Newsjournal of the Society for Industrial and Applied Mathematics

siam news

Volume 46/ Number 3
April 2013

Living Proof: MBI Celebrates the Future of Biomathematics

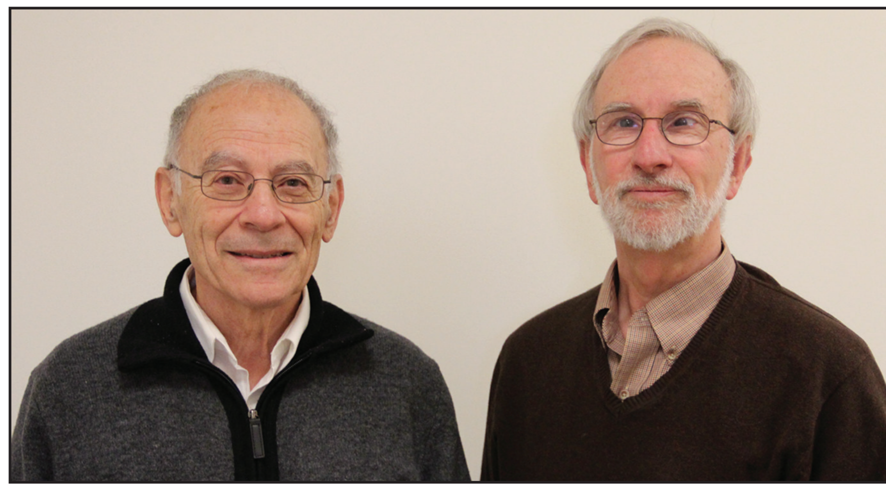
By Barry A. Cipra

The Mathematical Biosciences Institute is ten years old. To mark the milestone, it did what institutes do: It held a meeting. Rather than a self-laudatory review of past accomplishments, MBI set its sights on the *future* of mathematical biology.

The meeting, held September 19–21 at the institute's digs on the third floor of Jennings Hall at the Ohio State University, featured 11 lectures by experts in areas ranging from public health statistics to mathematical neuroscience.* It also featured a poster session with contributions from many of the nearly six dozen postdoctoral fellows who have passed through MBI over the years, and it concluded with a banquet honoring the institute's founding director, Avner Friedman.

"This century *is* the century of biology," Friedman said at the banquet. Much as math and physics partnered in the past, math and biology will partner well into the future. The burgeoning relationship is driven by the complexity of biological problems and by recent stunning technological developments in experimental and computational capabilities. Researchers are able to probe increasingly fine detail with increasing accuracy, presenting mathematical modelers and analysts with a firehose of data from which they can slake their simulated and theoretic thirst. The challenge for mathematical scientists will be to seize the opportunities presented by the partnership. "New math will not only advance biology," Friedman says, "but new math will come out of it."

*Readers can find "Modeling the Origin of Eusociality," William Kolata's report on Martin Nowak's keynote presentation at the MBI 10th-anniversary meeting, in the March 2013 issue of *SIAM News*.



Marking the 10th anniversary of the Mathematical Biosciences Institute, founding director Avner Friedman (left) and current director Marty Golubitsky look to a future of inspired teamwork that will continue to drive advances in both fields.

Slicing and Dicing the Data

Nicholas Jewell, a biostatistician at UC Berkeley, gave a glimpse of what lies ahead in the age of big data. "When I arrived in the school of public health at Berkeley [in 1981], there wasn't a single computer anywhere in the department," he recalls. Statisticians for the most part dealt with problems having a relatively small number of observations and variables. Nowadays, the pressing problems bristle with huge numbers of observations comprising ultra-high-dimensional data sets that often feature non-standard data (such as video or Tweets), giving rise to non-standard questions (e.g., Can you recognize an incipient flu epidemic based on Google searches?).

"Statisticians are at the fundamental core of every endeavor," Jewell says. Of all the data currently stored, he points out, an estimated 90% was created in the last two years. It's sometimes proposed that science can now dispense with theory and "just let the data speak." Nothing could miss the mark more widely, Jewell says: "The more data we have, the more mistakes we seem to make."

The core tasks haven't changed. Statistics is vital in obtaining data, and cleaning and exploring it, as well as in modeling data, and drawing inferences in the presence

See **Math Biology** *on page 4*

Applying Math to Myth Helps Separate Fact from Fiction

By Pádraig Mac Carron and Ralph Kenna

Although the distinctions between them are not always sharp, myths differ from legends and folktales. Mythological epics frequently entail timeless narratives with abundant characters, outside documented history. Legends, on the other hand, are couched in a definite historical timeframe, and folktales are intentionally fictional.

A frequently asked question is whether some mythological narratives may contain traces of historicity. Despite layers of obfuscation added by distortive political and religious forces to epic tales as they were passed orally through the centuries, could remnants of reality still lurk within the pages of the ancient manuscripts in which they were eventually preserved?

In the absence of firm archaeological evidence for or against the existence of societies described by some mythological narratives, we turned to applied mathematics. In a recently published pioneering study, we looked at the social-network structures of three iconic mythological narratives and found indications that could point to elements of historicity.

Homer's *Iliad* and the Anglo-Saxon *Beowulf* are believed by many antiquarians to be partly historically based. The ancient Irish epic *Táin Bó Cúailnge*, by contrast, is often believed to be mostly fictional. By comparing the social-network structures underlying these three tales, we find that the apparent artificiality of the *Táin* is not pervasive; appropriately interpreted, the Irish narrative may hold a degree of historicity similar to that of the other two.

■ ■ ■

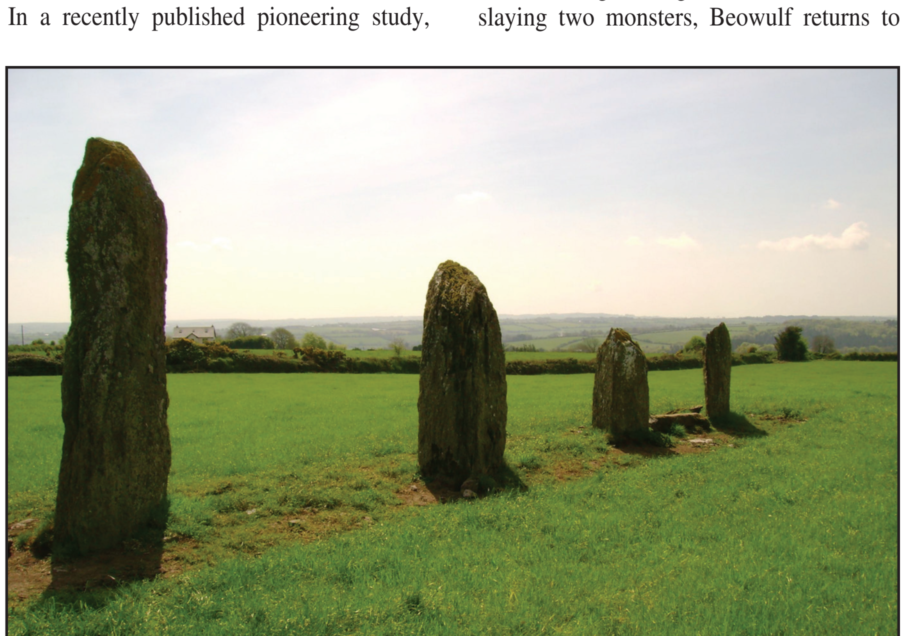
Beowulf is an Old English heroic epic, set in Scandinavia. Estimated dates of the single surviving codex range from the 8th to the 11th centuries. The story tells of the coming of Beowulf, a Geatish hero, to assist Hrothgar, king of the Danes. After slaying two monsters, Beowulf returns to Sweden to become king of the Geats and, following another fabulous encounter many years later, is fatally wounded. Although the poem is embellished with obvious fictional elements and creatures, archaeological evidence in Denmark and Sweden offers support for historicity associated with some of the human characters. The character Beowulf himself is mostly believed not to have existed in reality.

The *Iliad*, an epic poem attributed to Homer, dates to the 8th century BC. It is set during the final year of the war between the Trojans and a coalition of besieging Greek forces. It relates a quarrel between Agamemnon, king of Mycenae and leader of the Greeks, and Achilles, their greatest hero. Although much debated through the years, evidence suggests that the story may be based on a historical conflict that occurred around the 12th century BC, interwoven with elements of fiction.

