A conversation with Professor Rosa Cao, August 7, 2019

Participants

- Professor Rosa Cao - Assistant Professor of Philosophy, Wu Tsai Neurosciences Institute, Stanford University
- 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. Cao.

Summary

Open Philanthropy spoke with Prof. Rosa Cao 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 computational models of biophysical processes in the brain.

Computation in the brain

Prof. Cao does not believe that there is a privileged description of *the* computations that the brain is performing. We can imagine many different possible computational models of the brain, which will replicate different types of behavior, to within a given error-tolerance, in a given circumstance. In order to determine which biophysical processes are important, and what level of precision and detail you need in a model, you first need to specify the particular type of input-output relationship that you care about, and how the relevant outputs need to be produced.

More generally, Prof. Cao thinks that the computational paradigm in neuroscience is conceptually underspecified. That is, the field is insufficiently clear about what it means to talk about the computations that the brain is performing.

Relevance of biophysical details

Digital computers achieve speed and reliability by ignoring many dimensions of what is happening in the system. In such a context, you only care about whether the voltage in the transistors is above or below a certain threshold, and designers try hard to shield this variable from disruptive physical fluctuations.

The brain is built on fairly different principles. Its functional processes are not shielded from the dynamics of the brain's biochemistry. Rather, the brain exploits this biochemistry to perform efficient computation.

This makes the brain difficult to simulate. In nature, biochemical processes like protein-protein interactions just happen, so they are "free" for the brain to run. Simulating them, however, can be quite computationally expensive.

Prof. Cao thinks that if we could be handed a model of the brain that only included the dynamics that matter to performing some set of cognitive tasks, it might not be very computationally expensive to implement. However, without more details about the specific case or a principled procedure for identifying the dynamics that really matter, this thought experiment is hard to have intuitions about.

The difficulty of computational modeling in biology

E. Coli, a comparatively simple, one-celled organism, exhibits fairly sophisticated behavior on the basis of carefully-tuned biochemical chains (for example, various rhythms at different timescales that allow the cell to survive in a range of environments). We have not yet been successfully able to capture this behavior in a computational model, despite throwing a lot of effort and computational power at the project.

Indeed, there was a lot of excitement about projects like this a few decades ago, but it seems to Prof. Cao that this energy has since died down, partly due to greater appreciation of their difficulty.

Similarly, efforts to build an artificial cell have proven very difficult. At some level, cells are simple, and we basically know what the components are. However, all of the biochemical processes are poised in a delicate balance with each other -- a balance that represents a vanishingly smaller percentage of all possible arrangements, and which is correspondingly difficult to replicate.

Efforts to create functional brain simulations might run into similar problems. For example, it may be that the brain's function depends on a particular type of relationship to the environment, which allows it to adjust and fine-tune its internal features in the right way.