Kripke vs. Reductive Physicalism

Introduction

- In the final lecture of his Naming and Necessity, Kripke presents an argument against type-identity/reductive physicalism. Many have thought of it as a sophisticated resurrection of Descartes's argument for the "real distinction between mind and body" in Meditation 6.

The historical context

Here is one way of representing Descartes's argument from the conceivability of disembodied existence (i.e., 'the conceivability argument'):

- P1. I can clearly and distinctly conceive of my mind existing without my body existing (or without anything else physical existing). (see Meditation 2 where Descartes is in the grips of hyperbolic doubt)
- P2. Whatever is clearly and distinctly conceivable is metaphysically possible.
- C1. It is possible for my mind to exist without my body existing (or without anything else physical existing).
- C2. My mind is not my body (or anything else physical).

From here it is easy enough to move (via some plausible assumptions) to the Cartesian conclusion that minds in general are nonphysical substances.

The historical context

The conceivability argument raises various questions. For example, what is "clear and distinct conceiving", and is it true that we can conceive of our minds existing in the absence of anything physical? And why think that conceivability – something apparently epistemic – is a guide to possibility – something apparently metaphysical?

But to understand the context for Kripke's discussion we need to focus on an aspect of the argument to which you might have paid little attention at first: why think that C2 follows from C1? One thing Descartes could say is this:

If it was possible for Clark Kent to exist without Superman existing, then clearly Clark Kent and Superman would have to be numerically distinct people. Likewise, if my mind could exist without my body existing (or anything else physical existing), then my mind would have to be numerically distinct from my body (or anything else physical).

The historical context

That does sound plausible. But what about cases like “Aaron Henry is the instructor of PHIL451”? It certainly seems possible for me to have existed without being the instructor of PHIL451 – that’s a way things could have turned out. So perhaps “my mind = my body” tracks this example, not the Clark and Superman one.

The historical context

It is arguably this sort of thought that led J.J.C. Smart, in his classic 1959 defense of the type-identity theory ("Sensations and Brain Processes"), to reject Cartesian conceivability arguments against mind-brain identity. Smart's view is that mental property-physical property identity statements like

(*) Pain = C-fibre firing (CFF)

are true but contingently so. In the actual world, pain is CFF, Smart believes, but there are possible worlds in which it isn't. That is why we can conceive of pain in the absence of CFF, even though (as a matter of empirical fact) pain is identical to CFF. (Likewise, in the actual world, Aaron Henry is the instructor of PHIL451, even though there are possible worlds in which he isn't, which is why we can conceive of Aaron Henry without the property of being the instructor of PHIL451.) According to Smart, all that follows from the conceivability of pain in the absence of CFF is that the words or concepts "pain" and "CFF" have different 'senses' or 'modes of presentation'. As Frege made famous with his example of Hesperus and Phosphorus, concepts with different senses sometimes co-refer.

The historical context

To clarify: Smart wants to concede premises like P1 and (maybe) P2 to Descartes while avoiding Descartes' conclusions by insisting that C2 does not follow from C1: mind-body identities are merely contingently true. In this respect, it is like any other historical example of theoretical identification (lightning and electrical discharge, water and H2O, heat and mean kinetic energy, etc.). In general, the negation of an empirically established identity statement is perfectly conceivable, and (maybe) metaphysically possible. (It could have been that the stuff we pick out with the word 'water' had a different underlying chemical composition than it turned out to have). But the identity statements are nevertheless perfectly true. Likewise, Smart proposed, for the identity theorist's identification of mental state-types with brain state-types.

- Many at the time were (rightly!) impressed with Smart's argument, and inferred that, whatever one feels about conceivability and possibility, Descartes's argument fails at the move from C1 to C2. Kripke changed all that.

Technical Background

Here is some important background information. (Kripke defends these and other ideas in the earlier lectures of Naming and Necessity.)

Technical Background: The necessity of identity

- Consider any object o. o is identical with itself. Moreover, o couldn't possibly have been identical to anything other than itself. Necessarily, o is identical to itself.

Technical Background: Rigid Designation

Some examples of referring terms: "Justin Trudeau" and "The prime minister of Canada."

Briefly and roughly, a referring term "T" is a 'rigid designator' if and only if it designates the same object in every possible world. A rigid designator "T" rigidly designates an object o if and only if it designates o in every possible world.

- Here is a helpful way to think about this. Take some possible world, w. Now consider the questions: "Who (or what) is T in w?" and "Who (or what) is T in the actual world?" If, for every world, w, these questions have the same answer – a certain object, o – then "T" is a rigid designator. But if the questions can be read so that the answer to one is "A certain object o" while the answer to the other is "A certain object o*," such that o and o* are distinct objects, then "T" is not a rigid designator.

- Exercise: Try the above using "Justin Trudeau" and "The prime minister of Canada."

Technical Background: Rigid Designators and Necessarily True Identity Statements

In Naming and Necessity, Kripke famously argues that where "A" and "B" are both rigid designators, then if the identity statement "A = B" is true, it is true necessarily.

Proof:

1. "A = B" is true. (Assumption)
2. If "A" and "B" are both rigid, then each refers to the same object in every possible world.
3. In the actual world, "A" refers to o and "B" refers to o.
4. In every possible world, "A" refers to o and "B" refers to o.
5. If in every possible world, "A" refers to o and "B" refers to o, then "A = B" is true in every possible world.
6. "A = B" is true in every possible world.
7. If "A = B" is true, it is necessarily true.

Technical Background: Rigid Designators and English

- Kripke argues that in English and other natural languages proper names (e.g., “Socrates”) and nouns for natural kinds (e.g., “water”) are rigid designators. He also argues that “pain” and other phenomenal terms, as well as scientific terms like “CFF,” are rigid designators.

