

## Lecture 2: The Knowledge Argument

This lecture we consider the case for a physicalist view. We look at how physicalists have responded to the arguments for property dualism set out in the previous chapter.

Let us then look at some physicalist responses to the arguments in support of property dualism, beginning with the knowledge argument.

### I. The Knowledge Argument

Recall:

*Premise 1:* If physicalism is true, then on leaving her room Mary will not learn any new facts about colour vision (since, by hypothesis, she already knows all the physical facts about it).

*Premise 2:* On leaving her room, Mary learns new facts about colour vision (namely what it's like to see various colours).

*Conclusion:* Physicalism is not true.

The Knowledge Argument derives a metaphysical conclusion from an epistemic premise. Is that a problem? Not obviously: knowledge implies truth, so it's not surprising that epistemic premises could have metaphysical consequences.

#### *The no-learning view: Denying Premise (2)*

According to Dennett, we have no good reason to think that Mary could not work out what it is like to see colours from the physical information available to her in her room. It is true that we have a strong intuition that she could not but, Dennett argues, we have no reason to believe that our intuitions about this case are reliable.

Dennett is generally suspicious of thought experiments, which he refers to as "intuition pumps". The function of intuition pumps is:

to entrain a family of imaginative reflection in the reader that ultimately yields not a formal conclusion but a dictate of 'intuition'. Intuition pumps are cunningly designed to focus the reader's attention on the 'important' features, and to deflect the reader from bogging down in hard-to-follow details. (Dennett, 1984, 12)

Dennett worries that some intuition pumps in particular can be seriously misleading. Often they encourage us to imagine a scenario that's much simpler than the one they officially describe. Dennett asks us to imagine that upon leaving the room, Mary is given a blue banana. According to Dennett, we can very well imagine that she would know that we are trying to trick her. That's because she knows what physical effects different colours would have on her nervous system, including what thoughts they would evoke in her. So she could tell which colour experiences she is having by noting these effects.

Some comments:

- Does Dennett really succeed in showing that Jackson's example is unreliable? While it's hard to imagine Mary's situation in detail, we have an idea of the sort of information Mary will have: it will be information about neurological mechanisms, the functions they perform, etc. Can such facts entail phenomenal facts?
- On the other hand, can we really be sure that they wouldn't?
- We can see how someone who has no colour vision might get some understanding of it: the experience of seeing red has a relation to the experience of seeing other colours: red is close to yellow, closer to orange, opposite to green, it's felt as being warm,... If Mary knew all this, maybe she would be able to recognize colours straight off. And if so, it might be hard to deny that she would know what colour vision was like.

The no-learning view is not popular amongst physicalists. Insofar as physicalists want to attack Premise (2), they focus on an alternative strategy.

#### *The Fregean Strategy*

## Part II, Consciousness

The third response concedes that Mary does acquire some factual knowledge, but maintains that this kind of knowledge does not pose a threat to physicalism: what she learns are not new non-physical facts but familiar physical facts known in a new way. This view has been defended by e.g. Michael Tye.

For this response, we need an account of modes of presentation and contents of belief that will be more fine-grained than distinctions between facts.

The Fregean view relies on an account of concepts:

- Concepts are mental representations.
- They are constituents of thoughts.
- Concepts represent or refer to things but they don't represent transparently. They represent under 'modes of presentation'. The concepts 'water' and 'H<sub>2</sub>O' both represent the same substance but under different modes of presentation.

Learning what an experience is like, Tye argues, is simply a matter of coming to conceptualize it in a certain way, by applying phenomenal concepts. These concepts are special in that they can be acquired only by someone who has undergone the experiences they represent:

Phenomenal concepts, I maintain, are conceptually irreducible concepts that function in the right sort of way. To possess the phenomenal concept red, for example, is to possess a simple concept that has been acquired by undergoing experiences of red...and that not only dispose one to form a visual image of red in response to a range of cognitive tasks pertaining to red but also is brought to bear in discriminating the experience of red from other color experiences in a direct and immediate manner via introspection. The functional role that the concept plays is what makes it perspectival. A person who is blind from birth or who is always restricted to an environment of things with achromatic colors cannot possess a concept with the requisite role and hence cannot possess the phenomenal concept red.

