

Experimental neuroscientists
and theorists at Janelia are joining
forces to make sense of – and improve –
the deluge of data coming out of labs.

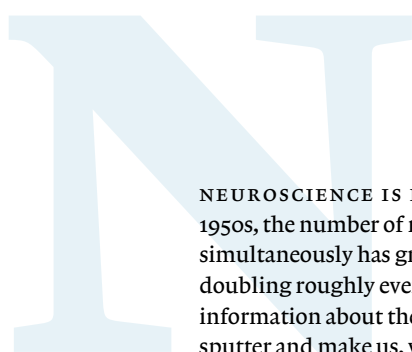
Navigating the Torrent

BY CARRIE ARNOLD



ILLUSTRATION BY CARMEN SEGOVÍA





NEUROSCIENCE IS DROWNING in data. Since the 1950s, the number of neurons that scientists can record simultaneously has grown at an exponential pace, doubling roughly every seven years. To utilize this information about the billions of neurons that spit and sputter and make us, well, human, researchers have had to cope with an exponential growth in data.

“What you would do, back in the day, is maybe look at a couple of neurons in one part of the brain during simple sensory stimulation – a very focused study,” says Jeremy Freeman, a neuroscientist and group leader at HHMI’s Janelia Research Campus.

Now, the pendulum has swung in the opposite direction. Neuroscientists can record the activity of nearly all the neurons in the brains of zebrafish larvae, and in ever increasing portions of mouse and *Drosophila* fruit fly brains.

A single set of experiments can generate terabytes of information. Simultaneous increases in computing power mean researchers can perform more sophisticated analyses, studying relationships between groups of neurons instead of analyzing one neuron at a time. To stay afloat in the deluge of data, scientists need to develop an entirely new way of thinking about experiments – and making sense of the resulting torrent of information.

“It’s really a big change from thinking about what single neurons do to

thinking about what large populations of neurons do. With a single neuron, an experimentalist could use his or her intuition and, in fact, people are quite good at that,” says Larry Abbott, a Janelia senior fellow and a theoretical neuroscientist at Columbia University. “When you have a population of neurons, and they’re all interacting, it’s almost impossible to have that intuition. You really have to make a model of it and figure out how you think it’s going to behave.”

Abbott and Freeman, together with other theoretical and computational neuroscientists at Janelia, are

working to build lifeboats and lighthouses for other scientists to help them navigate the swirling storms of data. Buried in this tsunami of statistics are the patterns and insights that will enable them to crack the biggest mystery in science: how the human brain works.

Computation as Partner

To Janelia Group Leader Kristin Branson, keeping your head above water as data pours in requires computer science as much as it does biology. She joined Janelia in 2010, intending to improve the computer tracking software known as Ctrax that she had built while she was a postdoc at the California Institute of Technology. The premise behind Ctrax was simple. At the time, measuring the effects of neural activity on behavior meant measuring how a fly’s behavior changed after a group of neurons was switched on or off. To make sense of this behavior, a scientist would have to determine what the fly was doing in each frame of video – an impossible task when a single experiment can yield days of video.

Enter Ctrax. Using a variety of computer algorithms that allow a machine to process and analyze images, Ctrax enables researchers to track individual flies even when they congregate in large groups. Unlike other programs available to biologists, Ctrax doesn’t require users to know how to code. Instead, Branson created a GUI (pronounced “gooey” and standing for “graphical user interface”) that allows even non-coders to use the program.

When she first arrived at Janelia, Branson had ideas for improving Ctrax. But the number of other Janelia scientists also working on tracking software gave her pause, as did her realization that the tougher problem would be analyzing the fly’s behavior, not just tracking it. So Branson developed JAABA, the Janelia Automatic Animal Behavior Annotator, which is freely available to all researchers through Branson’s website. Researchers can “teach” JAABA the relevant behaviors to recognize and record, which allows them to begin to figure out what happens to the animal when neural activity is altered.

