Philosophy of AI: David Chalmers and the Hard Problem of Consciousness

Stanley James
Mindbuilding Seminar
Winter Semester 2003
University of Osnabrück

The philosophy of David Chalmers is described and critiqued, with emphasis on what it means for the field of Artificial Intelligence. Topics covered are the Hard and Soft problems of consciousness, supervenience, arguments against materialism, and Chalmers' proposal of information as the building block of consciousness.

Introduction

The ultimate goal of artificial intelligence research is to create an artificial system that is as intelligent as a human, so called "Strong AI." But it is worthwhile for AI researchers to take at least a moment to consider some philosophical questions about their would-be creations. And foremost among these questions is, "Can machines be conscious?" In this paper I will consider how this question and related questions would be answered according the philosophy of David Chalmers, philosopher at the University of Arizona. He has achieved some notoriety in recent years for his controversial claim that consciousness is non-material.

This paper expands upon the ideas presented in a presentation I gave in a class entitled "Mindbuilding" at the University of Osnabrück in December of 2003. My purpose in writing is not so much to expound new ideas, but to present important ideas from the Philosophy of Mind to the artificial intelligence community, i.e. "Mindbuilders." The format will go like this: First I relate our question of consciousness to the ideas of computer science legend Alan Turing, and use this to illustrate Chalmers' distinction between the "Hard Problem" and "Easy Problem" of consciousness.. Then I give an overview of supervenience and Chalmers' arguments against materialism. And finally I consider the implications of his proposed idea of proto-consciousness emerging from information.

Turing and his games

The question "Can machines be conscious?" is a shade different from that posed by Turing (1950) in his classic article "Computing Machinery and Intelligence" where he considered the question “Can machines think?” Turing took the question to be unanswerable in principle because the definitions of “machine” and “think” could not be precisely defined. He went on to propose his famous imitation game (now known as “The Turing Test”) as a better means of considering the question.

It may seem that our question of consciousness is in the same boat. After all, “conscious” seems to have no clearer definition than “think.” However, within the last thirty years there have been some developments in the philosophy of mind that, arguably, allow us to give a principled definition of consciousness. This is owed to the work of Patrick Nagel (1974), who in his article “What is it like to be a Bat?” said that something is conscious if there is “something it is like to be it.” For example, it makes no sense to ask “What is it like to be a stone?” because a stone has no phenomenal states. But it does make sense to ask “What is it like to be a baker?” or “What is it like to be an only child.” This stuff, this “what it is like to be,” has been given the name qualia.

It is no surprise that most AI researchers, or at least the ones who pursuing Strong AI, are philosophical materialists. On the surface, this seems like the only logical choice. After all, if you believe that consciousness is made from something other than physical stuff, it doesn’t make much sense for you to try and create consciousness on a computer—something definitely known to be made only of physical stuff! However, Chalmers is a philosopher who firmly claims that Strong AI is possible, but also claims that materialism must be false.

On Chalmer’s view, people who find these views to be irreconcilable are confusing the “Hard” and “Easy” problems of consciousness. The easy problem is what AI Researchers and Neuroscientists do in their laboratories: Figuring out how matter, be it microprocessors and memory chips or neurons and glial cells, can be arranged in such a way that conscious behavior emerges. The hard problem is to understand *why* such arrangements give rise to the phenomenal states described by Nagel. This is quite controversial. What is it that is required above and beyond an explanation of how to build something with conscious behavior? Chalmers himself says “Informal surveys suggest that the numbers run two or three to one in favor of the former view [no hard problem exists] with the ratio fairly constant across academics and students in a variety of fields.” (1996, pg xiii) The remaining sections in this paper will try to make clear Chalmers’ arguments for the hard problem through explanations of supervenience and arguments against materialism.

Supervenience and Such

Descartes argued that consciousness was really a different sort of stuff—a different *substance* from matter. A more modern belief would be that it is possible to arrange matter in such a way as to bring about consciousness, one such example (some would say the only) being the human brain. Descartes' view is not very popular these days, but his argument set up the question of how consciousness relates to matter, and this question is applicable to the modern view as well.

Whenever one property is contingent upon another, there is some sort of *supervenience* relation. As Chalmers describes it, "B-properties *supervene* on A-properties if no two possible situations are identical with respect to their A-properties while differing in their B-properties." (Chalmers 1996, pg 33) For example, the property of color supervenes on the property of reflecting light. You will never find something which has a color but does not reflect light.

Philosophers distinguish between many types of supervenience, but of particular interest to questions of consciousness are logical, natural, and metaphysical supervenience. Their differences are best understood by considering possible worlds. For example, imagine a world in which there exist married bachelors. This world is impossible because it is not logically consistent. Bachelor-ness supervenes logically on the property of being not married. Now consider a world in which bits of matter in space were pushed away from each other instead of coming together, as they do in our world. Such a world is not logically impossible, but has a different set of natural laws. Thus the supervenience of gravity on is natural. "B-properties supervene naturally on A-properties if any time two situations which could naturally arise in our world share the same A-properties; they also share the same B-properties." (Raymore 2002, pg 3)

Metaphysical supervenience is more subtle, but it is a key component of the most successful criticisms of Chalmers' philosophy. It was "discovered" by Kripke in his book *Naming and Necessity* (1972). To understand it, consider a world in which water is not H_2O. Maybe it is XYZ instead. How does the property of being water supervene on the property of being made of two parts hydrogen and one part oxygen? "Therefore, even though it is imaginable that water is not H_2O, one can *discover* the identity of water and H2O *a posteriori*, and this identity is as valid an identity as those which can be known *a priori*. So, water is metaphysically identical to H2O." (Raymore 2002, pg 3)

So how does consciousness supervene onto matter? Or is Descartes correct in stating that it is wholly different from matter, eliminating any supervenience? If consciousness supervenes logically, it means that if consciousness arises from a given arrangement of matter[1] in our universe then, by *logical* necessity, an identical arrangement in any other universe must also give rise to consciousness.

