A conversation with Professor David Chalmers, May 20, 2016

Participants

- Professor David Chalmers – Professor of Philosophy, New York University (NYU)
- Luke Muehlhauser – Research Analyst, Open Philanthropy Project
- Nick Beckstead – Program Officer, Scientific Research, Open Philanthropy Project

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

Summary

The Open Philanthropy Project spoke with Professor Chalmers of NYU 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 possible approaches to the question of distribution and character of phenomenal consciousness, and ideas for stimulating more philosophical work in this area.

Phenomenal consciousness

(Some context for the reader.)

In what follows, when we say "consciousness" alone, we always mean phenomenal consciousness. Philosophers often say that a creature is phenomenally conscious when there is "something it is like to be" that system. If there is something it is like to be a bat, for example, then a bat is phenomenally conscious. States of phenomenal consciousness include the experience of feeling pain, the experience of seeing red, and the experience of tasting wine. Phenomenal consciousness is often distinguished from self-consciousness (consciousness of oneself) and reflective consciousness (consciousness of one's own thoughts). While it is arguable that self-consciousness and reflective consciousness are limited to humans and perhaps a few other species, it is much less obvious that phenomenal consciousness is so limited in its distribution.

The question of phenomenal consciousness is central for addressing the question of whether a certain taxon (say, trout) can feel pain, as opposed to merely exhibiting (unconscious) nociceptive processing that, to many people, has little or no moral significance. Many philosophers hold that some degree of phenomenal consciousness is required for an entity to have any moral standing.

At the same time, phenomenal consciousness is perhaps the least-understood aspect

of the mind. There is no current consensus theory of phenomenal consciousness, and we cannot yet explain how phenomenal consciousness is instantiated by the brain (in humans), and how it might be instantiated by other substrates in other systems. However, it is also widely agreed that certain kinds of neurobiological and behavioral processes are highly correlated with phenomenal consciousness, and thus it seems possible to approach the question of phenomenal consciousness in animals even without a complete explanation of consciousness, by studying physical processes that are widely thought to correlate with phenomenal consciousness.

Approaches to the distribution of consciousness

Potential approaches to the question of distribution and character of consciousness across cognitive systems (e.g. species of animals or types of software) include:

- Focusing first on developing a robust theory of consciousness, then applying it to the distribution question. (This would likely be a very long-term project.)
- Determining a reasonable credence distribution over leading theories of consciousness (including substantial credence that none are correct), then analyzing what predictions these theories make about the distribution of consciousness. For example, if this credence distribution over theories had substantially more credence on first-order theories than on higher-order theories, this would suggest a higher probability that animals with relatively simple nervous systems could be conscious than would be the case if more weight were given to higher-order theories.
- Generating candidates for necessary and/or sufficient conditions for consciousness, producing arguments for and against these candidates, and applying these conditions to the distribution of consciousness. It could also be useful to challenge proponents of particular necessary and/or sufficient conditions to defend them.

Conditions for consciousness

In the scientific and philosophical literature on animal consciousness, by far the most common strategy for arguing that certain non-human animals are or are not conscious is the third: argue that a certain condition is necessary for consciousness and argue that certain animals do not satisfy the condition and so are not conscious; or argue that a certain condition is sufficient for consciousness, and argue that certain animals satisfy the condition and so are conscious. The most questionable part of these arguments is typically the claim that a certain condition is necessary or sufficient for consciousness.

Professor Chalmers thinks it is particularly difficult to make strong arguments in favor of necessary conditions for consciousness. Some features seem very unlikely to be necessary for consciousness, while for other conditions (for example, that the

system possess at least some specific level of complexity) there are not yet conclusive grounds for considering them necessary or unnecessary.

Sufficient conditions may be somewhat easier to argue for. For example, certain sorts of whole functioning human brains are plausibly sufficient for consciousness. This is a very strong sufficient condition, but one can then weaken it by identifying and setting aside features that seem very unlikely to be necessary.

As another example, it would be relatively easy to argue that certain biological conditions are sufficient for consciousness (because they appear to enable self-reported consciousness in humans), even though those biological conditions may not be necessary (if, for example, there are underlying functional conditions that would be sufficient to produce consciousness regardless of whether or not they are implemented in a biological substrate).

It may feasible to conclude that some systems are very likely to be conscious (e.g. based on a comparison of features with systems known to be conscious), even if the theory used does not provide strong grounds for ruling out other systems as not conscious. For instance, observing that a system has abilities like perception**,** planning, decision-making, language, reflection, creativity, etc. may, in practice, be sufficient to conclude the system plausibly has phenomenal consciousness.

