower-SES backgrounds (as shown here), they themselves are likely to produce more meanings in gesture, which then promotes the development of more vocabulary words in speech (*27*). This scenario parallels findings in speech—children from high-SES backgrounds are exposed to more diverse vocabulary than children from lower-SES backgrounds and, in turn, produce more vocabulary words of their own (*2*, *12*). Thus, SES appears to relate to child vocabulary in two ways: (1) through parent speech and (2) through parent and child gesture.

We have shown here that variation in children's early gesture use, independent of early spoken vocabulary, explains a portion of the robust relation between SES and later vocabulary skill. However, the specific nature of the relation between early child gesture and later child vocabulary is not addressed in this study. Child gesture could play an indirect role in word learning by eliciting timely speech from parents; for example, in response to her child's point at a doll, a mother might say, "yes, that's a doll," thus providing a word for the object that is the focus of the child's attention. S. Goldin-Meadow *et al*. (*28*) found that when mothers "translated" their child's gestures into words in just this way, those words tended to become part of the child's spoken vocabulary several months later. Gesture could also play a more direct role in word learning by giving children an opportunity to practice generating particular meanings by hand, at a time when those meanings are difficult to produce by mouth (*27*).

Whether or not early gesture plays a direct or indirect role in word learning, it is clear that gesturing partially accounts for the relation between SES and later vocabulary skill. Of course, gesture is not the sole factor mediating the robust relation between SES and child vocabulary. Other environmental factors (including parent speech) and child factors probably influence child vocabulary as well. Nonetheless, given our findings, it seems fruitful for future research to explore whether parents and children can be encouraged to increase the rate at which they spontaneously gesture when they speak [as has been shown in older children, either by directly telling them to gesture (*29*) or by modeling gesture for them (*30*)]. If so, the next step is to explore whether increases in gesturing lead to vocabulary gains in early childhood.

References and Notes

1. E. Hoff, *Dev. Rev.* **26**, 55 (2006).
2. B. Hart, T. Risley, *Meaningful Differences in the Everyday Experience of Young American Children* (Brookes, Baltimore, 1995).
3. G. Farkas, K. Beron, *Soc. Sci. Res.* **33**, 464 (2004).
4. R. C. Anderson, P. Freebody, in *Reading Comprehension and Education*, J. T. Guthrie, Ed. (International Reading Association, Newark, DE, 1981), pp. 77–117.
5. C. E. Snow, S. Burns, P. Griffin, *Preventing Reading Difficulties in Young Children* (National Academy, Washington, DC, 1998).
6. G. J. Duncan, W. J. Yeung, J. Brooks-Gunn, J. R. Smith, *Am. Sociol. Rev.* **63**, 406 (1998).
7. E. Hoff, in *Socioeconomic Status, Parenting, and Child Development*, M. H. Bornstein, Ed. (Erlbaum, Mahwah, NJ, 2003), pp. 147–160.
8. E. Hoff, *Child Dev.* **74**, 1368 (2003).
9. E. Hoff, B. Laursen, T. Tardiff, in *Handbook of Parenting, Volume II: Ecology and Biology of Parenting*, M. H. Bornstein, Ed. (Erlbaum, Mahwah, NJ, 2003), pp. 161–188.
10. M. H. Bornstein, M. O. Haynes, K. M. Painter, *J. Child Lang.* **25**, 367 (1998).
11. S. B. Heath, in *Cultural Psychology: Essays on Comparative Human Development*, J. W. Stigler, R. A. Shweder, G. Herdt, Eds. (Cambridge Univ. Press, Cambridge, 1990), pp. 496–519.
12. J. Huttenlocher, M. Vasilyeva, H. R. Waterfall, J. L. Vevea, L. V. Hedges, *Dev. Psychol.* **43**, 1062 (2007).
13. B. A. Pan, M. L. Rowe, J. D. Singer, C. E. Snow, *Child Dev.* **76**, 763 (2005).
14. E. Bates, *Language and Context* (Academic, Orlando, FL, 1976).
15. E. Bates, L. Benigni, I. Bretherton, L. Camaioni, V. Volterra, *The Emergence of Symbols: Cognition and Communication in Infancy* (Academic, New York, 1979).
16. M. L. Rowe, S. Özçalıskan, S. Goldin-Meadow, *First Lang.* **28**, 182 (2008).
17. M. L. Rowe, *First Lang.* **20**, 305 (2000).
18. J. M. Iverson, O. Capirci, E. Longobardi, M. C. Caselli, *Cogn. Dev.* **14**, 57 (1999).
19. L. L. Namy, L. Acredolo, S. Goodwyn, *J. Nonverbal Behav.* **24**, 63 (2000).
20. M. L. Rowe, S. Goldin-Meadow, *Dev. Sci.* **12**, 182 (2009).
21. R. M. Baron, D. A. Kenny, *J. Pers. Soc. Psychol.* **51**, 1173 (1986).
22. K. McCartney, M. R. Burchinal, K. L. Bub, *Monogr. Soc. Res. Child Dev.* **71**, 88 (2006).
23. K. J. Preacher, A. Hayes, *Behav. Res. Methods Instrum. Comput.* **36**, 717 (2004).
24. L. M. Dunn, L. M. Dunn, *Peabody Picture Vocabulary Test* (American Guidance Service, Circle Pines, MN, ed. 3, 1997).
25. G. Butterworth, L. Grover, in *Thought Without Language*, L. Weiskrantz, Ed. (Carendon, Oxford, 1988), pp. 5–24.
26. M. Carpenter, K. Nagell, M. Tomasello, *Monogr. Soc. Res. Child Dev.* **63**, 52 (1998).
27. J. M. Iverson, S. Goldin-Meadow, *Psychol. Sci.* **16**, 367 (2005).
28. S. Goldin-Meadow, W. Goodrich, E. Sauer, J. Iverson, *Dev. Sci.* **10**, 778 (2007).
29. S. C. Broaders, S. W. Cook, Z. Mitchell, S. Goldin-Meadow, *J. Exp. Psychol. Gen.* **136**, 539 (2007).
30. S. W. Cook, S. Goldin-Meadow, *J. Cogn. Dev.* **7**, 211 (2006).