Táin Bó Cúailnge (Cattle Raid of Cooley), the most well known epic of Irish mythology, describes the invasion of Ulster by the armies of queen Medb of Connacht and the defence by Cúchulainn, Ireland's most famous hero. A number of related pretales and tangential tales (remscéla) give the backgrounds and exploits of the main characters of the *Táin* itself. The *Táin* has come down to us in three recensions. The first has been reconstructed from partial texts contained in *Lebor na hUidre* (the Book of the Dun Cow, dating from the 11th or 12th century) and *Lebor Buide Lecáin* (the Yellow Book of Lecan, a 14th-century manuscript) and other sources. The second, later recension is found in *Lebor Laignech* (the Book of Leinster, a 12th-century manuscript formerly known as *Lebor na Nuachongbála*, or the Book of Nuachongbáil). A third recension comes from fragments of later manuscripts and is incomplete.

Medieval scholars dated the *Táin* to the first centuries BC, but this may have been

See **Math and Myth** *on page 6*



Megalithic standing stones at Castlelack, County Cork, Ireland.

Nonprofit Org
U.S. Postage
PAID
Permit No 360
Bellmawr, NJ

siam®
SOCIETY for INDUSTRIAL and APPLIED MATHEMATICS
3600 Market Street, 6th Floor
Philadelphia, PA 19104-2688 USA

siam news
Volume 46/ Number 3/ April 2013

Questioning Big Data

At a recent meeting of the SIAM Committee on Science Policy (December 3 and 4, Washington, DC), a few themes emerged from the wide-ranging discussions generated by the agenda and slate of visitors. "Big data," the overarching theme, was closely connected to two others: programs of most of the U.S. agencies that fund science and the increasingly interdisciplinary nature of research on important problems.

With a single comment, CSP member Fred Roberts threw all those themes into relief. A lot of agencies, he said, don't understand "big data"—do not, in fact, even know what it is. Of interest to the CSP, he continued, are the "huge opportunities and roles for mathematics with data," both in the use of existing methods to analyze data and in the development of new methods.

Roberts, director emeritus of DIMACS, the NSF-supported Center for Discrete Mathematics and Theoretical Computer Science at Rutgers, is currently director of CCICADA: Command, Control and Interoperability Center for Advanced Data Analysis (a Center of Excellence of the Department of Homeland Security). During a break in the meeting, he outlined for *SIAM News* the view of "big data" that he has honed in thinking about the issues in his various academic and government positions. His points, in summary, are:

■ With respect to a definition of "big data," you have a big data question if you have so much data that you don't know what to save and, in some cases, need to make that decision instantaneously. This occurs especially in certain disciplines, e.g., astrophysics.

■ It is often necessary to determine the normal state of a system in order to be able to quickly detect departures. An example is the smart grid: Operators now get data every 2–4 seconds; with new phasor technology, updates might come 10 times/second; a human won't be able to detect the state of the grid without algorithms.

■ Data now comes from a variety of sources—sensors, audio, video, among many others—and a variety of media. How do you make sense of data coming from the many different sources?

■ How do you store, query, and search data when there's so much of it?

■ How can you trust the data you have? How do you define "trust"? Social media data is an example—can Twitter and Facebook data be considered accurate?

■ You would like to make inferences and hypotheses from large amounts of data. How do you do that?

■ The problem is not just the size of the data set, but also its complexity.

■ Large data problems now come from many disciplines. Examples are NEON (National Ecological Observatory Network), a project of the National Science Foundation, and GBIF (Global Biodiversity Information Facility), an international effort to digitize all information about all living species (estimated number: between 2 and 10 million). These projects are striking for both the size and the complexity of the data sets. The intelligence community has been dealing with big data questions for some time; the Department of Homeland Security tries to do so. The financial sector grapples with huge amounts of data.

Just about every U.S. federal agency that funds science currently supports at least one major national initiative on data. Announcing a $200 million R&D initiative in big data in March 2012, the White House described the program as a way to enhance "our ability to extract knowledge and insights from large and complex collections of digital data." Several of the agencies most important to the interests of the SIAM community were represented at the CSP meeting by (knowledgeable) visitors, who described their programs and engaged in discussion with the committee members.

With articles by Barry Cipra, careers columnist Tanya Moore, and reviewer James Case, this issue of SIAM News *offers a variety of perspectives on big data. Under way for future issues are reports from the unprecedentedly well-attended 2013 SIAM Conference on Computational Science and Engineering (Boston, February 25 to March 1), where big data was featured in many sessions, including the forward-looking panel discussion, "Big Data Meets Big Models."*

Cardiff Students Hold First SIAM Chapter Day

The First Annual SIAM Chapter Day, held at Cardiff University on January 21, 2013, marked the formation of the first SIAM Student Chapter in Wales. This launching event brought together more than 70 research students and faculty in mathematics and engineering, and offered a platform for discussions of a wide range of topics—from computational accuracy to mathematical modeling in science and industry to engineering applications.

The formation of the Cardiff SIAM Chapter came as a natural development in an ongoing effort to strengthen the ties between mathematics and engineering at Cardiff; the chapter's goal is to contribute to the education of a new generation of scientists who can interact with each other in a significant and effective manner, from the beginning. Currently, the number of applied mathematicians is almost equal to that of engineering students in the chapter, and this is suitably reflected by the fact that the president and the vice-president are a mathematician and an engineer, respectively.

The programme included guest lectures by Nick Higham (University of Manchester), Alain Goriely (University of Oxford), Simon Cox (Aberystwyth University), and Matthew Gilbert (University of Sheffield), and poster presentations by doctoral students from Cardiff University and their guests from the Universities of Manchester, Aberystwyth, and Swansea.

Student poster presenters and guest lecturers at the First Annual SIAM Chapter Day at Cardiff University, UK: Front row, from left: Chris Rowlatt, Tom Croft, Darong Jin, Chang-Kye Lee, Mike Walters, Daniel Alves Paladim, and Danas Sutula; back row, from left: Hadrien Courtecuisse, Piotr Kusmierczyk, Lewis Pryce, Olivier Goury, Matthew Gilbert, Alain Goriely, Ross McKenzie, Nick Higham, Mary Aprahamian, Sam Relton, and Simon Cox.

Nick Higham's discussion of the accuracy and stability of numerical algorithms, rich in examples from everyday life, was complemented by posters by research students providing further technical details. Alain Goriely explored the paradigm of chirality in mechanical biology, showing how many important biological systems, such as bio-gels, DNA, and plant stems, exhibit counterintuitive phenomena that challenge our understanding of fundamental concepts in mechanics and material modeling. Simon Cox talked about the underlying bubble-scale mechanisms in the cellular structure of liquid foams and the emerging field of discrete microfluidics, wherein foams are pushed through narrow channels. This work has found many industrial applications, from oil extraction to food and cosmetic products, although more accurate numerical algorithms are needed to predict the correct mechanical behaviour required for industrial processes. Matthew Gilbert described the so-called layout optimization technique, as applied to truss design problems and to the problem of identifying the critical layout of slip-line discontinuities in a solid body at the point of collapse; the technique employs numerical schemes for solving these and other large-scale mathematical programming problems. Slides for all the guest lectures can be downloaded from the event's official website: http://www.cardiff.ac.uk/maths/research/researchgroups/applied/siam/siamday2013.html.

This one-day event might have seemed ambitious, not least because of an amber weather warning for southeastern Wales the preceding week-end. The aim was to give participants the opportunity to interact across areas, and the speakers and students had to work very hard to make their contributions accessible to a varied and demanding audience.

Everyone was up to the challenge, and the resulting event was thoroughly enjoyed by all involved.—*Angela Mihai, Faculty Adviser, SIAM Student Chapter at Cardiff University.*

siam news

ISSN 1557–9573. Copyright 2013, all rights reserved, by the Society for Industrial and Applied Mathematics, SIAM, 3600 Market Street, 6th Floor, Philadelphia, PA 19104–2688; (215) 382–9800; siam@siam.org. To be published ten times in 2013: January/February, March, April, May, June, July/August, September, October, November, and December. The material published herein is not endorsed by SIAM, nor is it intended to reflect SIAM's opinion. The editors reserve the right to select and edit all material submitted for publication.

