A conversation with Professor E.J. Chichilnisky, January 23, 2020

Participants

- Prof. E.J. Chichilnisky - John R. Adler Professor of Neurosurgery and Professor of Ophthalmology at Stanford University
- Joseph Carlsmith - Research Analyst, Open Philanthropy

Note: These notes were compiled by Open Philanthropy and give an overview of major points made by Prof. Chichilnisky.

Summary

Open Philanthropy spoke with Prof. E.J. Chichilnisky of Stanford University as part of its investigation of what we can learn from the brain about the computational power ("compute") sufficient to match human-level task performance. The conversation focused on the compute necessary to replicate the information-processing performed in the retina.

Prof. Chichilnisky's work

The Chichilnisky Lab at Stanford is working on developing an artificial retina. They've focused thus far on an experimental lab prototype using ex vivo stimulation and recording, and they're building an in vivo prototype in 2020. Their long-run goal is to help restore impaired vision, increase neuroscientific understanding, and augment human visual capabilities.

Necessary modeling detail

Discussion of the compute sufficient to replicate the brain's information-processing is very speculative. We don't know enough about the brain to give answers with confidence, and different people with neuroscientific expertise will answer differently.

Neuron models typically discussed in neuroscience lie on a spectrum of increasing complexity. There are very simple models (for example, generalized linear models and linear-non-linear models) on one end, and detailed biophysical models on the other. Prof. Chichilnisky believes that the best models are probably somewhere in the middle -- e.g., each cell is doing something more complicated than a linear-non-linear model, but we don't need to understand every biophysical detail to understand and replicate the computation.

For this reason, we need models that more effectively balance the tradeoff of explanatory power and complexity.

Biophysical modeling

Prof. Chichilnisky does not believe that detailed biophysical modeling (for example, modeling every ion channel) is necessary for replicating the brain's information-processing. He also does not know anyone in the field who thinks that such modeling is strictly required, though there are people who think that focusing on biophysics is the most productive way to make neuroscientific progress. Even if ion channels turn out to be an interesting type of function approximator, that wouldn't mean that there isn't another way to perform the same computation.

However, no one has been able to prove one way or another whether detailed biophysical modeling is necessary. It's hard to know, and there isn't a lot of evidence. There are high-quality experimental and computational efforts underway to understand this.

One example of a context in which simplification is possible comes from the work of Prof. Eve Marder of Brandeis University on the crab stomatogastric ganglion, a network of neurons in the crab stomach. Prof. Marder showed that the network can exhibit the same behavior with different ion channels and different gene expression. In this case at least, you just need to replicate the high-level computation, and since artificial neural networks are universal function approximators, they may be able to do this.

One point in favor of focusing on biophysics, though, is that it centers attention on the biophysical components that actually produce a neuron's input-output function, as opposed to the types of models that appeal to humans on pen and paper. This might ultimately make summarizing what a neuron does much more efficient. As an analogy: you can approximate a non-polynomial mathematical function like cos(x) as a polynomial with a very large number of terms, but it would be much simpler to model it using the actual function.

People's views about the right level of biophysical detail to focus on are sometimes shaped by what they're good at (e.g., computational simplifications, vs. detailed biophysical analysis). And some people find just biophysical complexity intrinsically interesting.

Challenges in replicating retinal computation

We are not that close to correctly reproducing the neural code of the retina. We're somewhat close, and we might be able to see the end of the road, but we cannot replicate every single spike in every single condition. Challenges include:

- It's hard to know when to stop fine-tuning the details of your model. A given model may be inaccurate to some extent, but we don't know whether a given inaccuracy matters, or whether a human wouldn't be able to tell the difference (though focusing on creating usable retinal prostheses can help with this).
- There are many different types of cells in the retina. There are about 20 different types of retinal ganglion cells (the output cells of the retina) in humans, and in mice there are maybe 40. They all do very different things, and some might be simpler than others.
- In addition to capturing the behavior of an individual cell, you need to capture correlations between cells in the retina.
- One of the biggest challenges is the world of possible stimuli. It would take lifetimes to present all possible stimuli, so we don't know if we're missing something. Prof. Chichilnsky's lab has the biggest trove of data in the world from retinal ganglion cells. They've recorded from something like 500,000 retinal ganglion cells (roughly half the retina), and they have about 50 billion spikes. But even this may not be enough data.
- There is variability in retinal function both across species and between individuals of the same species. Mouse retinas are very different from human retinas (a difference that is often ignored), and there is variability amongst monkey retinas as well; Chichilnisky's lab has recently made important progress on understanding this variability. Individual human retinas may use different neural codes as well.
- Most of these points very likely apply to the brain as well.

In the context of retinal prostheses to be implanted in your eye or worn on your head, the large compute burdens of artificial neural networks also cause practical challenges, since these require complex hardware that consumes a lot of power.

Speculating, Prof. Chichilnisky believes that it is possible to build an adequate computational model of the retina, though we may need to rely on a neural network model we don't understand until we can build a model that better captures what the retina is doing.

Generalizing from the retina to the brain as a whole

Prof. Chichilnisky is a retina expert, and so is biased towards the view that studying the retina is generally very valuable. He thinks that the history of science shows that what we learn about the retina now, we discover in the brain 20-40 years later. For example, the importance of distinct computations in different cell types became clear to the retina

community a couple of decades ago. A similar realization of the profound and critical differences between cell types in the cortex has been emerging much more recently and is now widely viewed as a big deal, even though it was largely ignored 1-2 decades ago. The retina is usually better understood than the brain for technical reasons, and it provides a very reasonable null hypothesis for what we need to understand in the brain in the future.

The level of modeling detail necessary in the retina provides a good test of the level of modeling detail necessary in the brain as a whole. However, the data on the retina aren't in, and they won't be in for a while.

The brain is probably a lot more plastic than the retina, though this is likely a quantitative rather than a qualitative difference. Indeed, Prof. Chichilnisky's approach to building retinal prostheses assumes less plasticity in the retina and brain than other approaches do. Plasticity is obviously important. However, building plasticity into a circuit (biological or electronic) is generally expensive and creates greater risk of malfunction. Thus, we should assume that not every neural circuit is highly plastic. Instead, we should assume that the brain is plastic in those instances where plasticity confers an evolutionary advantage.

We are much further along in mapping all of the cell types in the retina than we are in the brain as a whole. Differences between cell types matter a lot in the retina. We don't know how much these differences matter in the rest of the brain. Some people think that they don't matter very much, but Prof. Chichilnisky disagrees, and certainly the field has been moving in the direction of emphasizing the cell type differences in the brain.

However, there's no reason to think that some neuron types in the brain/retina will be radically simple and some will be radically complicated. There will be some variations, but perhaps not a big gulf.

AI and neuroscience

Prof. Chichilnisky thinks we should be careful in drawing inferences about neuroscience from AI systems. We don't really understand how artificial neural networks work. We can tweak different knobs and get them to do what we want, but the internals are still opaque.

We see some hints of similarity between artificial neural network models of the retina, and the internal biological behavior of the retina (for example, in recent work by Prof. Stephen Baccus of Stanford and his collaborators). But there are also strong warning signs that such networks work in a manner different from human vision, as revealed by so-called adversarial approaches. For example, by tweaking the pixel values of an image in a manner

indistinguishable to humans, you can cause an artificial neural network to classify a school bus as an ostrich.

Relevance of hardware vs. software

The hard part, in replicating what the brain does, is not the computational hardware, but rather figuring out how to organize software to take in and process information in the right way.

All Open Philanthropy conversations are available at
http://www.openphilanthropy.org/research/conversations