An objection:

Conceptually possible worlds are characterizations of a world that are conceptually coherent: those that a thinker who is competent with the constitutive concepts would judge to be possible. Metaphysically possible worlds meet a further substantive condition: constraints on the compatibility of properties that are picked out by the concepts. So one can be fully competent with each of two concepts and be able to form a clear and coherent conception of a situation in which one is exemplified and the other is not, even though there is no metaphysically possible situation in which one is exemplified and the other is not.

But how do we explain conceptual capacities if not in terms of the ability to form accurate conceptions of possible situations in which they are exemplified?

### *The Ability Hypothesis*

This response to the knowledge argument concedes that Mary gains new knowledge on leaving her room, but denies that she learns new facts. The most important advocates of this view are David Lewis and Laurence Nemirow.

Distinguish two kinds of knowledge:

- i. knowledge of facts (propositional knowledge)
- ii. knowledge of how to do something (practical knowledge)

Learning lots of facts about swimming is not the same as learning how to swim. Learning how to swim is to gain the ability to do it.

According to this response, knowing what it's like to see colours is also knowledge of the second kind, not the first. It is a matter of possessing abilities, rather than grasping facts: the ability to remember, imagine and recognize colour experiences.

## Part II, Consciousness

One nice feature of the ability hypothesis is that it explains the ineffability of experience. If knowing what an experience is like is pure no how, then it is not surprising that we find it impossible to put it into words. An expert swimmer might be completely unable to describe how to swim; they just know how to do it.

Does the ability hypothesis work? Does knowing what an experience is like involve only the abilities Lewis describes? Some objections:

- When we talk of knowing what X is like, we mean knowing that X has certain properties.
- Learning what an experience is like brings also the ability to think new thoughts about it. She can now think “So this is what an seeing yellow is like. For all I knew it could have been like *that* (pointing to something green)”. Mary can represent to herself the possibilities that she previously could not distinguish between. So, it seems that Mary has acquired new cognitive abilities that require some information.

These objections suggest that possessing the abilities Lewis describes is not sufficient for knowing what an experience is like. Are such abilities necessary? The richness of some experiences may outstrip our abilities to remember, recognize or imagine them.

### *The Self-Location Strategy*

This strategy notes that Mary’s case is similar to cases of self-location. Consider the following case: Professor Lingens is a famous professor who is trapped in the Widener Library. He is also an amnesiac. He picks up a book from the shelf that happens to be his autobiography. As he reads the book he learns many facts about Professor Lingens: where he was born, where he grew up, etc. But he is missing one crucial piece of information: namely that *he* is Professor Lingens.

We can imagine similar cases in which the agent is missing information about where she is located, or what time it is even as she knows all physical facts. These cases suggest that our knowledge of ourselves and our place in the world cannot be reduced to impersonal objective knowledge about what the world, as it is in itself, is like. But we are not tempted to draw a strong metaphysical conclusion in those cases: that there are non-physical, objective facts about the self.

Stalnaker suggests that this is one way we might tackle the Mary argument: we need to develop an account about the nature of self-locating knowledge.

## IV. Readings

Carruthers, P. (2000). *Phenomenal Consciousness: A Naturalistic Theory*. CUP.

Dennett (1991) *Consciousness Explained*. Penguin.

Lewis, D. (1990). *What Experiences Teaches*.

Lycan (2003) *Perspectival representation and the knowledge argument*.

Papineau (2002) *Thinking about Consciousness*. OUP.

Stalnaker (2007). *Our Knowledge of the Internal World*. OUP.

Tye (1995) *Ten Problems of Consciousness*. MIT Press.