Sam Wang

Sam Wang is a pioneer in using statistical methods to analyze U.S. presidential election polls. In this election season interview with *Neuron*, he discusses diagnosing partisan gerrymandering, his research on the cerebellum, and how he analyzes polls with high accuracy.

Sam Wang is a professor of neuroscience and molecular biology at Princeton University. His laboratory studies the cerebellum's role in action, cognition, and social thought processes using in vivo optical imaging, viral tracing, neuroanatomy, and statistical analysis. They are especially interested in the cerebellum's contributions to normal mental development and to autism spectrum disorder and are investigating how transient disruption of cerebellar activity in early postnatal life can have lasting effects on cognitive and social processing. Dr. Wang earned a BS with honor in physics from Caltech and a PhD in neurosciences from Stanford University with Stuart Thompson. Before coming to Princeton, he did postdoctoral research at Duke University with George Augustine and at Bell Labs Lucent Technologies with David Tank and Winfried Denk. He also worked on education and research policy for the U.S. Senate Committee on Labor and Human Resources with Senator Edward M. Kennedy as a Congressional Fellow of the American Association for the Advancement of Science (AAAS). Dr. Wang is a past National Science Foundation Young Investigator, an Alfred P. Sloan Fellow, a W.M. Keck Foundation Distinguished Young Investigator, and recipient of a McKnight Technological Innovations in Neuroscience Award. He has co-authored two popular books, *Welcome to Your Brain* and *Welcome to Your Child's Brain*, which have been translated into over 20 languages. Dr. Wang is also noted for pioneering the use of statistical methods to analyze U.S. presidential election polls with unusually high accuracy. He publishes data analytics and election commentary at the Princeton Election Consortium, http://election.princeton.edu. In 2012 his methods correctly predicted all state Presidential races, the national popular vote, and every Senate race.

Sam Wang
Neuroscience Institute and Department of Molecular Biology, Princeton University

What big question would you like to see answered next in your field?

Currently there is a lot of excitement about brute-force mapping of a connectome—but little emphasis on how the circuitry becomes wired up in the first place. It takes years of postnatal experience to build a brain. It would be wonderful to know, in mechanistic detail, how experience acts to help establish a functional circuit. Many people think of the brain as a computational object that is programmed to make sense of incoming information and act appropriately. But contrary to the brain-as-computer metaphor, the brain does not come out of a box ready to go. Somehow, a baby's brain starts with reward learning and other principles of plasticity such as Hebbian learning and eventually ends up with sophisticated circuits for social processing, language, and cognitive acts. How on earth does that happen?

To tackle that question, is there a tool that either needs to be developed or is currently available that could be implemented in a novel way?

As a first step, many tools are available now and simply need to be applied in an artful manner. Pharmacogenetics and optogenetics can be used to perturb brain function at specific developmental periods, and these can be combined with measurements of behavior and neural activity. That is largely a matter of using existing technology.

To look into the future a bit, a problem with current connectomic approaches is that they kill or destroy the sample. It would be cool to be able to map circuitry repeatedly over time, with enough detail to link experience and activity to the consequent changes.

How do you view the level of crosstalk between disciplines? And where do you see the strongest potential for progress and new breakthroughs in neuroscience?

Usually when people talk about such crosstalk, they often mean "outside" areas such as physics or molecular biology or social science. I am excited about the possibilities for cross-fertilization *within* neuroscience, which contains multiple intellectual traditions. For example, I gained an appreciation for the importance of experience-dependent CNS development in graduate school, and for biophysical and computational approaches in postdoctoral work. Both areas are intellectually rich, and I see an opportunity there to bridge a gap.

In one project in my lab, we are extending Wiesel and Hubel's discovery of sensitive periods beyond sensory input to include within-brain communication. Cerebellar injury at birth confers more risk for autism than genetics or environmental factors. We are using pharmacogenetics, optogenetics, and behavioral testing to determine when and how cerebellar activity influences forebrain processing and structural plasticity. Our technical approach uses in vivo two-photon microscopy, miniscopes, pharmacogenetics, and optogenetics, and behavioral monitoring. We are also tracing long-distance anatomical pathways from the cerebellum using viral approaches. These projects require an arsenal of BRAIN initiative-type tools.

Bridges are also starting to be built between systems and cognitive neuroscience. In some sense, cognitive neuroscience is a successor of psychology, which identified many important problems a long time ago. It feels like time to start working out these problems with circuit-level rigor.

When you get into a new research area, how important is it to understand the concerns of experts?

I like to get into a new subject every 5 years or so, which in academic circles is considered a short attention span. I think outsiders bring fresh perspectives. While my first lab was being renovated at Princeton, I got into brain evolution. I found an exciting area where computational tools and theoretical analysis allowed me to take a fresh look at how brains scale up across phylogeny. That was great fun.