Advertisers: For display advertising rates and information, contact Kristin O'Neill at marketing@siam.org.
One-year subscription: $66.00 U.S., Canada, and Mexico, $76.00 elsewhere. SIAM members and subscribers should allow 8 weeks for an address change to be effected. Change of address notice should include old and new addresses with zip codes. Please request address change only if it will last 6 months or more.

Printed in the USA.

siam. *is a registered trademark.*

Patternmaster: The Inner World of a Mathematician

Careers in the Math Sciences

By Tanya Moore

I love science fiction. *Patternmaster* by Octavia Butler, one of my favorite authors, is in part the story of Doro, a character born during ancient Egyptian times. With the ability to have his consciousness transferred from one body to another, he is immortal. Over time, he develops a network of telepaths that he can control with his thoughts, and this network becomes a matrix of minds that he is deeply attuned to, sensing changes or disturbances like a tug on a rope. The more minds in his web, the stronger he becomes and the more he is able to influence and create the experiences he desires.

Although my current career choice does not involve mind control or telepathy, some aspects of Doro's ability remind me of the types of skills mathematicians regularly call on to do their work.

As a young girl I always noticed patterns. Whether seeing the symmetry in the numbers of a street address, or observing that a set of related ideas was simultaneously communicated from completely different sources, or having fun guessing the next number in a series or sequence, identifying patterns became one way I oriented myself to my environment. Looking back, it's not surprising that I would ultimately be drawn to mathematics, but at the time I never would have imagined that it would figure so prominently in my life.

My aspirations included becoming an actress, a psychologist, and a revolutionary (I grew up in Berkeley, California). With all the theatre performances, political protests, and cheerleading practices, math remained an unglorified background singer to the main headliners in my life, despite a lot of A's and high marks on standardized tests. The only pattern I didn't recognize was the path I was on—I had so many interests but no one clear desire for a future career.

Had I not gone to Spelman College, I doubt that I would have majored in mathematics. I arrived at Spelman feeling diminished and limited by the words of my high school AP calculus teacher, who told me at the end of the course that she didn't think I belonged in a math class at that high a level. As a Spelman freshman, I initially decided to major in psychology. I had scored high enough on placement exams in math to be exempt from any math requirement for my major. But it felt weird not to take any math classes. What had been a staple in my academic schedule for many years couldn't now be casually tossed aside, and I decided to enroll in a calculus class.

Before the end of my freshman year, urged by my calculus professor to change my major to math or engineering, I decided to consider a different path for myself, one that explicitly included mathematics. That encouragement from my professor was the first time my ability and potential in mathematics were recognized by a teacher. Twenty years later, the words of many of my teachers continue to affect my view of what is possible for my life.

Spelman College is a Historically Black College for women. As a student there, I had the incredible opportunity to be under the tutelage of women who were mentors, role models, and great teachers. At that point I had no real appreciation for what it must have meant for them to earn doctorates in math at a time when not many women, let alone African American women, had the opportunity to earn doctorates, let alone in math!

Upon graduating from Spelman, I started graduate school at Johns Hopkins University, and I quickly realized that success would require more than A grades in undergraduate math courses. I struggled to feel connected to most of my professors, and I was intimidated by my peers, some of whom had started the program with master's degrees in related fields or had taken graduate-level math courses as undergrads. Comments by students and faculty—that I was sure to pass my qualifying exams because the department needed more black students, that I had performed better than expected in class—made me self-conscious and shy about asking questions in class or seeking additional support; they challenged my willingness to try to be the best student I could be.

I decided to transfer to the University of California, Berkeley (Cal), and place myself in an environment where I felt I could be successful. Besides transferring schools and taking more responsibility for being a good student, I was able to complete my doctorate in large part because of the connections I maintained with my former Spelman classmates and professors. Their example and encouragement during graduate school were instrumental to my success.

I completed my doctoral studies in biostatistics at Cal, where I was fortunate to work under the supervision of Mark van der Laan. In my dissertation research, in the area of statistics called causal inference, I worked on statistical models that could be applied in analyzing data that arose from "non-textbook" situations. Van der Laan had a strong foundation in theoretical mathematics; for me, his ability to explain things and make connections was amazing—it was as if the depth of his mathematical understanding allowed him access to a playground from which he could create new models and theories. He gave me an appreciation for what one can discover by investing in learning and understanding as much information as possible.

The challenging and instructive experiences I had in graduate school, combined with the understanding that math provides a path to many careers, inspired me to share what I had learned to support other minority women mathematicians. After

For Tanya Moore (front row, fifth from right), a PhD in biostatistics (from UC Berkeley) links her interest in math with her growing interest in improving health outcomes in society. Giving back to the community in another way, she has been an organizer of the national Infinite Possibilities Conference for minority women mathematicians; shown here are participants in the IPC held at the University of Maryland, Baltimore County, in March 2012.

completing my doctoral studies, I worked with other women mathematicians and some of my former Spelman professors to create a national conference for minority women mathematicians, the Infinite Possibilities Conference (IPC). Nagambal Shah, Leona Harris, Kimberly Weems, Lily Khadjavi, Erika Camacho, and I have led teams of women to organize and host four IPCs, in partnership with Spelman, North Carolina State University, the University of California, Los Angeles, the University of Maryland, Baltimore County, and the NSF-supported math institutes SAMSI and IPAM; the National Security Agency, the National Science Foundation, and a number of corporations provided generous support. The conferences are modeled after experiences I took for granted

My personal and professional networks have helped me over the years with everything from identifying job opportunities to maintaining hope when I feel discouraged. We are made stronger by our connections.

as an undergraduate; they provide students and professionals with opportunities to connect, share their research and experiences as mathematicians, and interact with the numerous role models in attendance. As Nagambal Shah said in the freshman math course I took at Spelman, "You may have to make the journey alone, but you don't have to be lonely." My personal and professional networks have helped me over the years with everything from identifying job opportunities to maintaining hope when I feel discouraged. We are made stronger by our connections.

As much as I enjoyed learning math in school, there was a part of me that longed to be involved with work that would positively impact the lives of people. Statistics is a powerful discipline, and as the technological capacity to gather and store data becomes greater and greater, those who understand how to make sense of and interpret large amounts of information will be in demand. Biostatistics was one way to link my interest in math with my growing interest in improving health outcomes in our society. Still, I yearned to do work that would have a more direct impact on improving the lives of people.

I have carried with me a thought expressed again and again by my high school African American Studies teacher: No matter what we do in life, we should always remember to give back to our community. I don't think that he necessarily meant for us to take that literally; nonetheless, I currently work for the City of Berkeley. My first task for the city was to create strategies that would reduce health disparities; more recently, I've transitioned to a role that is focused on education.

How does this story lead to where I am now?

My primary role in my job is to work in partnership with the school district, our post-secondary institutions, and community agencies on initiatives that will eliminate the achievement gap that persists in our public schools. Despite Berkeley's reputation for a commitment to social justice and equity, there is still a need to end the racial predictability of student achievement in our schools. Looking back, I see that every experience, every interest I've had plays a role in what I do today. My experience performing in plays assists me when I have to speak publicly, my interest in psychology translates into trying to really listen to what people are communicating, spoken and unspoken, and my inner revolutionary is alive and well as I have the opportunity to do work that supports all of our youth in succeeding academically. But at the core of all that I do is my training as a mathematician. I make direct use of my math skills in different aspects of my work, whether I'm tutoring students in algebra at one of our recreation centers, or assisting with evaluation design, or analyzing data on student achievement. But mostly it's in non-obvious ways that my math training assists me in my work.

One of the reasons I like science fiction is that it dabbles in the realm just outside the world we know, just beyond the knowledge we currently possess. I remember reading an interview with Octavia Butler in which she mentioned that to create her futuristic worlds, all she did was extrapolate from values, priorities, knowledge, and technologies currently in play. As mathematicians, we often reach for that solution, that model, that theory that lies just outside our current understanding. We take the universe of knowledge and facts we have learned and, using our creativity and logic, embark on a journey to uncover the true relationship (that in some dimension already exists). It's a process and a skill that can be valuable in many fields and disciplines.