“The idea is you want to automatically be able to say for every frame and for every fly, Is this fly doing this certain behavior or not?” Branson explains. “Is this fly chasing another fly? Is it walking? Is it turning?”

Branson has begun using Ctrax and JAABA to screen all of the neurons in the fruit fly brain, to link specific groups of cells to behaviors. To begin, she took advantage of the 10,000 lines of fruit flies, created by Janelia Executive Director Gerry Rubin and his lab group, that express the protein GAL4, which scientists use to affect gene expression in different cells. Scientists can select the flies that express GAL4 in specific sets of neurons and then use the fluorescent marker GFP to visualize these neurons with a microscope. Branson crossed 2,000 different lines of GAL4 flies with flies containing *TrpA*, a temperature-sensing gene, which allows her to activate these neurons by raising the fly’s body temperature a few degrees. Placing these flies in an enclosed arena, Branson tracks them with Ctrax and records their behavior with



Jeremy Freeman built an open-source computer platform to help researchers analyze and share massive data sets.

“If you could give me the structure of the circuit, it would serve as crucial inspiration for hypotheses regarding what the circuit is trying to do.”

—SHAUL DRUCKMANN

JAABA. Used together, the systems give her a way to link the activity of specific neurons in flies with measurable behavior differences, such as walking speed.

Machine learning systems have benefits beyond just managing tidal waves of data; Branson has also found that these systems can recognize effects that might go unnoticed by humans.

“It might be a subtle behavioral change that’s happening, but you’re observing it lots and lots of times. So if you analyze a big enough data set, it’s going to come out,” she says. “Maybe it’s not a huge phenotype that you’re seeing – it’s not something that a human would typically notice – but if you compound this with how many times you’ve seen it, it’s something that can’t be ignored.”

Massive Computation

Ctrax and JAABA may be able to help neuroscientists identify and catalog relevant behaviors, but researchers also have other types of data to analyze. The problem, Freeman points out, is that some of these data sets are far too big for a single computer to handle. So researchers are turning to cluster computing, using multiple computers working together to make sense of their findings.

Historically, Freeman says, scientists have worked to solve these types of problems on a lab-by-lab basis. “Individual labs build their own little local solutions to solve problems that are totally customized to their labs and are meant to work on a single machine,” he says. “We’re at

a point where an individual lab is starting to reach a limit of what it can do in a reasonable amount of time.”

This process of continually reinventing the wheel didn’t seem productive, nor did it allow scientists to easily share their data with each other. So Freeman, in collaboration with others at Janelia and elsewhere, built an open-source computer platform that would allow researchers to analyze and share massive data sets. He utilized the Apache Spark platform – an open-source framework able to process large data sets on computer clusters – to create Thunder, a library designed specifically to analyze neural data. Detailed in *Nature Methods* in September 2014, the library is freely available for researchers to use and contribute to. Freeman estimates that Thunder is now being used by some 10 to 20 labs around the world. He’s using it to look at large data sets to gain a more holistic understanding of neural function.

In one of his collaborations, he has been working with fellow Janelia Group Leader Misha Ahrens to image the entire brain of the zebrafish using two-photon microscopy, which can record fluorescing cells at a depth of up to one millimeter. Freeman and Ahrens display a visual pattern in front of the fish. They want to know which neurons fire when the fish sees the pattern, which ones are active as it tries to swim, and whether the animal’s movement is feeding back into the neural activity. Freeman hopes to answer these questions both at the level of the entire brain and on the level of individual neurons. Just an hour of these recordings, however, can generate more than a terabyte of data (the rough equivalent of 16 million books), making the use of Thunder or another type of cluster computing software a necessity.

The analyses performed by Thunder, however, are only as good as the experiments that provide the data. To Freeman, theory and computation aren’t just things you do after you get your results – they need to be integrated into every aspect of the scientific process.