31. Supported by the National Institute of Child Health and Human Development: grants K99HD055522 to M.L.R. and P01HD40605 to S.G.-M. We thank J. Huttenlocher, S. Levine, W. Meadow, J. Rowe, and three anonymous reviewers for comments on earlier versions; the families for their participation; K. Brasky, L. Chang, E. Croft, K. Duboc, J. Griffin, S. Gripshover, K. Harden, L. King, C. Meanwell, E. Mellum, M. Nikolas, J. Oberholtzer, C. Rousch, L. Rissman, B. Seibel, M. Simone, K. Uttich, and J. Wallman for data collection/transcription; E. Sauer for gesture coding; and K. Schonwald and J. Voigt for administrative/technical assistance.

Supporting Online Material

www.sciencemag.org/cgi/content/full/323/5916/951/DC1
Materials and Methods
SOM Text
References

9 October 2008; accepted 15 January 2009
10.1126/science.1167025

New Products Focus: Essential Lab Equipment

Gradient Thermal Cycler

The TC-3000G gradient thermal cycler enables eight different temperatures to be tested simultaneously within a small footprint of 17 cm by 33 cm. The instrument offers two easily interchangeable block options. The 48-well thermal block is arranged to provide eight columns for temperature optimization and six rows for optimizing reagents and primer concentrations. Annealing temperatures can be optimized over a 15°C gradient between 20°C and 80°C with a fast heating rate of 3.3°C per second. The system can help optimize annealing temperatures for polymerase chain reactions particularly in life science and molecular biology laboratories with insufficient throughput to justify the expense of a 96-well gradient system. A thermal block option of 30 x 0.5 ml offers significantly greater capacity than the usual 12 or 20 tubes.

Bibby Scientific
For information +44-(0)844-936-0234
www.bibby-scientific.com

Sample and Chemical Storage and Retrieval

The comPOUND system is for the storage and retrieval of chemical compounds and biological samples. It offers flexible storage capacity with high-speed cherry-picking to select library subsets. It can deliver any sample in 5 seconds. Sample integrity is maintained through a hermetically sealed storage chamber with an inert nitrogen atmosphere kept at -20°C. Tracking is performed using a barcode on each microtube. The system's innovative use of pneumatics allows samples to be delivered up to 15 meters away. Multiple modules can be connected to expand libraries, with units connected in parallel, so tubes can be accessed from all connected modules at the same time. The modular design allows expansion to up to 1.2 million samples.

TTP LabTech
For information +44-1763-262626
www.ttplabtech.com

Real-Time PCR System

The BioMark Real-Time PCR System for high throughput genetic analysis performs on-board thermal cycling and fluorescence detection. The Biomark system integrates thermal cycling and detection of polymerase chain reaction (PCR) assays. It acquires data for each reaction chamber of the integrated fluidic circuit (IFC) chip simultaneously and can operate in either end-point or real-time detection mode. It includes integral data collection and data analysis software for genotyping, digital, and real-time PCR. Powerful real-time PCR analysis software displays the analyzed data in multiple formats, including color-coded maps of every reaction chamber on the IFC chip, amplification curves, and numeric tables. Because the system is designed for proprietary licensed 5' nuclease assays, laboratories can easily switch to dynamic arrays and digital arrays for PCR while continuing to use trusted reagents and protocols.

Fluidigm Europe
For information +33-44-259-3861
www.fluidigm.com

DNA Purification and Sequencing

The BioCel platform is a sample processing system for assembling genomics reactions, dispensing samples for compound preparation, or processing biological samples in screening applications. It is available in three different sizes with various options for enclosures and environmental control. With the addition of the high-speed VPrep precision pipetting system, the BioCel becomes an automated system for DNA purification and sequencing. The VPrep's ability to perform complex pipetting sequences while delivering high throughput overcomes the problem of the long and complex pipetting protocols needed for DNA preparation. The BioCel system is capable of this demanding application thanks to its powerful dynamic scheduling software, flexible liquid handling options, high-speed central robot, and ability to integrate third-party instruments. The combination of the BioCel and the VPrep can provide unprecedented throughput for DNA purification and sequencing.

Velocity11
For information +44-1763-269110
www.velocity11.com

Refrigerated Microcentrifuge

The Prism R refrigerated microcentrifuge is less than 11 inches wide. Programming and operation are simple, through one control knob and a large LCD. The brushless motor can achieve a maximum speed of 13,500 rpm/17,133 x g. The refrigeration system can cool the centrifuge to 4°C in less than eight minutes. The Prism R is supplied with a 24-place rotor for 1.5/2.0 ml tubes. Adapters are available for smaller tubes. A unique StripSpin rotor for polymerase chain reaction strips is available. The unit's vibration-free isolation system helps in achieving smooth and quiet operation. Other features include momentary spin and imbalance detection.

Labnet International
For information 888-522-6381
www.labnetlink.com

Electronically submit your new product description or product literature information! Go to **www.sciencemag.org/products/newproducts.dtl** for more information.

Newly offered instrumentation, apparatus, and laboratory materials of interest to researchers in all disciplines in academic, industrial, and governmental organizations are featured in this space. Emphasis is given to purpose, chief characteristics, and availabilty of products and materials. Endorsement by *Science* or AAAS of any products or materials mentioned is not implied. Additional information may be obtained from the manufacturer or supplier.