I'm in the business of solving problems (how to close the achievement gap), and if I assume that a solution already exists, then my job is to collect as much information as possible so that I can reveal the underlying connection. So I try to listen thoughtfully to teachers, principals, youth providers, public health nurses, students, and parents. I read as many articles as I can, and speak with educational experts about best practices. I analyze the data we have on student achievement and well-being and identify strategies for evaluating the impact of our efforts. Then, using logic, reasoning, and intuition, I try to uncover the connections, patterns, and possibilities so that as a community collaborative interested in the future of its youth, we can find and act on meaningful solutions. Part of the reason I am so passionate about education, and in particular about the need for all youth to have solid foundations in mathematics, is my hope that they can not only acquire practical skills, and access to numerous career options, but can become Patternmasters of their own making.

Sue Minkoff (sminkoff@utdallas.edu), of the University of Texas at Dallas, is the editor of the Careers in the Math Sciences column. She was the local organizing chair of the 2012 Infinite Possibilities Conference that was held at UMBC.

Announcements

Send copy for announcements to: Advertising Coordinator, SIAM News, *3600 Market Street, 6th Floor, Philadelphia, PA 19104–2688; (215) 382–9800; marketing@siam.org. The rate is $1.75 per word (minimum $250.00). Announcements must be received at least one month before publication (e.g., the deadline for the June 2013 issue is April 30, 2013).*

GSMMC 2013: Graduate Student Mathematical Modeling Camp

June 11–14, 2013
Rensselaer Polytechnic Institute

The Department of Mathematical Sciences at RPI is pleased to announce the 10th annual GSMM Camp. The camp is a four-day informal workshop in which graduate students work in teams on problems brought by invited faculty mentors. The problems, inspired by real problems that arise in industrial applications, span a wide range of mathematics and are designed to promote problem-solving skills while the team approach is designed to promote scientific communication.

Graduate students at all levels are invited to participate. General information and an online application form can be found at http://www.rpi.edu/dept/math/GSMMCamp/. Financial support for travel and local accommodations is available. The application deadline is April 26, 2013.

Graduate students attending the GSMM Camp are also invited to participate in the Mathematical Problems in Industry (MPI) Workshop, to be held at Worcester Polytechnic Institute during the week following the camp. For further information about MPI 2013, students should visit http://www.math.wpi.edu/MPI2013/.

Math Biology

continued from page 1

of variation. Jewell described an analysis he and colleagues have done of an AIDS study known as the MIRA trial (Methods for Improving Reproductive Health in Africa). The trial, conducted in 2003–2006, assigned initially healthy women at random to a "control arm," in which subjects were given only condoms and intensive condom counseling (the "gold standard" for HIV prevention), or an "intervention arm," in which subjects received a diaphragm and contraceptive gel in addition to condoms and condom counseling.

From an "intention to treat" standpoint, the study was a wash: 158 of the 2472 women in the intervention arm were infected with HIV by the end of the trial, versus 151 of the 2476 women in the control arm. The researchers noted, however, that the intervention arm reported significantly less condom use (53.5%) than the control arm (85.1%), which might have cancelled a protective effect of the intervention. Indeed, a crude analysis adjusting for condom use suggests exactly that. But the Berkeley group found that a more detailed look at the data did not support such a conclusion. Their "direct effects" analysis, which took into account confounding factors and the time-dependent nature of the data, showed that the information in the study was not sufficient to tell whether the differential condom use was masking a benefit from the diaphragm.

New, sophisticated methods, such as direct effects analysis, will help wring the most out of the welter of data facing public health officials and better establish the limits on how the data can be interpreted. If you want to solve 21st-century problems, Jewell says, "you can't use 20th-century statistics."

This Is Your Brain on Math

If public health is beset by confounding factors, the human brain is a tangle of differential equations. Nancy Kopell of Boston University and Emery Brown of MIT and Harvard Medical School gave talks on mathematical aspects of neuroscience, the subject of MBI's current annual program. Kopell has spent decades studying the rhythms of thought; Brown has spent recent years trying to understand exactly what happens when he makes people *stop* thinking.

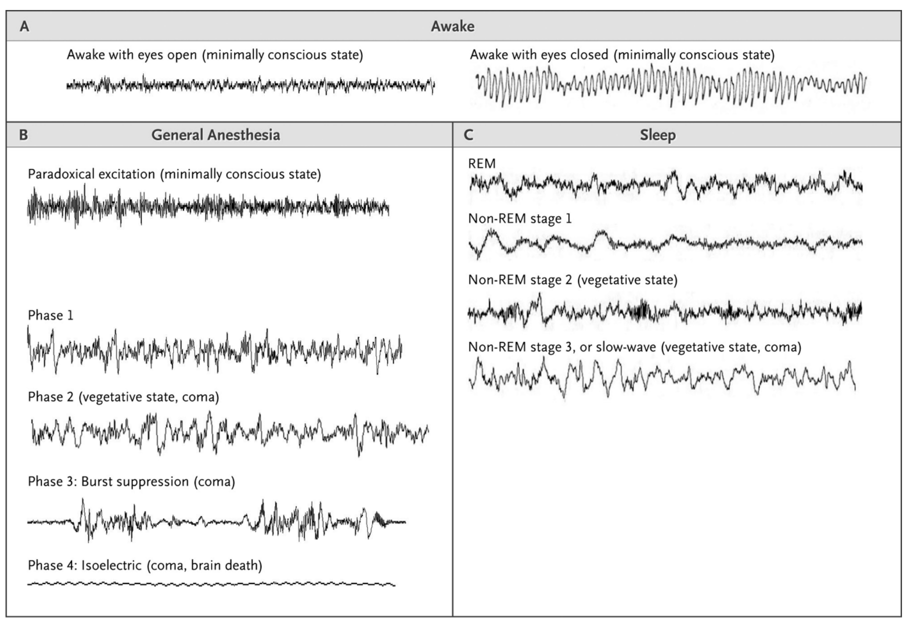
Somewhat more precisely, Kopell and colleagues are trying to understand the brain as a huge system of coupled oscillators. Some aspects of its dynamics are clarified with simple models based on a handful of excitatory and inhibitory cells, while other aspects demand large-scale computer simulations. Kopell described an ambitious project spearheaded by Michelle McCarthy of BU to establish the origin of pathological "beta" oscillations associated with Parkinson's disease. Beta waves span a range of frequencies between 12 and 30 Hertz; everybody has them, but they are pronounced in people with Parkinson's. One theory is that the abnormal oscillations are caused by structural changes in the network connecting fast-spiking interneurons and a class of inhibitory cells called medium spiny neurons in a portion of the brain known as the striatum. Such changes certainly occur. But the BU group has shown, both mathematically and experimentally, that the exaggerated dynamics can result more simply from neuromodulation in a normal network involving only medium spiny neurons.

Their work has implications for schizophrenia as well, for which the relevant oscillations are gamma rhythms (30–90 Hz). Structural changes are also evident in schizophrenics' brains, but the mathematical biologists propose that, again, these may be secondary, the product of pathological dynamics in an otherwise normal network. Viewing things this way could suggest new treatments for brain disorders. A possible testbed, Kopell notes, can be found in studies of anesthesia, in particular of the paradoxical "beta buzz" observed when low (non-knock-out) doses of the anesthetic propofol are administered. Because the effects of anesthetics are clearly not structural in origin, anesthesia might be a good setting for modeling the pathological dynamics that can arise as perturbations of normal ones. "There is a huge amount of work left to be done," Kopell says.

For his part, Brown, who is an anesthesiologist at Massachusetts General Hospital, would like to know how drugs like propofol work. The physiological effects of general anesthesia—chiefly unconsciousness and analgesia (inability to feel pain), but also amnesia and akinesia (inability to move—surgeons don't want patients thrashing about on the operating table)—are well known, but nobody knows how they are achieved. Nor is it understood how the most important part occurs: getting patients to return to consciousness.

"It's rather embarrassing," Brown says, "when you walk into a room and you say you do something every day and you don't know how it works."

Data-gathering opportunities abound: In the United States alone, general anesthesia is induced in nearly 60,000 patients every day. Brown and colleagues have worked with neurosurgeons who operate on epileptic patients who previously had high-density electrode grids implanted in their brains. "It gives us an opportunity to see a fair amount of natural brain," Brown says. (Only a neuroscientist, accustomed to electrodes attached to isolated cells or brain slices, would describe tissue with implants as "natural.") Their studies, which combine a wide array of signal processing techniques, have begun to offer specific hypotheses on how changes in activity level in neural circuits relate to unconsciousness. One of the promising possible applications is to the all-important step of helping patients regain consciousness: Instead of simply letting patients passively come to as the anesthetic wears off, with the common side effects of grogginess and disorientation, not to mention nausea, it might be possible to actively induce emergence in a way that returns the brain to a normal, alert state. Advances in understanding the neural pathways involved in anesthesia will likely further benefit the millions of people every year who go under the knife without having to watch.



