A conversation with Professor Eric Jonas, September 17, 2019

Participants

- Professor Eric Jonas - Assistant Professor of Computer Science at the University of Chicago
- Joseph Carlsmith - Research Analyst, Open Philanthropy

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

Summary

Open Philanthropy spoke with Prof. Eric Jonas of the University of Chicago 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 state of relevant neuroscientific knowledge.

High-level sources of uncertainty

Attempting to estimate the compute sufficient to replicate the brain's task performance is an extremely challenging project. It's worthwhile (indeed, it's a common thought experiment amongst neuroscientists), but the error bars will be huge (e.g., something like ten orders of magnitude).

Task-definition

One level of uncertainty comes from the difficulty of defining the high-level task that neural systems are trying to perform (e.g., the "computational level" in the hierarchy proposed by David Marr). Our attempts to capture cognitive tasks with objective functions we can fit machine learning models to are all extreme simplifications. For example, Prof. Jonas is fairly confident that the visual system is not classifying objects into one of *k* categories.

Which systems to focus on

We also aren't sure what boundary to be drawing around neural systems. For example, active dendritic dynamics mean that using the neuron membrane as the boundary between computational units is somewhat inappropriate. Instead of asking how much computation

a neuron could be doing, maybe we should be asking how much computation a lower-level component like an NMDA receptor could be doing.

Alternatively, given that we know that there is some functional localization and modularity in the brain, maybe we should be asking how much computation a larger area, like the lateral geniculate nuclear, could be doing. That said, a lot of the interesting things happen in the prefrontal cortex, which is less specialized and so less amenable to this treatment.

It's also possible that the computational capacity of a chunk of cortex is several orders of magnitude higher than the computational complexity of the underlying task that it's performing.

Limitations of model animals

In an experiment with a model animal like a rat, which has a very complicated brain, the number of input/output bits we can control/observe is extremely small. This makes it very hard to do informative, high-throughput experiments. Even if you had a billion rats doing your experiment 24/7, you'd still only have a small number of bits going in and out.

Uncertainty about neuron computation

It's reasonable to ask what a synthetic analog of a neuron would need to be able to do in order to replicate the function of a real neuron in the brain.

However, many electrophysiologists would say that we don't know what neurons are doing. And they would ask: how can we start making claims about the computational capacity of networks of neurons, if we don't know how individual neurons work? Prof. Jonas is sympathetic to this.

There are a variety of complexities that make the computations performed by a neuron extremely difficult to quantify. Examples include: dendritic spiking, the complex dynamics present in synapses (including large numbers of non-linearities), the diversity of ion-channel receptors, post-translational modification, alternative splicing, and various receptor trafficking regimes.

Some people attempt to draw comparisons between neurons and transistors. However, even with a billion transistors, Prof. Jonas does not know how to create a reasonable simulation of a neuron.

Dendritic computation

Active dendritic computation could conceivably imply something like 1-5 orders of magnitude more compute than a simple linear summation model of a neuron. And if dendritic morphology is evolving over time, you also need to be thinking about the space of all possible dendrites that could have formed, in addition to the current dendritic tree.

That said, it's reasonable to think that at the end of the day, simplified dendritic models are available. For example, Prof. Jonas has heard arguments suggesting that post-synapse, there is very little plasticity in dendrites, and that dendritic computation mostly involves applying random features to inputs.

State retention

State retention is also a source of uncertainty. Prof. Jonas is not convinced by any arguments he's heard that attempt to limit the amount of state you can store in a neuron. Indeed, some recent work explores the possibility that some information is stored using DNA. If there are actually molecular-level storage mechanisms at work in these systems, that would alter compute estimates by multiple orders of magnitude.

Learning

Prof. Jonas thinks that estimating the complexity of learning in the brain involves even more uncertainty than estimates based on firing decisions in neurons.

Neuroscientists have been studying things like spike timing dependent plasticity and long-term plasticity for decades, and we can elicit versions of them reliably *in vitro*. But it's much harder to understand the actual biological processes occurring *in vivo* in a behaving animal, because we have so much less experimental access.

The machine learning community has multiple theories of the computational complexity of learning. However, these don't seem to capture the interesting properties of natural systems or existing machine learning systems.

Prof. Jonas does not think that there is a clear meaning to the claim that the brain is a deep learning system, and he is unconvinced by the argument that "the brain is doing optimization, and what is deep learning but optimization?". He also has a long-term prior that researchers are too quick to believe that the brain is doing whatever is currently popular in machine learning, and he doesn't think we've found the right paradigm yet.