Now I am relatively new to another field, autism research, where there is a large body of clinical, genetic, and other biological observations. There is a lot of opportunity to bring in mechanistic and systems-level thinking. An outsider can bring a fresh look at old data, and I hope contribute to a systems-level conceptual framework to bridge gaps between genes, molecules, and the autistic person.

Learning from experts can be essential. In the domain of statistical politics, I have become interested in the question of partisan gerrymandering. My interest arose in 2012, when I correctly predicted that Democrats would win the national House popular vote, but I missed the fact that Republicans would retain control, in part because of gerrymandering. Simple statistical tools, some involving just pencil and paper, can help diagnose gerrymandering. This is important because legal precedents interfere with obvious criteria like the shapes of districts. I have recently published an article in the Stanford Law Review (http://www.stanfordlawreview.org/wp-content/uploads/sites/3/2016/06/3_-_Wang_-_Stan._L._Rev.pdf) outlining how courts can easily apply these tools. This required learning enough law to come up with a gerrymandering standard that judges might actually adopt.

What lessons from statistics do you wish most neuroscientists would implement in their research?

I think significance testing is overrated. The basic problem is that unless you have a well-founded reason for why you're doing the experiment, there is a risk of false positives. This is similar to the "replication crisis" that has been brewing in social sciences, where often they don't know what the mechanisms are. An interesting finding happens to be statistically notable, but then fails to be confirmed.

For many experiments it is often better to look first at the effect size (the effect, measured in units of standard deviation). The effect size puts into perspective whether your finding is robust enough to be probably important, or whether it is susceptible to being contradicted by the next study to come along.

What advice do you find yourself giving to your students and postdocs?

First, two contradictory apothegms: be open to technical and intellectual criticism—but also listen to your internal voice of what excites you. Second, think about how your skills will be useful both in the lab today and also in the future: data analysis, project management, public communication, and human relations. These skills turn out to be useful in many situations.

You've written for popular audiences about autism risk, gerrymandering, and the myth of brain training. How much statistical training does it take for a scientist to contribute to a public discussion?

Not much at all! Many readers of this column could do what I do. It's not limited to the political domain. News coverage of science often fails to put a finding's true importance into perspective. Simple tools like odds ratios and effect sizes can go a long way.

How did your interest in politics emerge?

I have long been interested in the power of politics to improve—or worsen—people's lives. In 1978, as a sixth-grader in a southern California public school, I got interested in a ballot initiative called Proposition 13, which drastically cut property taxes, which fund public schools. I liked school! So I handed out literature against Prop. 13. Our side lost—deeply wounding California schools and setting the stage for anti-tax battles nationwide.

Nearly two decades later, I had an even more direct experience. As AAAS Congressional Science and Engineering fellow, I spent a year working in Congress. The experience broadened my horizons tremendously. I analyzed education and research policy analysis and wrote a lot of speeches. It was suggested to me that I consider doing policy work for the White House. However, as much as I enjoyed legislative work, I realized that what I really wanted was to come back to neuroscience and make some concrete contributions in my field.

Was there any lasting impact of that foray?

Yes, it exposed me to a wide range of ideas and topics, both scientific and nonscientific. It is easy to forget the pleasure of doing new things but that year restored it for me. For me, stepping away from the laboratory was an extremely valuable experience.

You have a side activity that has attracted a lot of media attention. Can you describe for the *Neuron* audience the Princeton Election Consortium and your experience providing analysis of U.S. election polls?

First, the arcane process of picking a President: the Presidency is determined by a rule in which each of the 50 states, plus the District of Columbia, is assigned electors who vote on the outcome. Generally, each state is winner-take-all.

Therefore state-level opinion gives direct information about how electors will be assigned on Election Day. And state-level general election polls are, on average, highly accurate.

The Princeton Election Consortium converts publicly available polling data to a win probability for each of the 51 state/D.C. races. Then it calculates the exact probability distribution of an election based on those probabilities. That snapshot provides a way to follow the race, not unlike a hurricane tracker. It is then possible to make a prediction by assuming a combination of bounded random drift, in conjunction with a Bayesian prior that says the race will tend to stay in a range where it's already been. It's the opposite of the Star Trek aphorism; I assume that opinion will go "where every [poll] has gone before."

Similar methods are used to follow Senate and House races. The site also shows readers where their contributions are most effective in driving the outcome probabilities. The big goal of the site is to show where things are probably headed and help people leverage their time and money intelligently.

How did you get started on poll aggregation?