EEG-wise, general anesthesia has more in common with being in a coma than with being asleep, but patients prefer the latter, more reassuring term. Emery Brown is among the researchers who are studying mathematically what happens in the brain when drugs like propofol are administered. Reprinted with permission from "General Anesthesia, Sleep, and Coma," December 10, 2010, New England Journal of Medicine *© 2013.*

"A Place Where People Are Shaping a Field"

MBI is one of eight mathematical research institutes now funded by the National Science Foundation. (Its siblings, in alphabetical order, are the American Institute of Mathematics (AIM), in Palo Alto, California; the Institute for Advanced Study (IAS), in Princeton; the Institute for Computational and Experimental Research in Mathematics (ICERM), at Brown University; the Institute for Mathematics and its Applications (IMA), at the University of Minnesota; the Institute for Pure and Applied Mathematics (IPAM), at UCLA; the Mathematical Sciences Research Institute (MSRI), at UC Berkeley; and the Statistical and Applied Mathematical Sciences Institute (SAMSI), in Research Triangle Park, North Carolina.) MBI's first program, in 2002–03, was in mathematical neuroscience, followed by programs on cell processes (2003–04), bioinformatics (2003–04), ecology and evolution (2004–05), and systems physiology (2005–06). The institute is revisiting neuroscience in its current program. It will dive into ecosystems this fall, look at imaging for the life sciences in the spring of 2014, and tackle cancer in 2014–15.

"The institute is about bringing people together," says Marty Golubitsky, who took over as director of MBI in 2008. In remarks at the banquet honoring his predecessor, Golubitsky noted that more than 9000 visitors have attended over a hundred workshops and other events since MBI opened shop. A substantial portion of them are MBI's postdocs, whom Golubitsky calls "the lifeblood of the institute."

Two postdocs also spoke at the banquet. "We were here when MBI was new and exciting," says Janet Best (to laughter at the implication that MBI is now old and exciting). Best, who is now in the mathematics department at Ohio State, was in the second cohort of postdocs, back when the institute was in an old space "centered around an espresso machine—a highly caffeinated but very, very warm experience." She thanked Friedman for being "both a scientific and a professional mentor," and described MBI as "a place you can come and take a chance."

"I had a really fantastic time," enthused Marisa Eisenberg, who was a postdoc from 2009 to 2012 and is now an assistant professor of epidemiology at the University of Michigan. "I feel like I really grew up as a researcher." Eisenberg especially liked the blend of independence and mentoring she experienced. "The community at MBI is amazing," she says. "You get to be in a place where people are shaping a field."

And what shape is the field taking? Mike Reed, the senior scientific adviser of MBI, pointed out that mathematical bioscience has broadened in two important directions. On the one hand, the application of mathematics to medical and field biology is coming to the fore, he says, while on the other hand, "pure mathematics has discovered all the fantastic problems that arise in mathematical biology." As for MBI's role in the field's future, Golubitsky summed it up this way: "We'll just try to do better what we already do well."

Barry A. Cipra is a mathematician and writer based in Northfield, Minnesota.

ICIAM 2015: Call for Prize Nominations and Minisymposium Proposals

The International Council for Industrial and Applied Mathematics invites nominations for the five ICIAM Prizes that will be awarded at ICIAM 2015, in Beijing, China.

All five ICIAM Prizes are international, and each prize has its own special character. Descriptions of the prizes can be found at http://www.iciam.org/council/PrizeDescriptions.pdf.

Nominations should reflect the specifications for each prize, and should contain the following: the full name and address of nominees; web pages, if any; the name of the particular ICIAM Prize for which the nominee is to be considered; a proposed citation (outstanding contributions of the nominee in fewer than 250 words); justification for nominations (reasons for the nomination, including explanations of the scientific and practical influence of the nominee's work and his or her publications); a CV of the nominee; and the name of and contact information for the nominator.

Nominations, preferably in electronic form, should be sent to: Barbara Keyfitz, president of ICIAM, bkeyfitz@math.ohio-state.edu. The deadline for nominations is October 31, 2013. All nominations will be acknowledged.

Members of the ICIAM Prize committee are Barbara Keyfitz (committee chair), Donatella Marini (chair of the Collatz Prize subcommittee), Felix Otto (chair of the Lagrange Prize subcommittee), Pam Cook (chair of the Maxwell Prize subcommittee), Takashi Kako (chair of the Pioneer Prize subcommittee), and Philippe Ciarlet (chair of the Su Buchin Prize subcommittee).

Readers can find more information about the prizes and a call for proposals for thematic and industrial minisymposia for ICIAM 2015 in the first issue of the new *ICIAM Newsletter*, posted on the ICIAM website at www.iciam.org/news.

Foxes, Hedgehogs, and the Art of Prediction

The Signal and the Noise: Why So Many Predictions Fail, but Some Don't. *By Nate Silver, The Penguin Press, New York, 2012, 544 pages, $27.95.*

BOOK REVIEW
By James Case

Nate Silver is a bona fide celebrity. A Google search on his name returns more than 35 million items; his Wikipedia bio is 14 pages long and concludes with a 163-item bibliography. His book (released in August 2012) rose to second place on *The New York Times* best-seller list once it was clear that he had correctly predicted the outcomes in all 50 states in the 2012 presidential election. He had missed by just one in 2008, when Obama bested McCain in Indiana by a single percentage point. He also predicted the outcome of 31 of 33 senatorial races in 2012, missing only in Montana and North Dakota, where lightly regarded Democrats scored narrow upset victories. His 2013 Super Bowl predictions, though incorrect, were breathlessly reported from coast to coast.

Silver got his start as a professional prognosticator by selling a computerized player-evaluation system called PECOTA to *Baseball Prospectus*, an annual publication of little interest to anyone save the hardest of hardcore baseball fans. The acronym, which stands for Pitcher Empirical Comparison and Optimization Test Algorithm, was named after Bill Pecota, a journeyman infielder with the Kansas City Royals during the 1980s. A lifetime 0.249 hitter, Pecota became a recurrent thorn in the side of Silver's beloved Detroit Tigers, against whom his batting average soared to 0.303. As a condition of purchase, Silver was required to extend the PECOTA system to evaluate hitters as well as pitchers, thus more than doubling the reach of the program without changing the acronym.

Growing up in Lansing, Michigan, Silver became hooked on baseball at an early age. He was only six when the 1984 Tigers won 35 of their first 40 games, romped virtually unchallenged to the American League pennant, and defeated the National League Champion San Diego Padres in the World Series. He learned that summer to decipher the morning box scores, and permanently succumbed to the lure of what he still considers "the world's richest data set." Only later, while majoring in economics at the University of Chicago (he subsequently did graduate work at Chicago and the London School of Economics), did he become familiar with *Baseball Prospectus* and begin applying his newly acquired quantitative skills to the wealth of data contained therein. Drawn initially to the raw numbers, he was surprised to find the writing sharp and entertaining.

After college, Silver spent nearly four years in Chicago, working as a "transfer pricing consultant" for the accounting firm KPMG. Despite the good money, congenial colleagues, and apparently secure future, he found the work boring in the extreme. To amuse himself, he began to develop the "colorful spreadsheet" that eventually became PECOTA. Only later, when *Baseball Prospectus* abandoned the computerized system it had been using to identify major league players likely to enjoy notable success in the coming season—information dear to subscribers—did he seek to capitalize on his efforts.

The title of the book is borrowed from radio engineering, practitioners of which struggled for years to design receivers capable of eliminating the noise—due mainly to atmospheric effects—that inevitably corrupts the signal transmitted to listeners at home. Silver argues that all who seek to interpret time series data face a similar problem, as all such data can be viewed as disguising a (wanted) signal with an (unwanted) quantity of apparently meaningless noise. His argument applies with particular force to financial data, which tends to embed a faint or non-existent signal within a cacophony of distracting noise, resulting in theories that model the past very well, yet fail to prove predictive.