“If people handed me data, and I went away for six months and analyzed it, it would be boring,” he says. “That’s not really the way to progress. You need to be constantly interacting and finding cool stuff in the data together.”

Modeling Networks

The close marriage of computation, theory, and experiment isn’t unique to biology – it’s long been a feature of physics. Perhaps that explains why so many theoretical neuroscientists, including many of those at Janelia, got their start not in biology but in physics. Shaul Druckmann, for example, was all set to start a graduate program in high-energy physics but changed his focus to neuroscience after



Kristin Branson uses computation and machine learning to understand behavior.

This visual representation of neurons from the larval zebrafish brain shows calcium responses in the cells – an indicator of activity – measured using light sheet microscopy. Each circle marks a neuron’s position, with colors based on a functional categorization. Curves connecting the circles reflect a measure of neuronal coupling.



hearing a lecture by a theoretical physicist who worked on neural networks. After completing his PhD in computational neuroscience at Jerusalem’s Hebrew University, where he concentrated on detailed models of single neurons, Druckmann moved to Janelia for a postdoc position in 2013 and was subsequently hired as a group leader. Since coming to the research campus, he has focused on developing models of working memory, the process that holds information for brief periods of time.

Researchers had previously believed that if you asked mice to remember something for a short period of time, their neural activity would be relatively constant, since all they were doing was remembering. But when scientists actually recorded what the neurons were doing, they found that neural activity in the mice was all over the map, constantly changing. Druckmann wanted to build a model to show why such activity was always shifting. During his postdoc, Druckmann was able to show that a simple model, whereby neurons exchange information between themselves in a specific manner, shows how shifting activity is perfectly consistent with representing information that isn’t changing over time.

“The problem in the brain is that typically you have networks of neurons that are

extremely strongly connected to each other, so A triggers B, B triggers C, and then C goes back and triggers A. But it also triggers D, E, and F, and each one of these also triggers X, Y, and Z, and they all feed back into each other,” says Druckmann. “And, it’s all nonlinear, which means it is considerably more difficult to intuit causes from effects, and all of these models become harder to think through.

Once one does the math, and understands how this can come about, it is straightforward to explain the intuition behind it, but without going through the calculations, it would be hard to reach that intuition.”

At the same time, another Janelia group leader, Karel Svoboda, was testing short-term memory in mice, which gave Druckmann a chance to see if his ideas would match experimental data. The memory task presented to the mice was relatively straightforward. First, the researchers trained a mouse to respond to a sensory cue – a pole it could locate with its whiskers. Depending on the location of the pole, the mouse would respond by moving right or left after hearing a beep. Importantly, the sound was not played immediately after the mouse found the pole – the animal had to wait seconds before the beep sounded. This meant that the mouse had to remember both the location of the

pole and the direction in which it needed to move while it waited. Svoboda’s team also used optogenetics – a technique that allows scientists to control neural activity with light – to strategically turn off different groups of neurons in the mouse brain cortex. They found that activity in an area of the brain called the anterior lateral motor cortex was crucial to the animal being able to perform the memory task. When scientists switched off that area, performance dramatically declined.


Now that he has this information, Druckmann is revising his models to more accurately reflect Svoboda’s data. By understanding these fluctuations in neural activity on time scales of just a few hundred milliseconds, Druckmann hopes to understand what computations are going on during this time period and what kind of mechanisms may be responsible for it. A key piece of information that is hard to get at is the structure of the brain’s neural networks, which can tell scientists a lot about how the different parts of the brain function. “If you could give me the structure of the circuit, it would serve as crucial inspiration for hypotheses regarding what the circuit is trying to do,” Druckmann says.

Unifying Principles

Cracking just one neural network can give scientists insight into many other networks. Theoretical neuroscientist Sandro Romani, also a group leader at Janelia, spent the early years of his career modeling how primates perform working memory tasks. During his postdoc at the Weizmann Institute of Science in Israel, he worked to understand how humans remember long lists of words. Later, at Columbia University, he went on to study neural circuit dynamics in the hippocampus, a seahorse-shaped region deep in the brain



Shaul Druckmann focuses on developing models of working memory, the brain’s ability to hold information short-term.