Efficiency of biology

Compute estimates based on the brain are premised on the assumption that the brain's computation is relevant to the algorithms and tasks we care about. This raises an orthogonal set of questions. After all, biology is a mess, and its mechanisms are likely to be far from maximally efficient.

You can also argue that biological systems are so complicated because they need to be robust, and that fundamentally the units are very simple. But Prof. Jonas does not think so.

Various discoveries in biology have altered Prof. Jonas's sense of the complexity of what biological systems can be doing. Examples in this respect include non-coding RNA, the complexity present in the three-dimensional structure of the cell, histone regulatory frameworks, and complex binding events involving different chaperone proteins. The class of computation that Prof. Jonas can imagine a single cell doing now seems multiple orders of magnitude more complex than it did 20 years ago.

Upper bounds

Upper bounds could come from limits imposed by the physics of computation on the number of bits expressible in a given unit of matter. Obviously, though, this will be a massive overestimate: the brain is not maximally efficient in this sense.

Prof. Jonas's skepticism does not extend so far as to take seriously the possibility that quantum dynamics are relevant to the brain's computation. And he does not give credence to dualist theories of the mind.

History of over-optimism

There is a history of over-optimism about scientific progress in neuroscience and related fields. Prof. Jonas grew up in an era of hype about progress in science (e.g., "all of biology will yield its secrets in the next 20 years"), and has watched the envisioned future fail to arrive. Indeed, many problems have been multiple orders of magnitude more complicated than expected, to such a degree that some people are now arguing that science is slowing down, and must rely increasingly on breadth-first search through possible research paths.

In biology, for example, there was a lot of faith that the human genome project would lead to more completeness and understanding than it did. And many in the neuroscience community feel that some neuroscientists made overly aggressive claims in the past about what amount of progress in neuroscience to expect (for example, from simulating networks of neurons at a particular level of resolution).

It's also fairly rare for neuroscientific discoveries to lead to mechanistic understanding sufficient for direct clinical impact. A lot of drugs are discovered serendipitously, and a lot of our animal models are wrong in clinically-relevant ways.

Comparisons with deep neural network vision models

Deep neural network vision models are vulnerable to adversarial examples, and Prof. Jonas was involved in a project that showed that it is relatively easy to find things that neural nets misclassify in naturally occurring data. These problems suggest that these systems aren't learning what we think they're learning, that we haven't found the right task metric, and that we haven't yet captured the aspects of vision we should be caring about.

One can also argue that for the more interesting ImageNet tasks (e.g., combined detection and classification tasks), performance has plateaued over the past five years.

We can also get very good performance on object-classification tasks using support vector machines or kernel methods (you can even get pretty far using k-nearest-neighbors). The performance is not as good as neural networks, but conceptually, the underlying computation these systems are performing is very simple, and much simpler than the complicated cascade of linearities, non-linearities, filters, etc performed by a deep neural network. This makes it hard to think about the underlying complexity of the task.

In general, Prof. Jonas is skeptical of the hype surrounding deep learning, though he does not ignore the progress that has been made. As an interesting class of function-approximation technologies for working with real-world data, deep neural networks appear to be useful, and stochastic gradient descent does a shockingly good job of training them. But in the context of the brain, once you get outside of early stage perceptual tasks, their value is less clear, and Prof. Jonas is very skeptical about deep reinforcement learning as a paradigm. In general, we still don't understand the underlying computational limits of deep learning systems, or the complexity of the tasks they are solving.

FLOP/s

FLOP/s is an imperfect metric for thinking about computational power. A FLOP is a very specific type of operation, and not everyone agrees on how to define it. Generally, people have in mind a multiply-accumulate operation, using some sort of binary code, usually some version of IEEE floating point. This is fine in many contexts, but we also have an actual theory of what it means to compute, which doesn't depend so much on the contingencies of computer architectures.

You can also do floating point arithmetic using a giant-look-up table. This will take up a lot of silicon, but the underlying computational complexity is still the same. So it's worth thinking carefully about what it means to do this type of computation.

That said, for the purposes of using the brain to estimate the compute sufficient to match human-level task performance, and at the level of precision those estimates currently admit, FLOP/s is a fine metric. The error bars will be dominated by other factors.

Disagreement amongst experts

Prof. Jonas wishes there were ways of encouraging people to make bets about these sorts of issues, and to quantify their uncertainty, but it's hard to define the questions (e.g., "when will we simulate a neuron?") precisely. It might be helpful to rely more heavily on opinions expressed by people whose critical research pathway depends on them getting this question right (Prof. Jonas is not such a person).

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