Arguments against Materialism

Arguments against materialism fall into three basic categories: explanatory, conceivability, and knowledge. The explanatory argument is motivated by the intuitions that drive the hard/easy distinction. Namely, a physical description can explain "at most structure and function." (2002, pg 248)

The second flavor of argument is based on the conceivability of beings that appear conscious, but really are not. Imagine a person who behaves perfectly normal, but in fact has no phenomenal states. As the saying goes, "The lights are on but nobody's home." Another way to think about it is to ask if God had a choice about consciousness when he created the world.[2] Would it have been possible to have a world just like ours, but lacking any phenomenal experience? If such a world is conceivable, then consciousness must be an *extra* fact that is not entailed by (is not supervened by) the physical facts.

The third and final flavor of argument is based on epistemology. The idea is that knowledge of all the physical facts does not exhaust all the facts. This argument was made most famous by Frank Jackson with his illustration about Mary. Mary is a neuroscientist who knows everything there is to know about human vision, but unfortunately has lived her entire life in a black in white room. Having never seen red, is she lacking some knowledge? Chalmers presents it as follows: (2002, pg 250)

(1) Mary knows all the physical facts

(2) Mary does not know all the facts

(3) The physical facts do not exhaust all the facts.

[1] To be complete, we should also include the set of physical laws. This is often expressed as "Any arrangement of physical properties…"

[2] Of course, I am using "God" here only in thought-experiment way and take no stand as to how the world really began or the existence of supernatural beings.

Objections

Someone may object even if Chalmers is right, that it doesn't matter: if we know how to build something with consciousness, then there is nothing more to know. An objection would be to consider the world of the Matrix[3]. Suppose that we are in fact living in a simulated world. Now consider what we can know about the "real world" in which our simulation is running. We can know very little about this world. Perhaps it has four dimensions, perhaps the laws of physics are completely different, or perhaps it is completely immaterial. Perhaps something like that imagined by the famous idealist Bishop Berkeley, who claimed that only minds existed and that the world consisted of nothing more than sense-perceptions within these minds. This scenario, while certainly far-fetched, is not logically impossible. Therefore one can see that it is flawed to say an understanding of the physical configurations that will give rise to consciousness constitutes a complete knowledge of consciousness. The laws of physics that we are operating within are not logically necessities, and any knowledge we have which is contingent on them is not complete knowledge.

Information and Proto-Consciousness

If consciousness does not supervene on physical properties, then what is its nature? Chalmers writings are long on attacks on materialism, but short on offering alternatives. In the final chapters of *Consciousness Explained*, he submits a rough proposal that it has something to do with information. In a nutshell, his idea is that any time there is information being processed, there are some phenomenal properties. Or to be more precise, that there are at least "protophenomenal properties." (1996, pg 298)

This view is a hard pill to swallow, as it borders on panpsychism and demands that we accept that things like thermostats have some sort of experience. One counter view that attempts to escape these conclusions is that consciousness may be more like a software program. There are many things that can do computations—computers, calculators, wristwatches and so on—but not all of them can run Microsoft Word. There are certain properties a system must possess before it can run this program. In other words, it's not a continuum: a given device either runs Microsoft Word or it doesn't. It makes sense that consciousness may be the same way. Therefore, the

[3] *The Matrix* (1999) was film about a world in which aliens had imprisoned all of humanity, but hid this fact by connecting everyone to a virtual world run on a computer.

existence of consciousness in certain information-processing systems like brains and maybe some computers, does not demand the existence of “proto consciousness” in simpler systems.

Conslusion

In this paper I have tried to shown that the philosophy of David Chalmers’, while controversial, is relevant to anyone interested in artificial intelligence, and especially those pursing “Strong AI.” The concepts of supervenience phenomenal states have been introduced, and then used in a survey of the most common arguments against materialism. We are living in times of exciting ideas, and age-old questions about what it means to be alive are only becoming more poignant in this high-tech age. The issues presented in the paper should be kept in mind for anyone building a mind.

References

Chalmers, David (1996). *The Conscious Mind*. Oxford University Press.

Chalmers, David (2002). Conciousness and Its Place in Nature. In *Philosophy of Mind: Classical and Contemporary Readings.* Oxford University Press

Nagel, Thomas (1974). “What is it like to be a bat?”, *The Philosophical Review* LXXXIII, 4 (October 1974): 435-50.

Raymore, Paul (2002). *A Materialist Response To David Chalmers' The Conscious Mind.* Retrieved December 28, 2003, from http://www.stanford.edu/group/dualist/vol4/pdfs/raymore.pdf

Turing, A.M. (1950). “Computing Machinery and Intelligence,” *Mind*, 59, 433-460.