Discussions with Dr. Stephen Larson, 2019-2020

Participants

- Dr. Stephen Larson - CEO of MetaCell and Co-founder of OpenWorm
- Joseph Carlsmith - Research Analyst, Open Philanthropy

Note: These notes were compiled by Open Philanthropy and give an overview of the major points made by Dr. Larson. Some of these points were made in conversation, and some via electronic communication, at various points in 2019 and 2020. Dr. Larson also read a draft of Mr. Carlsmith's report on computation and the brain.

Summary

Open Philanthropy reached out to Dr. Stephen Larson of MetaCell and OpenWorm as part of its investigation of the computational power ("compute") sufficient to match the human-brain's task-performance. The discussions focused on the compute required to model different biophysical processes in nervous systems.

C. elegans

Dr. Larson co-founded OpenWorm, a project that aims to simulate *C. elegans*, a nematode worm with 302 neurons and ~1000 cells. Despite its small size, we do not yet have a model that captures even 50% of the biological behavior of the *C. elegans* nervous system. This is partly because we're just getting to the point of being able to measure what the worm's nervous system is doing well enough.

It is possible to replicate certain kinds of worm behaviors, such as a crawling forward motion, using a very simple neural network. However, the same model cannot be used to make the worm shift into crawling backwards. Rather, you have to re-train it, and even then, you don't know if the model makes the decision to crawl backward with the same frequency, and for the same reasons, that the real worm does. In general, evolution has equipped the worm to respond to a very wide range of conditions, and the worm's biology has all of these intricate and complex mechanisms that could potentially be involved in the behaviors you care about.

Note that the simulation standards here aren't simply about reproducing black-box behavior. OpenWorm also aspires to be able to identify mappings between what's going on inside the simulated worm and measurements you can make of the real nervous system.

On the basis of his experience at OpenWorm thus far, Dr. Larson thinks it unlikely that very simplified neuron models (e.g., integrate-and-fire neurons, or models akin to the artificial neurons used in deep neural networks) are going to be sufficient to describe the information-processing dynamics involved in the worm's behavior. OpenWorm's approach is to throw as much complexity into the neuron models as they think is necessary (this is currently roughly at the level of a Hodgkin-Huxley model, plus some additional features), in an effort to really nail down that their model is capturing the worm's behavior across many conditions and timescales.

Success in such a project would allow you to bound the complexity necessary for such a simulation (indeed, this is one of Dr. Larson's motivations for working on it). After that, you could attempt to simplify the model in a principled way. However, the jury is still out on how much simplification is available, and Dr. Larson thinks that in this kind of uncertain context, you should focus on the worst-case, most conservative compute estimates as your default. This means worrying about all of the information-processing present in cell biology. In general, in studying complex biological mechanisms, Dr. Larson thinks that the burden of proof is on those who want to say that a given type of simplification is possible.

Mr. Carlsmith asked Dr. Larson whether he thought it relevant to the compute capacity of the *C. Elegans* nervous system that fairly small artificial neural networks can be trained to engage in motor-control behaviors like running through environments or wrestling. Dr. Larson was interested in this idea, and he thought it possible that, at the end of the day, the 302 neurons in the *C. elegans* aren't doing some very large amount of behaviorally relevant computation. However, the algorithms they're using, and the unique computational properties of biological systems, are still worth understanding.

Complexity of neuron computation

Gene expression in cells can be viewed as a type of computation, and cells may be capable of storing memory in the combinations of proteins in the cell. These may be slower activities, but they could be complex regardless (for example, these proteins could interact with receptors on the membrane, or they could up-regulate or down-regulate synapses).

It's also easy to dismiss the relevance of factors like the temperature of the brain, or the pH of the cells, but these can have effects on neural mechanisms and on cognitive processing

more broadly, and Dr. Larson has not been able to show (despite some effort) that they do not need to be accounted for in functional models.

