A conversation with Dr. Derek Shiller, January 24, 2017

Participants

- Dr. Derek Shiller – Software Engineer, DailyPay
- Luke Muehlhauser – Research Analyst, Open Philanthropy Project

Note: These notes were compiled by the Open Philanthropy Project and give an overview of the major points made by Dr. Derek Shiller.

Summary

The Open Philanthropy Project spoke with Dr. Shiller as part of its investigation into which types of beings should be of moral concern, and thus a potential target for the Open Philanthropy Project's grantmaking. This conversation focused on one particular factor plausibly relevant to whether a being should be of moral concern or not — namely, whether that being is phenomenally conscious, and what the character of its conscious experience is. Conversation topics included the possibility of "hidden" qualia, the implications of hidden qualia for the distribution of phenomenal consciousness, and the possibility of multiple conscious subjects being implemented in the human brain.

Hidden qualia and the distribution of consciousness

Dr. Shiller does not think that accepting the possibility of hidden qualia — i.e. non-introspectable qualia — implies much about whether, e.g., rats, chickens, or fish have qualia. However, if one thinks that hidden qualia are incoherent or impossible, and that introspection is necessary for qualia to exist, this might imply a very limited distribution of consciousness, since the ability to introspect on experiences seems quite sophisticated. For instance, one could argue that, from an evolutionary perspective, chickens would be unlikely to benefit from the capacity to introspect their mental states.

If hidden qualia do not exist, this would also imply that at least some kinds of highly sophisticated information processing (e.g., as occur in non-introspectable parts of the human brain) are not sufficient to produce qualia.

Potential scenario that would prohibit hidden qualia

One scenario in which hidden qualia might turn out to be impossible is if the instantiation of qualia requires a certain type of self-model, and the introspection process connects particular perceptual processes which are the focus of attention to an internal self-model, and this is what causes a stream of information processing to have "something it is like" to be it.

Possibility of the brain implementing multiple conscious subjects

Some have proposed that the human brain may implement multiple conscious subjects, rather than just the central narrative "self" of introspection (e.g., perhaps the cerebellum is a conscious subject). Any qualia associated with those centers of consciousness would therefore be unavailable to ordinary introspection.

Dr. Shiller has proposed the stronger claim that some qualia, rather than being "hidden" in the sense of belonging to a separate conscious subject in the brain, may in fact contribute to our phenomenal experience without us being able to accurately introspect about them. The experimentally-demonstrated fact that color data is not sampled along the edges of human peripheral vision, despite our intuitive belief that our whole visual field is in color, is somewhat analogous to what Dr. Shiller means by qualia being "hidden" in this way.

Evidence from split-brain cases

Dr. Shiller thinks the best interpretation of the evidence from split-brain cases is that human brains contain (at least) two separate conscious subjects, but that we are usually unaware of this. Furthermore, while the corpus callosum is a particularly easy bundle of fibers to sever, there might in principle be many ways of severing neural connections that would reveal a variety of separate, internal conscious subjects.

It seems possible to imagine a large cognitive system (or sufficiently complex computer program) that happens to instantiate phenomenal consciousness in multiple "places," not all of which are linked to any visible outputs, and which produces the behavioral appearance of a single agent, despite in fact containing multiple sub-agents. Something like this may be observed in split-brain patients, in which the brain hemispheres seem to continue running in synchrony until they receive different information from each other.

Other people to talk to

- Eric Schwitzgebel (University of California, Riverside)
- Peter Carruthers (University of Maryland)
- Chris Eliasmith (University of Waterloo)