Many of humankind's misfortunes result from failure to predict clearly foreseeable events. The recent financial crisis is a case in point.

Many of humankind's misfortunes, Silver writes, result from failure to predict clearly foreseeable events. The recent financial crisis is a case in point. He identifies four separate failures: that of homeowners and investors to foresee that housing prices could not continue to rise indefinitely, that of the ratings agencies to foresee that a fall in housing prices would cause a crisis in U.S. financial markets, that of economists to foresee that a financial crisis in the U.S. was tantamount to a global financial crisis, and that of leadership to foresee that financial crises would produce an unusually long and deep recession. None of the four qualifies as a "black swan"—all were readily apparent to anyone who cared to look "under the hood" of the models then in use.

■■■

In the first half of the book, Silver surveys the current state of the art of prediction. Though his own prognostications have concerned mainly the outcomes of baseball and political contests, he seems to be fascinated by every phase and aspect of the predictive process, eager to learn what works for others and what leads so many astray. The second half of the book examines the ways in which predictions are used in various activities, including competitive ones like chess, war, and portfolio management.

A recurrent theme in the book concerns the need to revise predictions as additional information becomes available. Silver uses the 9/11 attacks to illustrate the process, applying the simplest version of Bayes' theorem, but twice in succession. On 9/10, one might have judged the probability that terrorists would launch such an attack to be $x = 0.005\%$. Once the first plane hits, one should reflect that the probability of such a collision if an attack actually is under way approaches $y = 100\%$, while the probability if no such attack is in progress—as Silver estimates from historical data—is a mere $z = 0.008\%$. The probability that such an attack is under way jumps instantaneously, as soon as the first plane hits, from $x = 0.005\%$ to $xy/(xy + z(1 - x)) = 38\%$. A second application of the theorem reveals that the probability jumps again when the second plane hits, from 38% to 99.99%!

A second recurrent theme is the thinness of the market for accurate predictions. Whereas forecasts obtained from the National Weather Service are meant to be the best possible, local weather shows often purposely overstate the probability

See **Prediction** *on page 6*

Math and Myth

continued from page 1

an attempt to artificially synchronise oral traditions with biblical and classical history. The historicity of the *Táin* is questioned. While some argue that such narratives corroborate Greek and Roman accounts of the Celts and offer us a window on the Iron Age, others object that such tales have no historical basis whatsoever.

■ ■ ■

Social networks have been widely studied in recent years; researchers have looked at the interconnectedness of Hollywood actors, corporate directors, and scientific co-authors, amongst many other examples. These real-world networks all share certain properties. They are highly connected and small-world, a concept related to the famous six degrees of separation notion in sociology. They are also assortative, which means that people tend to associate with people similar to themselves. Their degree distributions (the pattern of links between nodes) are often scale-free, meaning that only a small number of people tend to know a very large number of people.

To construct the social networks underlying each of the mythological narratives, we created databases for the characters and their interactions (see Figure 1). Relationships between characters were categorised as hostile or friendly. The mythological networks were found to have characteristics of real-world social networks, including the small-world property and structural balance (related to the idea that the enemy of my enemy is my friend). The worlds of intentionally fictional narratives, such as *Harry Potter* and *The Lord of the Rings*, also exhibit these properties, as does the society underlying Marvel Comics.

The *Iliad*, however, is assortative, a feature lacking in these intentionally fictional networks. *Beowulf* is also assortative, but only if the eponymous character is removed from the network. The *Táin* is disassortative. These, and other social-network features, appear to corroborate antiquarians' interpretations of the historicity of the tales—the societies underlying the *Iliad* and *Beowulf* have traces of reality, whereas that of the *Táin* appears artificial.

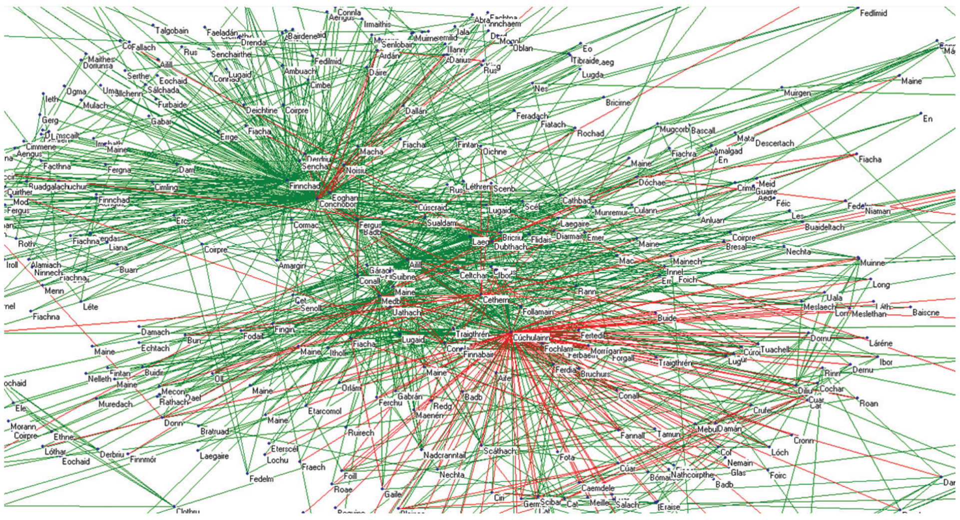


Figure 1. *The central clusters of the social network for* Táin Bó Cúailnge. *Green lines represent friendly links, red lines hostile links, where two characters meet only in combat.*

To investigate the *Táin*'s apparent artificiality, we looked further into the degree distributions of the social networks (Figure 2). This relates to the distribution of popularity amongst characters. Like real social networks, the networks in all three myths are scale-free. We observe a striking similarity between the *Táin* and *Beowulf* for all but the top six characters of the Irish myth (Figure 2, left). The degrees associated with these characters appear inflated, perhaps exaggerated. The beauty of the mathematical approach is that we can identify this anomaly as local—associated with only 6 of the 404 characters in the text. We can then correct for the anomaly by removing the weakest links (i.e., characters they encounter only once) from the six strongest characters. The resulting *Táin* network (Figure 2, right) is assortative, similar to the *Iliad* and to other real social networks and very different from *The Lord of the Rings*, *Harry Potter*, and other examples of fictional extremes, which also lack the scale-free property.

This quantitative investigation is, of course, very different from traditional approaches to comparative mythology—it tells us nothing about qualitative aspects, only about how realistic, or otherwise, the societies appear when viewed as interconnected networks. However, the application of mathematics delivers new perspectives. In particular, it allows us to speculate that, if the *Táin* contains echoes of a bygone age, the six anomalous characters may be based on amalgams of a number of entities and proxies that became fused as the narrative was passed orally through the generations.

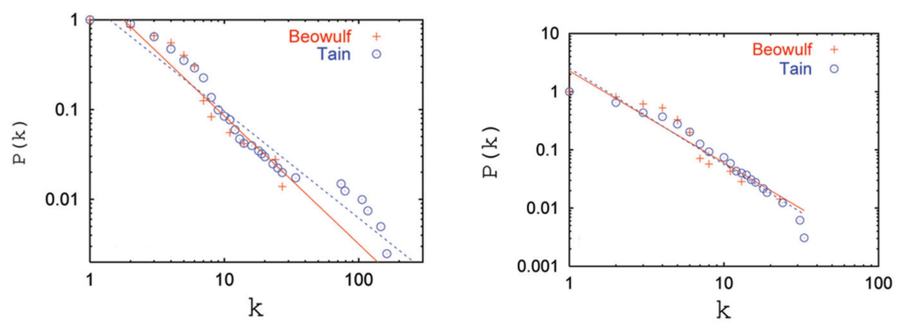


Figure 2. *The degree distribution as a power law for (left) the entire, unadjusted* Beowulf *and* Táin *networks and (right) the adjusted networks. Here, k is the degree and P(k) is its cumulative frequency. Removal of the weakest links of only the six most strongly connected* Táin *characters renders the Irish network similar to that of* Beowulf*, which is believed to be based upon a real society.*

Acknowledgments

We thank Chris Budd for encouraging us to write this article for *SIAM News*.

The work described here is supported by The Leverhulme Trust under grant F/00 732/I.

Related Publication

P. Mac Carron and R. Kenna, *Universal properties of mythological networks*, EPL, 99:2 (2012), 28002; http://iopscience.iop.org/0295-5075/99/2/28002.