In 2000, Ryan Lizza compiled polls at *The New Republic*. His data showed that the Presidential race would come down to Florida. That year the Society for Neuroscience meeting spanned Election Day, and my friends and I threw a large party in the Faubourg Marigny district in New Orleans. The owner of the club hauled a television into the bar so people could watch returns. As our hired brass band played on, Florida was resolved—or so I thought. Back at the meeting the next day, Marla Feller was the first to alert me that the election had gotten un-resolved. The bad news was that the dispute was a national disaster in the making. The good news is because of opinion polls, we saw it coming.

In 2004, I created a website to analyze the contest between Democratic Senator John Kerry and Republican President George W. Bush. That year, I wrote MATLAB scripts to calculate poll medians for each state, converted the medians to probabilities, and calculated an exact distribution of electoral-vote outcomes. To my surprise, the site went viral, eventually getting onto the front page of The Wall Street Journal. Apparently there was an unmet need for quantitative insights into elections! Since 2008, some Princeton students and I have automated the process at http://election.princeton.edu. And in 2012 we added long-term predictions, with excellent results.

Do you think that so much media attention to polls is a good thing?

Actually, no. Originally, I thought my hobby could cut through the noise by reducing the amount of horserace coverage. It would make room for journalists to write about Senate and House races, and even the policies that make politics important in the first place! But my heart has been broken—so much coverage still focuses on who is up and who is down. Sad!

How is web publishing different from peer review?

Academic writing has rules of engagement that involve giving credit to those who came earlier. Other geeks and I were the vanguard for what is now a whole cottage industry. If there were any justice, you'd know more about Andrew Tanenbaum at http://electoral-vote.com, a foundational figure from 2004—and not just that famous guy who came 4 years later.

On the up side, I get my own version of peer review through commenters at my web site. I have legal scholars, string theorists, financial traders, political activists—it's one of the smartest comment sections on the Web. I have learned a lot from these people! If the Princeton Election Consortium ever collided with a YouTube comment thread, the resulting annihilation would leave nothing but photons.

Describe for neuroscientists the types of statistical insights you use for analyzing polls with such high accuracy

Just like single data points in the lab, single polls are not all that great by themselves. Pollsters survey enough people to make their sampling error smaller than their systematic error—but both kinds of uncertainty still exist. It's necessary to take the median of the margin between the candidates. Averaging isn't as good because of the possibility of outliers. State polls are better than national polls, probably because it's easier to get a representative sample of a small part of the country.

Conversion to a win probability is not hard. Divide the median by the standard error to get a Z score and convert that to a win probability. Do that 51 times, compound the distribution, and you have a snapshot of the Presidential race.

Do you see any similarity between neuroscience and polling analytics?

Yes! For example, if you're trying to understand the properties of neurons, you get a clearer result by sampling one brain region or cell type than by attempting to survey the entire brain. In this analogy, Ohio could be a pyramidal neuron and Florida could be a Purkinje cell. Anyway, a well-defined sample is absolutely central, in neuroscience and in polling.

Once you have data, there's just not a big distinction. Both domains involve extracting a useful signal out of lots of noisy measurements. Whether it's axon diameters, calcium signals, or behavioral measurements, single observations are part of a larger picture. In all cases, I look for an analysis that reduces the observations to a small number of parameters.

In what ways has your professional experience influenced how you think about elections and statistics?

I find much journalistic coverage of polls to be remarkably bad. No scientist would ever want to look at a graph one data point at a time, but this is common in the press. And when a news organization has multiple polls available to it and chooses to report on the outlier, it is an act of reporting malpractice.

Very few journalists choose their craft out of a love of math. Nonetheless, math can help them do their jobs accurately and well. It would do political reporters a world of good to follow some remedial principles. If multiple polls are available, take the median. Unlike averaging, this does not even require addition or division!

When you put it that way, poll aggregation does not sound very complicated.

It's not! But it does require some patience: a willingness to wait for more than one data point before reaching a conclusion.

Going in the other direction, does work on the Princeton Election Consortium illuminate your neuroscience research?

I have come to value transparency in my data analysis more. When we analyze data in the lab, sometimes the first approach we take is fairly technical. If the analysis has fewer steps or the original observations are in some way recognizable, then it is easier to maintain quality control and to convince readers of a finding. Whether it's in vivo two-photon fluorescence imaging or animal locomotion, I prefer descriptive measures where I can discern the original units or measurements in some way.

Data analysis challenges in my lab include calcium imaging measurements from two-photon microscopy and the monitoring of behavioral states. I have great postdocs and students who come armed with all kinds of mathematical tools. I often find myself in the position of encouraging them to put away the fancy analysis and take a simpler, more transparent approach.

What do you do when you're not in the lab?

Some people like to garden. Some people restore old cars. I calculate political poll medians and try to make it sound interesting. This is my one hobby.

http://dx.doi.org/10.1016/j.neuron.2016.10.010