 To see more of the data being analyzed by Janelia's theorists, go to hhmi.org/bulletin/spring-2015.

“Theory is one of the tools in designing and getting the most out of an experiment.”

—LARRY ABBOTT

that controls navigation and long-term memory. Specifically, Romani was looking at recordings from place cells in the hippocampus of the rat, each of which maps a specific place in the animal's world. Ask a rat to run a familiar maze, and researchers can track where it is located by watching which place cells fire in its brain.

When scientists looked at this firing pattern more closely, however, they found that the signals were even more intriguing than they'd expected. In particular, they noticed a regularly repeating sequential activation of the place cells that moved more quickly than the rat did, indicating that the animal's movement wasn't driving this activity. Instead, the cells' activities were likely predicting the rat's future movements.

Just as a rat explores nearby places – which are similar but not the same – in a sequence, items in memory tend to be recalled according to how similar they are. For example, grabbing a box of cake mix at the grocery store reminds you to pick up frosting as well, instead of triggering your memory that you also need steak. Scientists study this kind of memory by asking people to remember a long list of words in what's known as the free recall task. As the list becomes longer, people can recall more words but the fraction of the total gets smaller.

Romani and colleagues realized that the nervous system may use similar dynamics for both sequences of places and sequences of words, and they developed a neural network model to show that this can be done with a realistic neural architecture.

At Janelia, Romani has worked with Group Leader Eva Pastalkova on a modified version of the model to account for some of her experimental results on a group of hippocampal neurons she has dubbed “episode” cells. These neurons are active in a sequence when a rat stops its motion through the environment while it is engaged in a memory task.

“One contribution that theoretical neuroscience can bring to the table is to step back a little from the phenomena and try to find unifying mechanisms and principles behind them,” Romani says.

Theoretical models, Abbott points out, are another set of tools for people designing and carrying out experiments – and should be treated as such. “In the past, you generated a whole bunch of data and you just threw it at a theorist and said, ‘Well, what does it mean?’ But it's better to have people with different expertise *in* the process,” he explains, “the same as you might have an expert microscopist in the process and say, ‘How do we design this experiment so we can get

the best use out of the tools?’ Theory is one of the tools in designing and getting the most out of an experiment.”

It's a process Abbott himself uses while trying to understand how an animal responds to stimuli. For example, the olfactory system in a fly helps it to recognize different odors in its world, everything from nutritious food to the presence of potential mates. But the fly's brain doesn't stop at just recognizing odors; it also helps drive the fly's behavior in response to those odors. A fly might move toward healthy food but away from food that is less nutritious – responses that Abbott, working with the lab groups of Gerry Rubin at Janelia and HHMI Investigator Richard Axel at Columbia University, has begun to identify in the olfactory circuit.

“You see this transformation from a representation that's dominated by the outside world – many odors come in and activate cells – to something that's internal to the animal,” Abbott says. “It depends on their experience, whether they're hungry or not, those kinds of things.”

The task now, he says, is to work with other scientists to design experiments to identify those experiences and internal factors that drive choices. All the theoretical and computational neuroscientists at Janelia agree that their work is iterative.

“What I hope will happen at Janelia is that there will really be this close shoulder-to-shoulder interaction with experimental labs: the development of models based on experimental observations, which in turn generates predictions to be tested with new experiments,” Romani says.

This partnership among theory, computation, and experiment in neuroscience gives researchers an enhanced ability to devise effective ways to peer into the mysterious depths of the brain. As better experiments yield a clearer understanding of neural processes, these scientists are helping researchers to navigate the turbulent sea of neural activity and row toward the truth. ■



Sandro Romani studies the ability of the brain's hippocampal cells to store sequences of places and words.