A conversation with Professor Keith Frankish, January 24, 2017

Participants

- Professor Keith Frankish – Visiting Senior Research Fellow, The Open University
- 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 Professor Keith Frankish.

Summary

The Open Philanthropy Project spoke with Professor Frankish of The Open University 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 focused on "illusionism" about consciousness and its implications, as well as the more general question of how to think about which systems might be conscious in a way that warrants moral concern.

Two theses of illusionism

Illusionism can be seen as involving two theses:

- The *negative* thesis that there is no "intrinsic subjectivity" – that is, that phenomenal properties do not exist in a substantive sense. Instead, illusionism views consciousness as a fundamentally psychological phenomenon (what might be termed "introspective subjectivity"). This rules out non-physicalist theories of consciousness, such as panpsychism, and it entails that consciousness is restricted to creatures with suitable psychological states, functionally defined. It is, however, compatible with a wide range of theories as to the nature of those states and therefore with a variety of views about the distribution of consciousness.
- Some *positive* thesis as to what, exactly, constitutes the "illusion of consciousness" – that is, why we are inclined to *feel as if* there is intrinsic subjectivity. If illusionism is correct, questions about the complexity and likely distribution of consciousness will depend almost entirely on the details of this positive thesis. Professor Frankish suspects that human consciousness involves a complex, multi-faceted, introspective illusion, which depends on a variety of sensory, affective, evaluative, cognitive, and cultural components, some of which are evolutionarily ancient and some distinctively human. There are probably different kinds of consciousness, as well as different degrees of it.

Avoiding a strong "consciousness" concept

Instead of asking whether a system fits some concept of "consciousness," it might be helpful to consider more granular features of the system, as informed by science. For instance, rather than asking broadly whether fish "feel pain," it might be more useful to ask what abilities fish have, what preferences they display, etc., and what moral status we should grant them on that basis.

Having a better model of the complex suite of abilities, dispositions, reactions, etc. that contribute to humans' moral status might allow us to make more confident judgments about how much moral concern to extend to other, simpler cognitive systems.

Moral importance of preferences about internal states

For a system to have moral weight, Professor Frankish thinks that the system must have preferences with respect to its own internal states (rather than simply preferences about external states of affairs). Professor Frankish does not think a chess-playing computer, for example, has moral weight: while it can be viewed as exhibiting a "preference" for winning a match (an external state of affairs), it does not have preferences about its own internal states.

Usefulness of example programs

Luke has suggested it might be helpful to write example computer programs to try to clarify which features of a system we morally care about. Professor Frankish thinks that the fine-grained, low-level features of such programs are unlikely to affect our moral intuitions about the program much; instead, he thinks our moral intuitions are likely more sensitive to the program's behavior (external or internal). For instance, one salient feature of human consciousness is our internal verbal monologue; *whether* a program implements something analogous may seem morally relevant, but the lower-level details of *how* the program implements it may not.

Examining moral intuitions by hypothetically adding or removing features of a system

One potential way to get a sense of one's intuitions about the moral status of simpler cognitive systems is to assume the moral status of a human-level cognitive system as given, then hypothetically remove features until the system no longer seems to have moral weight (a “top-down” approach). Alternately, one could start by considering a very simple system that one doesn’t assign moral weight to and add features until it does seem morally relevant (a “bottom-up” approach). Both methods may be useful, though the latter promises to offer a more general, less anthropocentric perspective.

Other people to talk to

- Nick Humphrey (London School of Economics)
- Daniel Dennett (Tufts University)

- Peter Carruthers (University of Maryland)

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