Pádraig Mac Carron is a PhD student, and Ralph Kenna is a professor of theoretical physics at Coventry University in England.

Prediction

continued from page 5

of rain. The theory is that a type I error, such as failure to predict rain on a day when it does rain (with its potential to spoil somebody's picnic or golf outing), is more disruptive to viewers than a type II error, such as a prediction of rain on a day when it doesn't rain. In another example, guest commentators on TV talk shows are more likely to be invited back if they venture welcome predictions (think upset victory by the home team over a ranked opponent, or a construction project that should bring jobs to the region) that turn out to be wrong than a dour prediction that turns out to be right. In short, there are many reasons for issuing predictions that are less than the most accurate possible. Silver confines his interest to accurate predictions, and the people who make them.

■ ■ ■

Silver claims to have discovered, in the writings of social scientist Philip Tetlock, the psychological profile of a successful prognosticator. Beginning in 1987, Tetlock started collecting predictions from a broad array of government and academic experts on a variety of topics, including, among other things, the immediate fate of the Soviet Union. Virtually none of those questioned foresaw the troubled nation's impending collapse. Puzzled, Tetlock broadened his horizons, asking experts to venture predictions on the Gulf War, the Japanese real estate bubble, the potential secession of Quebec from Canada, and a host of other then-timely issues. His studies, which he continued for more than fifteen years, were eventually published in the 2005 book *Expert Political Judgment*. His conclusions were damning.

Tetlock's experts proved all but clueless, grossly overconfident, and unable to calculate probabilities. Fully 15% of the events they judged to have no chance of occurring did soon occur, while almost 25% of those they deemed to be sure things never happened! Whether they were predicting events in economics, domestic politics, or international affairs, their collective judgment was utterly unreliable. A few individuals, however, performed conspicuously well.

From their answers to a battery of questions lifted from standard personality tests, Tetlock was able to separate his respondents into two groups: foxes, who demonstrated genuine predictive ability, and hedgehogs, who did not. The names were inspired by a line from an ancient Greek poem: "The fox knows many little things, while the hedgehog knows one big thing." Foxes, he found, tend to be multidisciplinary, adaptable, self-critical, tolerant of complexity, cautious, and empirical, whereas hedgehogs are more likely to be specialized, stalwart, stubborn, order-seeking, confident, and ideological. Foxes provide the more accurate forecasts, but hedgehogs tend to be more entertaining as TV guests. The more interviews the experts had done, the less reliable Tetlock found their predictions to be.

Although he has plenty of other points to make, Silver returns to the fox-versus-hedgehog theme repeatedly. Indeed, he devotes much of the book to arguments that, in activities like forecasting baseball player performance, election results, and hurricane paths, where predictions are made by faceless foxes, the quality of predictions tends to be relatively high and gradually improving. But in fields like economics and climate science, where predictions are more likely to be made by publicity-seeking hedgehogs, prediction quality tends to be and remain low. He has thought long and hard about the art of prediction, and his book on the subject is well worth reading!

James Case writes from Baltimore, Maryland.

Professional Opportunities

Sharing Code

continued from page 8

for many readers or referees, being able to inspect the relevant part of the code is often more valuable than being able to run the code.

If you do publish code, in a paper or on the web, it is worth thinking about the type of copyright or licensing agreement you attach to the code. Your choice may affect the ability of others to reuse your code and the extent to which they must give you credit or propagate your license to derivative works [12].

The code may run only on certain systems today, and nowhere tomorrow. Even apart from the question of proprietary software, many codes have certain hardware or software dependencies that may make them impossible for the average reader to run—perhaps a code runs only on a supercomputer or requires a graphics package that's available only on certain operating systems. Moreover, even if everyone can run it today, there is no guarantee that it will run on computers of the future, or with newer versions of operating systems, compilers, graphics packages, etc. In this way a code is quite different from a proof. So what is the value of archiving code? As in the case of proprietary software dependencies, I would argue that being able to examine code is often extremely valuable even if it cannot be run, and it is critical in making research independently reproducible even if the tables or plots in the paper can't be reproduced with the push of a button.

Of course, authors should attempt whenever possible to make it easy to run their code. From a purely selfish standpoint, any effort put into cleaning up code so that it does reproduce all the plots in the paper with a push of the button often pays off for the author down the road—when referees ask for a revision of the way things are plotted, when the author picks up the research project again years later, or, in the worst case, when results in the paper come into question and the co-author who wrote the code has graduated or retired and is no longer available to explain where they actually came from. To minimize difficulties associated with software dependencies and versions, authors might consider using such techniques as a virtual machine to archive the full operating system along with the code [5], or a code-hosting site that simplifies the process of running an archived code without downloading or installing software [4].

Code is valuable intellectual property. It is true that some research groups spend years developing a code base to solve a particular class of scientific problems; their main interest is in "doing science" with these codes and publishing new results obtained from simulations. Expecting such researchers to freely share their full code might be seen as the computational equivalent of requiring experimentalists publishing a result not only to describe their methods in detail, but also to welcome any reader into their laboratory to use their experimental apparatus. This concern should be respected when advocating reproducibility in computational science, and I don't claim to have a good solution for all such cases.

For many research codes developed by applied mathematicians, however, the goal is to introduce and test new computational methods in the hope that others will use them (and cite their papers). For such codes I see little to be gained by not sharing. The easier it is for readers to understand the details and to implement the method (or even borrow code), the more likely they are to adopt the method and cite the paper. Some people worry that they will not receive proper credit from those who adapt code to their own research. But if everyone were expected to share code in publications, it would be much easier to see what code has been used, and to compare it to the code archived with earlier publications. Citing the original source would then be easy and would become standard operating procedure, leading to more citations for the original author. Readers of mathematics papers can judge for themselves the originality of the ideas in a published proof, and if code development were equally transparent, those developing the original algorithms and code would ultimately receive more credit, not less.

Today, most mathematicians find the idea of publishing a theorem without its proof laughable, even though many great mathematicians of the past apparently found it quite natural. Mathematics has since matured in healthy ways, and it seems inevitable that computational mathematics will follow a similar path, no matter how inconvenient it may seem. I sense growing concern among young people in particular about the way we've been doing things and the difficulty of understanding or building on earlier work. Some funding agencies and journals now require sharing code that is used to obtain published results (see the *Science* guidelines for authors [11], for example). SIAM journals are not currently contemplating such a requirement, but the capability is now available for accepting and publishing unrefereed supplementary materials (including code) in conjunction with papers for some SIAM journals (see "SIAM Journals Introduce Supplementary Materials," *SIAM News*, March 2013). I believe there is much to be gained, for authors as well as readers and the broader scientific community, from taking advantage of this capability and rethinking the way we present our work. We can all help our field mature by making the effort to share the code that supports our research.

References

[1] K.A. Baggerly and D.A. Barry, *Reproducible research*, 2011; http://magazine.amstat.org/blog/2011/01/01/scipolicyjan11/.

[2] N. Barnes, *Publish your computer code: It is good enough*, Nature, 467 (2010), 753; http://www.nature.com/news/2010/101013/full/467753a.html.

[3] S. Fomel and J.F. Claerbout, *Guest editors' introduction: Reproducible research*, Comput. Sci. Eng., 11 (2009), 5–7; http://csdl2.computer.org/comp/mags/cs/2009/01/mcs2009010005.pdf.

[4] J. Freire, P. Bonnet, and D. Shasha, *Exploring the coming repositories of reproducible experiments: Challenges and opportunities*, Proc. VLDB Endowment, 4 (2011), 1494–1497.

[5] B. Howe, *Virtual appliances, cloud computing, and reproducible research*, Comput. Sci. Eng., 14 (2012), 36–41.

[6] J. Kovačević, *How to encourage and publish reproducible research*, Proc. IEEE Int. Conf. Acoust. Speech, and Signal Proc., IV (2007), 1273–1276.

[7] R.J. LeVeque, I.M. Mitchell, and V. Stodden, *Reproducible research for scientific computing: Tools and strategies for changing the culture*, Comput. Sci. Eng., 14 (2012), 13–17.

[8] J. Mesirov, *Accessible reproducible research*, Science, 327 (2010), 415–416.

[9] M. Nielsen, *Reinventing Discovery: The New Era of Networked Science*, Princeton University Press, Princeton and Oxford, 2012.