In general, anything that influences a neuron's membrane potential is a candidate for computational relevance. In highly non-linear systems like the brain, small differences can have large effects, and these can cascade from affecting one neuron, to affecting large groups of neurons.

Results in the retina

Dr. Larson is resistant to generalizing from modeling results in the retina to the rest of the brain. Other parts of the brain have different cell types, circuits, and functions, and hence may process information in more complex ways. Given this heterogeneity in neural tissue, it's safer to assume that computation in the brain is more complex than in the retina than to assume that the two are comparable.

Alternative signaling mechanisms

Neuromodulation

Prof. Eve Marder at Brandeis University has done a lot of work on neuromodulation, which can shift the entire input-output relationship of a neural circuit. This could be analogous to changing the program you're running on your computer. Alternatively, though, it could involve a shift in the circuit's computational capacity.

Gap junctions

Dr. Larson thinks that gap junctions can contribute to non-linear dynamics and near-chaotic dynamics within neural networks. As a rough rule of thumb: the more non-linear a system is, the more computationally expensive it is to simulate.

Non-standard axon signaling

In *C. elegans*, almost all neurons do not spike, but important signaling takes places at chemical synapses regardless. This suggests that information-processing can occur in a purely subthreshold regime, without action potentials.

Blood flow

Cells in the blood-brain barrier regulate additional blood flow to areas that are particularly active (though areas with a lot of inhibitory cells can be active even when they are reducing activity elsewhere). It's generally thought that blood flow is more of an epiphenomenon/a sign that other forms of information processing are occurring (akin to the heat generated by a CPU), than a mechanism of information-processing in itself.

Landauer's limit and the brain

Dr. Larson is not persuaded that Landauer's limit can be used to upper-bound the FLOP/s necessary to replicate the brain's task-performance, as it seems possible to him that there could be computational processes occurring in the brain that do not require bit-erasures.

Uncertainty in neuroscience

Dr. Larson does not think that there is strong evidence that spikes and synaptic inputs are the most informative processes for studying information-processing in the brain. And it's very possible that a streetlight effect is at work, in which scientists give undo theoretical prominence to processes they are better able to measure.

In general, Dr. Larson thinks that we should be careful about inferences from "X assumption is common in neuroscience" to "X assumption is true of the brain's biology." Such inferences could easily lead to overconfidence about the reliability of certain simplifications.

Similarly, there may be selection bias at work in appeals to the success of simple models in some contexts as evidence for their adequacy in general. With respect to phenomena that simple models have thus far failed to explain, such explanation might not be possible.

In studying biological systems, it is very difficult to identify the causal link between one variable and another, because attempts to manipulate one variable can be confounded by compensatory mechanisms in the system (e.g., X gene is in fact linked to Y function, but if you get rid of X gene, other genes will compensate for it). As a result, biologists have come to expect that valid results will control for a huge number of possible interfering variables -- a standard that Dr. Larson thinks that some of the more computationally-oriented parts of neuroscience tend to employ with less rigor.

Dr. Larson believes that what we mean by "encode information" and "process information" is, at the most general level, an unsolved problem in neuroscience. Answers will vary widely depending on the expert you ask.

FLOP/s

Focusing on FLOP/s may be less informative than focusing on other hardware bottlenecks, such as memory bandwidth. The Blue Brain project in Switzerland, for example, believed that they needed a supercomputer in order to reduce latency between CPUs as much as possible, given the highly interconnected nature of the neurons they aimed to simulate. Dr. Larson thinks that going "beyond the FLOP" may be a helpful research direction, and he is intrigued by the possible use of TEPS as an alternative.

Overall uncertainty/agnosticism

Given the many uncertainties involved in estimates of this kind, Dr. Larson believes that the right conclusion is something like: there is insufficient evidence to justify concluding anything (as opposed to, e.g., "there is some moderate evidence in favor of X FLOP/s, so maybe let's believe that?"). In statistics, for example, one wants a P value less than .05, and Dr. Larson is not sure we have anything like that for these FLOP/s estimates.