

Session 3: The Hard Problem of consciousness

David Chalmers (1996), 'Can consciousness be reductively explained?' in *The Conscious Mind*. Oxford: Oxford University Press.

Physicalism

We can understand physicalism as a global (i.e. world-level) supervenience thesis: all the facts logically supervene on the physical facts, so that a world which is a physical duplicate of this world is a duplicate in every respect. Hence, at least in the actual world, the physical facts determine all the facts. It is clear that consciousness doesn't seem to be a physical fact:

“[C]onsciousness is surprising. If all we knew about were the facts of physics, and even the facts about dynamics and information processing in complex systems, there would be no compelling reason to postulate the existence of conscious experience. If it were not for our direct evidence in the first-person case, the hypothesis would seem unwarranted; almost mystical, perhaps. Yet we know, directly, that there is conscious experience. The question is, how do we reconcile it with everything else we know?” (Chalmers 1996, 20)

The easy and hard problems of consciousness

The *easy* problems of consciousness include those of explaining the following phenomena:

- ability to discriminate, categorise, and react to environmental stimuli;
- integration of information by a cognitive system;
- reportability of mental states;
- ability of a system to access its internal states;
- focus of attention;
- deliberate control of behaviour;
- difference between wakefulness and sleep.

The *hard* problem: “It is undeniable that some organisms are subjects of experience. But the question of how it is that these systems are subjects of experience is perplexing. Why is it that when our cognitive systems engage in visual and auditory information-processing, we have visual or auditory experience: the quality of deep blue, the sensation of middle C?” ... “Why should physical processing give rise to a rich inner life at all?” (Chalmers 1995)

But is it really the physical processing that gives rise to a rich inner life? Then we should assume that the rich inner life supervenes on the physical processing. Several familiar arguments are meant to undermine that assumption.

An explanatory gap?

This argument exploits the distinction between the ‘easy’ and ‘hard’ problems of consciousness.

1. Physical accounts explain at most physical structure, information-processing or psychological function.
2. Explaining physical structure, information-processing or psychological function does not suffice to explain phenomenal consciousness reductively in physical terms.
3. ∴ Phenomenal consciousness cannot be reductively explained in physical terms.

“A physicalist view of the mind must be reductive in one or both of the following senses: it must identify mental phenomena with physical phenomena (ontological reduction) or it must give an explanation of mental phenomena in physical terms (explanatory or conceptual reduction).” (Crane 2000)

Phenomenal zombies

What is a phenomenal zombie? Block: “a full-fledged phenomenal zombie, say, a robot computationally identical to a person, but one whose silicon brain does not support P-consciousness” (Block 1995, 233). Chalmers (1996, 94): physically identical, embedded in an identical environment, functionally identical, and psychologically identical.

1. It is conceivable that there are zombies.
2. If it is conceivable that there are zombies, then it is metaphysically possible that there are zombies.
3. If it is metaphysically possible that there are zombies, then phenomenal consciousness cannot be reductively explained in physical terms.
4. \therefore Phenomenal consciousness cannot be reductively explained in physical terms.

But is the notion of a phenomenal zombie really logically coherent?

Spectrum inversions

“...by the different Structure of our Organs, it were so ordered, that the same Object should produce in several Men’s Minds different Ideas at the same time; v.g., if the Idea, that a Violet produced in [Nonvert]’s Mind by his Eyes, were the same that a Marigold produced in [Invert]’s, and vice versa.” (Locke, 1689, II, xxxii, 15)

1. Nonvert and Invert are physically, functionally and psychologically alike.
2. Nonvert and Invert are not phenomenally alike.
3. If Nonvert and Invert are physically, functionally and psychologically alike, and not phenomenally alike, then phenomenal consciousness cannot be reductively explained in physical terms.
4. \therefore Phenomenal consciousness cannot be reductively explained in physical terms.

Does the argument assume that Nonvert and Invert are visually aware of different properties?

The Knowledge Argument

“Knowledge of all the physical facts will in principle allow Mary to derive all the facts about a system’s reactions, abilities, and cognitive capacities; but she will still be entirely in the dark about its experience of red.” (Chalmers 1996, 101)

1. Mary knows all the physical facts (and their entailments) while in the black-and-white room.
2. If phenomenal consciousness can be reductively explained in physical terms, then there are no phenomenal facts that someone who knows all the physical facts (and their entailments) does not know.
3. Upon leaving the black-and-white room, Mary comes to know a phenomenal fact she did not know before.
4. \therefore Phenomenal consciousness cannot be reductively explained in physical terms.

(Compare this argument with the argument, discussed by Avicenna (Ibn Sīnā, 980-1037) and others, that God does not have knowledge of particulars, even though God knows all the facts about the universe. See: Adamson, 2005)

Revisionary metaphysics

Suppose we take the arguments at face value. If phenomenal properties cannot be reductively explained in the standard ways, then we seem forced (at least) to accept a property dualism. Phenomenal properties are fundamentally non-physical properties. Now what?

For convenience, let's frame this in terms of events. A physical event is an event with physical properties or properties that can be reduced to physical properties. A phenomenal event is an event with phenomenal properties.

If phenomenal properties are fundamentally non-physical properties, then either (A) phenomenal events do not causally interact with physical events, or the physical world is slightly different from how we are used to conceive of it. In the latter case, either (B) the physical world is not causally closed (i.e. some physical events can have non-physical events as sufficient causes), or (C) at least some of the physical causes in the world turn out to have irreducible phenomenal properties.

- A. Epiphenomenalism
- B. Interactionism
- C. Pan(proto)psychism

This is what Chalmers says about C:

“The arguments do not lead us to a dualism such as that of Descartes, with a separate realm of mental substance that exerts its own influence on physical processes. The best evidence of contemporary science tells us that the physical world is more or less causally closed: for every physical event, there is a physical sufficient cause. If so, there is no room for a mental “ghost in the machine” to do any extra causal work. [...] The dualism implied here is instead a kind of property dualism: conscious experience involves properties of an individual that are not entailed by the physical properties of that individual, although they may depend lawfully on those properties. Consciousness is a feature of the world over and above the physical features of the world. [...] It remains plausible, however, that consciousness arises from a physical basis, even though it is not entailed by that basis. The position we are left with is that consciousness arises from a physical substrate in virtue of certain contingent laws of nature, which are not themselves implied by physical laws. [...] (Chalmers 1996, 124ff)