

Logic and Metaphysics

CONSCIOUSNESS

DAVID PAPINEAU SEMESTER 1 2001-2

Fridays 2pm Room 2B08, 2nd Basement, Strand, KCL

Introduction

The lectures will be about the relation between phenomenally conscious states (states that it is like something to have) and the physical world. In the first lecture I shall lay out the basic causal argument for the materialist view that conscious states are identical to material states. Then I shall consider two celebrated arguments against materialism: Jackson's knowledge argument (lecture 2) and Kripke's modal argument from the possibility of zombies (lecture 3). This will lead to an analysis of a special set of concepts humans use to think about conscious states, phenomenal concepts, and the confusions they can engender (lecture 4). Finally I shall consider the prospects for a scientific study of consciousness (lecture 5).

Lecture 1—The Causal Argument for Materialism

- (1) Mental occurrences have physical effects
 - (2) Physical effects are always entirely fixed by prior physical causes
 - (3) The physical effects of mental causes are not always overdetermined
- Therefore
- (4) The mental occurrences in (1) are identical with the physical causes in (2).

If you accept this argument, you can read it in a weaker or stronger sense, depending on whether you think causation is a relationship between events (Davidson) or facts (Mellor, Kim). But we can by-pass this issue, because even if you are a Davidsonian about causation, the argument can be reformulated to yield fact identity anyway. It follows that conscious properties must be identical to physical properties.

Should we accept the causal argument? You can deny (1), and be an

‘epiphenomenalist’; or deny (3), and adopt the ‘belt-and-braces’ view. But both of these seem less simple views than the identity theory (at least until we are given special arguments against the latter). Historically more common was to deny (2)—the causal completeness of physics (which is why materialism about the mind is so much a late 20C view).

Doesn’t the causal argument prove too much? For it identifies conscious states with strictly physical states, rather than multiply realized functional states (or disjunctions of physical states, or otherwise physically supervenient higher material states), and so implies that humans and octopuses can’t share the property of being in pain. Well, there is a version of the causal argument which reads causation weakly (as ‘supervenient causation’) and delivers the conclusion that conscious properties, if not strictly physical, must at least be physically realized higher properties (otherwise they wouldn’t even cause physical effects in the weak sense). I’m not sure which way to go here—but this won’t matter henceforth, since either way conscious properties will be identical with ‘material’ properties (ie strictly physical properties or physically realized higher properties). This is enough to conflict with Jackson’s and Kripke’s arguments.

Lecture 2—Jackson’s Knowledge Argument

(i) The Story of Mary. She is a super-duper future colour vision scientist, who knows everything there is to know about colour vision from a materialist point of view. But she’s never seen colours herself—it’s all book learnin’. Then one day she sees a red rose.

At this point she learns something new, says Jackson (she learns about the phenomenal side of *red* experiences). But note she already knew about every material property of red experiences. So there must be properties of conscious experiences which are non-material. (So, given the completeness of physics, Jackson opts for epiphenomenalism.)

(ii) The Inflationist Answer (short version). I say that Jackson’s argument shows only that Mary acquires a new (‘phenomenal’) concept for a property she could

always refer to in material terms, not that there are any extra phenomenal properties. (Cf. You know of all the children with names from the baptismal register who attended some school in Bristol 100 years ago, including ‘Archie Leach’. Then you learn something new. You learn that Cary Grant attended that school. So there must have been at least one child at that school with no name from the baptismal register.)

(iii) Denying the Difference. Not all materialist philosophers agree with this ‘inflationist’ conceptual response to Jackson. Dennett, for example, denies that Mary would learn anything new to start with, since he doesn’t accept there would be any before-after difference in Mary (towards the end of his Consciousness Explained). But that seems silly. Surely Mary acquires new abilities to visually imagine red experiences, and to introspectively classify other experiences as of that type.

(iv) The Ability Hypothesis. David Lewis allows that there is a real before-after difference in Mary. She does learn something new—she acquires imaginative and introspective abilities. But this is just new ‘know-how’, not knowing that anything. (Consider the omniscient but non-practising bicycle scientist, who then learns something new—how to ride a bicycle. No temptation here to infer there are non-material properties of bicycle-riding.) Similarly, Lewis points out that if we think of the before-after difference in Mary as matter of new abilities, there’s no need to posit any non-material properties of experience. There’s a natural materialist explanation of what Mary’s new abilities consist in (she has a new pattern stored in her brain), and moreover one which explains why you need an original experience to ‘know what such experiences are like’ in this sense.

(v) New Concepts. That’s all fine. But I say Mary’s new abilities also yield new concepts—she doesn’t just have new know-how, but also new know-that (though only at the level of sense, not reference). Suppose Mary is shown a sheet of coloured paper, but not told (in her old materialist terms) which colour it is. A bit later she might think (imagining) ‘I bet I’ll have that experience again today’. This isn’t a thought she can form using her old concepts (she doesn’t know which

one to use). Yet it's a perfectly good thought, which might be true or false. Concepts are components in truth-evaluable thoughts. Ergo, Mary acquires a new concept.

Lecture 3—Kripke's Modal Argument: The Possibility of Zombies

(i) A Posteriority is not Contingency. Early identity theorists (Smart, Place) used to say mind-brain identities were 'contingent', by which they meant that they were for science to discover, not to be figured out a priori (cf. heat = molecular motion). Kripke objected that this muddled the epistemological notion of a posteriority with the metaphysical notion of contingency. Mind-brain identities may be a posteriori. But, if true, they must be necessary, like all identities.