[10] R.D. Peng, *Reproducible research in computational science*, Science, 334 (2011), 1226–1227.

[11] *Science, General information for authors: Data and materials availability*, 2012; http://www.sciencemag.org/site/feature/contribinfo/prep/gen_info.xhtml#dataavail.

[12] V.C. Stodden, *The legal framework for reproducible scientific research: Licensing and copyright*, Comput. Sci. Eng., 11 (2009), 35–40.

Randall J. LeVeque (rjl@uw.edu) is a professor in the Department of Applied Mathematics at the University of Washington in Seattle. He is chair of SIAM's Committee on Supplementary Materials.

Top Ten Reasons To Not Share Your Code (and why you should anyway)

By Randall J. LeVeque

There is no . . . mathematician so expert in his science, as to place entire confidence in any truth immediately upon his discovery of it. . . . Every time he runs over his proofs, his confidence encreases; but still more by the approbation of his friends; and is raised to its utmost perfection by the universal assent and applauses of the learned world.
—David Hume, 1739*

I am an advocate of sharing the computer code used to produce tables or figures appearing in mathematical and scientific publications, particularly when the results produced by the code are an integral part of the research being presented. I'm not alone, and in fact the number of people thinking this way seems to be rapidly increasing; see, for example, [1–3, 6–8, 10].

But there is still much resistance to this idea, and in the past several years I have heard many of the same arguments repeated over and over. So I thought it might be useful to write down some of the arguments, along with counter-arguments that may be worth considering.

In this article I am thinking mostly of relatively small-scale codes of the sort that might be developed to test a new algorithmic idea or verify that a new method performs as claimed, the sort of codes that might accompany papers in many SIAM journals. It can be at least as important to share and archive large-scale simulation codes that are used as tools to do science or make policy decisions, but the issues are somewhat different and not all the arguments that follow apply directly. However, as computational mathematics becomes increasingly important outside the ivory tower because of these large simulation codes, it is also worth remembering that the way we present our work can play a role in the ability of other scientists and engineers to do credible and reliable work that may have life-or-death consequences. Reproducibility is a cornerstone of the scientific method, and sharing the computer code used to reach the conclusions of a paper is often the easiest way to ensure that all the details needed to reproduce the results have been provided.

This article grew out of a talk with the same title that I gave in a minisymposium, Verifiable, Reproducible Research and Computational Science, at the 2011 SIAM CSE meeting in Reno, organized by Jarrod Millman. (Slides from my talk and others are available at http://jarrodmillman.com/events/siam2011.html.)

A Treatise of Human Nature, http://www.gutenberg.org/files/4705/4705-h/4705-h.htm.

An Alternative Universe

Before discussing computer code, I'd like you to join me in a thought experiment. Suppose we lived in a universe where the standards for publication of mathematical theorems were quite different: Papers would present theorems without proofs, and readers would simply be expected to believe authors who state that their theorems have been proved. (In fact, our own universe was once somewhat like this, but fortunately the idea of writing detailed proofs grew up along with the development of scientific journals. See, for example, Chapter 8 in [9] for a historical discussion of openness in science; I highly recommend the rest of this book as well.)

In this alternative universe, the reputation of the author would play a much larger role in deciding whether a paper containing a theorem could be published. Do we trust the author to have done a good job on the crucial part of the research not shown in the paper? Do we trust the theorem enough to use it in our own work in spite of not seeing the proof? This might be troubling in several respects. However, there are many advantages to not requiring carefully written proofs (in particular, that we don't have to bother writing them up for our own papers, or referee those written by others), and so the system goes on for many years.

Eventually, some agitators might come along and suggest that it would be better if mathematical papers contained proofs. Many arguments would be put forward against the idea. Here are some of them (and, yes, of course I hope you will see how similar they are to arguments against publishing code, mutatis mutandis, and will come up with your own counter-arguments):

1. *The proof is too ugly to show anyone else.* It would be too much work to rewrite it neatly so that others could read it. Anyway, it's just a one-off proof for this particular theorem; my intention is not that others will see it or use the ideas for proving other theorems. My time is much better spent proving another result and publishing more papers rather than putting more effort into this theorem, which I've already proved.

2. *I didn't work out all the details.* Some tricky cases I didn't want to deal with, but the proof works fine for most cases, such as the ones I used in the examples in the paper. (Well, actually, I discovered that some cases don't work, but they will probably never arise in practice.)

3. *I didn't actually prove the theorem—my student did.* And the student has since graduated, moved to Wall Street, and thrown away the proof, because of course dissertations also need not include proofs. But the student was very good, so I'm sure the proof was correct.

4. *Giving the proof to my competitors would be unfair to me.* It took years to prove this theorem, and the same idea can be used to prove other theorems. I should be able to publish at least five more papers before sharing the proof. If I share it now, my competitors will be able to use the ideas in it without having to do any work, and perhaps without even giving me credit, since they won't have to reveal their proof technique in their papers.

5. *The proof is valuable intellectual property.* The ideas in this proof are so great that I might be able to commercialize them someday, so I'd be crazy to give them away.

6. *Including proofs would make math papers much longer.* Journals wouldn't want to publish them, and who would want to read them?

7. *Referees would never agree to check proofs.* It would be too hard to determine the correctness of long proofs, and finding referees would become impossible. It's already hard to find enough good referees and get them to submit reviews in finite time. Requiring them to certify the correctness of proofs would bring the whole mathematical publishing business crashing down.

8. *The proof uses sophisticated mathematical machinery that most readers/referees don't know.* Their wetware cannot fully execute the proof, so what's the point in making it available to them?

9. *My proof invokes other theorems with unpublished (proprietary) proofs.* So it won't help to publish my proof—readers still will not be able to fully verify its correctness.

10. *Readers who have access to my proof will want user support.* People who can't figure out all the details will send e-mail requesting that I help them understand it, and asking how to modify the proof to prove their own theorems. I don't have time or staff to provide such support.

Back to the Real World

Of course, sharing code and publishing proofs are different activities. So let's return to the real world and examine some of the arguments in more detail.

It's just a research code, not software designed for others to use. General-purpose software designed to be user-friendly obviously differs from research code developed to test an idea and support a publication. However, most people recognize this difference and do not expect every code found on the web to come with user support. Nor do people expect every code found on the web to be wonderfully well written and documented. The more you clean it up, the better, but people publish far more embarrassing things on the web than ugly code, so perhaps it's best to get over this hangup [2]. Whatever state it is in, the code is an important part of the scientific record and often contains a wealth of details that do not appear in the paper, no matter how well the authors describe the method used. Parameter choices or implementation details are often crucial, and the ability to inspect the code, if necessary, can greatly facilitate efforts of other researchers to confirm the results or to adapt the methods presented to their own research problems.

Moreover, I believe that it is actually extremely valuable to the author to clean up any code used to obtain published results to the point that it is not an embarrassment to display it to others. All too often, I have found bugs in code when going through this process, and I suspect that I am not alone. Almost everyone who has published a theorem and its proof has found that the process of writing up the proof cleanly enough for publication uncovers subtle issues that must be dealt with, and perhaps even major errors in the original working proof. Writing research code is often no easier, so why should we expect to do it right the first time? It's much better for the author to find and fix these bugs before submitting the paper for publication than to have someone else rightfully question the results later.

It's forbidden to publish proprietary code. It is often true that research codes are based on commercial software or proprietary code that cannot be shared for various reasons. However, it is also often true that the part of the code that relates directly to the new research being published has been written by the authors of the paper, and they are free to share this much at least. This is also the part of the code that is generally of most interest to referees or readers who want to understand the ideas or implementation described in the paper, or to obtain details not included in it. The ability to execute the full code and replicate exactly the results in the paper is often of much less interest than the opportunity to examine the most relevant parts of the code.

Some employers may not allow employees to share any code they write. However, if authors are allowed to publish a piece of research in the open literature, then in my view they should be allowed to publish the parts of the code that are essential to the research. After all, employers cannot forbid authors to publish proofs along with their theorems—referees would not put up with it. A change in expectations may lead to a change in what's allowed. Moreover, publishing code can take various forms. Some employers may forbid sharing executables or source code in electronic form, but impose far fewer restrictions on publishing an excerpt of the code (or even the entirety) in a pdf file. It is worth reiterating that,

See **Sharing Code** *on page 7*