Some of these conditions seem quite unlikely to be necessary: there is not much reason to think that language and creativity are required for consciousness, for example (deficits of language and creativity seem to have little effect on e.g. visual consciousness). So by dropping these one can then produce weaker sufficient conditions for consciousness. One might try to drop the other conditions too (plausibly there might be at least perceptual consciousness or pain without planning or reflection and maybe without decision-making), but then the remaining sufficient conditions become so weak that they may be satisfied by very simple organisms, and some will question their sufficiency.

In practice the least controversial necessary condition for consciousness is something like information processing, but this is so weak as to be useless without some specification of the sort of information processing required. Somewhat stronger candidates for necessary conditions that are relatively uncontroversial (at least compared to those above) involve capacities such as perception, information integration, and flexible use of information in the guidance of behavior, but these are all at least somewhat questionable (couldn't an extreme version of Helen Keller without any perception still be conscious? Couldn't an inflexibly behaving system still be conscious?) while also being very weak (arguably satisfiable even by some unicellular organisms). Many will think there should be stronger necessary conditions, but it is unclear what these are or how we can establish that they are necessary.

When considering the literature on consciousness, Professor Chalmers notes that

many theories of consciousness are intended as descriptions of the features that correlate with consciousness in humans, rather than specifications of necessary or sufficient conditions for consciousness in general. For example, one popular theory is that information in the brain is conscious if and only if it is part of a global workspace; information outside the global workspace is unconscious. But it would be a big leap to conclude from this that any system with a global workspace is conscious and that systems that lack a global workspace are not.

Because it is somewhat easier to make a case for sufficient conditions for consciousness than for necessary conditions, Professor Chalmers thinks that positive arguments that certain animals are conscious will, in the near term (and possibly in the long term), be on somewhat more solid ground than negative arguments that certain animals are not conscious.

For example, one can make a reasonable case that apes, say, are conscious on the grounds that the primary features that are possessed by humans but not apes (e.g. language, certain sorts of intelligence) are not plausibly necessary conditions for consciousness. This sort of reasoning might be extendible to argue that dogs or mice or fish or even insects are conscious. Caution is required, as the further we move down the scale of complexity, the more chance there is that we are bypassing some necessary condition. But doings things probabilistically: insofar as one can make a case for some significant credence that certain conditions are not necessary, one then has a case for some significant credence that animals that lack them (and are otherwise relevantly similar to humans) are conscious.

Satisfying conditions with a computer program

Some first-order theories of consciousness (e.g. Tononi's integrated information theory) might be satisfied by fairly short computer programs, but some higher-order theories of consciousness might include conditions that are difficult to unambiguously satisfy with a computer program, since these theories are often specified in terms of features (e.g. "thoughts") that we are not sure how to instantiate with code.

Degrees of consciousness

There are many potential ways to parameterize consciousness, and it is not straightforward to designate a single parameter (e.g. "intensity" or "amount of information") as capturing the overall "degree" of consciousness.

It is plausible that neuroscience research on the brain states and features that correlate with different intensities of pain in humans could be taken as indicating that analogous brain states in at least some other animals likely correspond to a similar phenomenal experience of pain. In practice, behavioral measures are also often used to draw conclusions about relative intensities of experiences like pain.

Potential to increase research in this area

The question of distribution of consciousness may receive less attention than would be optimal because many researchers see it as particularly intractable. Philosophers in particular are often most motivated to engage when there are strong arguments to analyze and debate, and it is difficult to come up with strong arguments about the necessary conditions for consciousness or the likely distribution of consciousness across different types of cognitive systems. Scientists are motivated to engage where there are experimental methods that one can use to help settle these questions, and it is hard to use experimental methods to settle questions about animal consciousness, as consciousness is not directly measurable. To draw conclusions about consciousness from empirical observations, one needs background principles about what conditions are necessary or sufficient for consciousness, and it can be very difficult to subject these background principles themselves to experimental test.

Potential strategies for incentivizing work

Offering a prize for the best paper of the year on animal consciousness, perhaps with an associated conference for presenting papers, might incentivize work on the topic.

Getting additional prominent philosophers with a track record of generating interesting arguments to work on the question of distribution of consciousness could lead to significant progress, as well potentially influence others to work on the topic. Professor Chalmers thinks that original and interesting arguments would likely attract philosophers to the topic and provoke fruitful debate (regardless of whether the arguments themselves are successful).

The new online journal *Animal Sentience* is already establishing itself as a good venue for work on these topics by both scientists and philosophers, although it also illustrates the fact that disagreements about the distribution of phenomenal consciousness in non-human animals are difficult to resolve.

Well-funded post-doc positions could potentially promote beneficial work on this topic.