(ii) The Appearance of Contingency. It might seem easy for materialists simply to reply to Kripke 'Yes, that's it, what you say—a posteriori yet necessary'. But the trouble now is that 'pains = C-fibres' doesn't seem necessary, like eg Cicero = Tully does. It's all too easy to imagine beings in whom mind and brain come apart. Zombies and ghosts. Why should these seem possible, if they aren't? Materialists need at least to explain the appearance of contingency.

(iii) The Unhelpful Analogy with Scientific Identities. Can't materialists draw an analogy with scientific identities? Thus, heat = molecular motion, yet molecular motion might not have been heat, at least in the sense it mightn't have caused heat sensations. Strictly speaking, this isn't the possibility that molecular motion not be heat, but still, we can see how someone might confusedly get the impression these things could come apart. So can't we tell the same story with pain and C-fibres? But it doesn't work. Consider. C-fibres might not have been pain, at least in the sense of not causing pain sensations. But this is a world in which C-fibres really aren't pain, not just one which confuses us into thinking they could come apart. And if the relation between C-fibres and pain is contingent in this way, concludes Kripke, they can't be identical.

(iv) Thinking Impossible Thoughts. Materialists shouldn't try to explain the appearance of contingency by drawing an analogy with scientific identities like

heat = molecular motion. (Look at it like this: as soon as materialists say that there is a possible world where C-fibres don't produce some conscious reactions that they do produce in us, then they've lost—for this admits the possibility of a being who has all our material properties but lacks one of our conscious properties, which wouldn't be possible if the conscious property at issue were identical with any of those material properties.) Rather, materialists should simply say we can have thoughts to which no possibility of any kind corresponds. For example, someone might think Cicero \neq Tully. (Alice laughed. "There's no use trying," she said: "one can't believe impossible things." "I daresay you haven't had much practice" said the Queen. "When I was your age, I always did it for half-an-hour a day. Why, sometimes I've believed as many as six impossible things before breakfast.") Those who support Kripke's argument insist that whenever you can conceive $a \neq b$, either a and b really aren't equal, or you must be thinking of the entity in question as lacking some description you associate with one of the terms. This assumes that for any two terms 'a' and 'b' that you don't associate with descriptions, you must know they co-refer if they do (and so find yourself unable to conceive $a \neq b$). But why suppose this? (If you believe Kripke's (!) causal theory of proper names, you can easily master two names which you don't associate with any descriptions, and yet not know they co-refer when they do.)

Lecture 4—Phenomenal Concepts

(i) "Pain" is ambiguous. I have it that we can think about pains etc in two ways: using phenomenal concepts (how it feels) and using material (ie non-phenomenal) concepts. A special case of the latter are psychological concepts, which pick out experiences in terms of their causal role. I take it that the English word "pain" expresses both a phenomenal and a psychological concept, and that we don't bother to disambiguate because we take it that both concepts refer to the same property. (Cf. the way "mammal" expressed both the property non-egg-layer and lactater.)

(ii) How do phenomenal concepts refer, if not by description? Well, the theory

of reference is a big topic. It is useful to think of phenomenal concepts as a species of demonstrative term: “this experience:--”, with the gap filled by an actual experience, or imaginative recreation thereof. The term refers to this experience (token or type).

(iii) Incorrigibility. Note how this model explains the incorrigibility of certain experiential self-ascriptions. When I have an experience, form a term “this experience: ----” for that type of experience, and then use this to judge “Now I am having this experience: ----”, there is no room for me to go wrong. Some Wittgensteinians might feel that this shows there is something wrong with phenomenal concepts. I think it is a definite virtue that phenomenal concepts explain this (limited) sort of incorrigibility.

(iv) The compelling illusion of mind-brain distinctness is also explained by the structure of phenomenal concepts. (“How can grey matter gives rise to this technicolour phenomenology?” McGinn) Note how on the left here we use a material concept (grey matter), and on the right a phenomenal one (technicolour phenomenology). When we use the phenomenal concept we activate the experience (or an imaginative copy). Not so with the material concept. So we judge that the material concept “leaves out” the experience. However, the fact the material concept doesn’t use the felt experience doesn’t imply it doesn’t refer to it.

Lecture 5—Prospects for the Scientific Study of Consciousness

(i) Phenomenal consciousness research aims to identify the material referents of phenomenal concepts. You can think of this as ‘hard problem’ research, in Chalmers’ sense. (NB there is a determinable phenomenal concept consciousness-as-such, in addition to the determinates like pain, seeing an elephant, and so on.) Consciousness research hinges crucially on subjects’ phenomenal reports about past and present experiences. To see this, note how such research is impossible with monkeys, say.

(ii) Too many candidates offer themselves as material referents for phenomenal

concepts, and the standard methodology can't decide between them. Most familiarly, if some strictly physical S coincides with P in humans, so will some structural F.

Similar problems arise with broad vs narrow candidates (assuming some representational view of consciousness), and again with HOT vs non-HO candidates.

(iii) Vagueness. Block (forthcoming, J Phil) takes this indecision to show that 'inflationism' generates insuperable epistemological problems. I say it is just an upshot of vagueness in our phenomenal concepts._