Technical Background: Scientific/theoretical identity statements

If Kripke is right about all of this, then it's a big deal. Why? For one thing, it entails that scientific identities like

(W) Water = H2O

are necessarily true. Why? Because they're true (let's assume) and because "water" and "H2O" are each rigid designators. But notice, they're not only necessarily true. They're necessarily true and a posteriori (i.e., knowable only through empirical means). Why? Because it takes empirical investigation to find out that water is one and the same thing as H2O. That means that science is in the business, not just of tallying up contingent regularities, but of discovering necessary truths – specifically, truths about the essences of things.

Technical Background: Scientific/theoretical identity statements

Why is this such a big deal? For one, it challenges the clean and tidy Humean distinction between "relations of ideas" and "matters of fact", which was resuscitated with gusto early in the 20th century.

On this tidy picture, there are truths that are: necessary (true in all possible worlds), a priori (knowable independent of empirical investigation), and analytic (truth in virtue of the meanings of the terms involved). And then there are truths that are: contingent (true in some but not all possible worlds), a posteriori (knowable only through empirical investigation), and synthetic (true not in virtue of the meanings of the terms involved).

The Kripkean picture threw a wrench in this familiar picture.

Kripke on the C1-C2 Inference

> We can now see that the move from C1 to C2 in the Cartesian conceivability argument is valid (i.e., from "It is possible for my mind to exist without my body existing (or without anything else physical existing)" to "My mind is not my body (or anything else physical).").

The "Aaron Henry is the instructor of PHIL451" counterexample isn't a genuine counterexample because "the instructor of PHIL451" is not a rigid designator. It operates as a definite description that picks out different objects in different possible worlds (depending on which object in that world satisfies the condition that the description imposes). By contrast, "my mind" and "my body" are both rigid.

- The moral of the story for Kripke: if P1 and P2 are both true, Descartes's argument goes through. Smart was wrong to dismiss the conceivability argument.
- Fine. So, are P1 and P2 true?

The Descartes-Kripke modal argument

(P1) It is conceivable that pain is not identical with CFF.

(P2) If it is conceivable that $x \neq y$, then it is possible that $x \neq y$.

(C1) It is possible that pain is not identical with CFF.

(P3) If it is possible that $x \neq y$, then $x \neq y$.

(C2) Pain is not identical with CFF.

Some notes about the argument

- P1 seems plausible on its face. It certainly seems as if there could be pain in the absence of CFF and vice versa. As Kripke says, the connection between pain and CFF "appears contingent" or non-necessary.
- P2 expresses the plausible-seeming view that conceivability is a reliable guide to metaphysical possibility.
- P3 follows from the necessity of identity, and C2 follows from P3 together with: (i) Kripke's claim that "pain" and "CFF" are rigid designators and (ii) his proof that identities involving rigid designators are necessarily true (if true at all).
- And each inference in the argument is valid.

The reductive physicalist's objection

Now for the crucial Kripkean contribution. As Kripke recognizes, reductive physicalists will reply by asking why, if the Descartes-Kripke Argument is sound, the following obviously unsound argument isn't also sound:

An Obviously Unsound Argument

- (P1*) It is conceivable that heat is not identical with mean molecular motion (MMM).
- (P2) If it is conceivable that $x \neq y$, then it is possible that $x \neq y$.
- (C1*) It is possible that heat is not identical with MMM.
- (P3) If it is possible that $x \neq y$, then $x \neq y$.
- (C2*) Heat is not identical with MMM.

The reductive physicalist's objection

P1* is plausible: the connection between heat and MMM "appears contingent." The rest of this argument goes through just as the previous argument did. But C2* contradicts something we know to be true: that heat is MMM. The reductive physicalist will insist that since the Obviously Unsound Argument is no good, the Descartes-Kripke Argument must not be any good either.

- Note: even if the reductive physicalist is right about this, we still need to know *where* the problem in the two arguments lies. Next week, will carefully examine both "a priori physicalist" and "a posteriori physicalist" ways of diagnosing the alleged flaw in arguments like this one.

Kripke's reply

Kripke's first move is to say that the "appearance of contingency" generating support for P1* – the sense we have that there could be heat in the absence of MMM (and vice versa) – can be *explained away*.

When we think we're imagining a possible world in which heat isn't MMM, what we're *really* imagining is a possible world in which something other than MMM is causally responsible for the "effects" that MMM is causally responsible for in the actual world. In particular, this is a world in which something other than MMM gives rise to the same heat *appearance* in us that MMM gives rise to in the actual world. Once we see this, we also see that (P1*) is false. We *think* a certain proposition – that heat exists without MMM – is conceivable, but we're confused. We are confusing the inconceivable proposition (that heat could exist without MMM) with a different, conceivable proposition (that something other than MMM could have the same sensory effects on us that MMM has). So, heat *is* identical with MMM.

Kripke's reply

The question that now confronts the reductive physicalist is whether they can pull off a parallel move with respect to P1 in the Descartes-Kripke argument. Can they show that when we *think* we are conceiving of pain in the absence of CFF we are *really* conceiving of a possible world in which something other than CFF gives rise to the same appearance of pain that CFF gives rise to in the actual world?

No, says Kripke. This is because, unlike with heat, we can't make sense of an appearance-reality gap for the case of pain (likewise for other phenomenally conscious states). There is no possible world in which something other than pain gives rise to the appearance of pain: pain *just is* the "appearance of pain." This means that the "appearance of contingency" generating support for P1 can't be explained away as was done in the heat case. So, P1 stands, and the Descartes-Kripke Argument is rescued. Reductive physicalism really is false.