

PRECIS OF KNOWLEDGE, POSSIBILITY AND CONSCIOUSNESS**JOHN PERRY**

[Philosophy and Phenomenological Research 68(2004): 172-181]

In *Knowledge, Possibility and Consciousness*¹ I argue that the Zombie Argument, the Knowledge Argument, and the Modal Argument do not provide people with broadly common-sensical views about consciousness and the mental, and an inclination towards physicalism, any reasons not to be physicalists. That is, they do not support the doctrine of *neo-dualism*, advocated by Chalmers², Jackson³, and others: although the mind may be the brain, qualia, the *what-its-like* properties of experiences that makes them experiences, are not physical properties.

I claim that as long as the physicalist maintains that these properties simply *are* physical properties, rather than holding some seemingly more sophisticated, but in fact more vulnerable, doctrine of supervenience, and adopts certain independently plausible views about knowledge and possibility, the force of the arguments can be evaded. Such a physicalist I call an “antecedent physicalist,” by which I have in mind someone like myself, who is inclined toward physicalism, and thus inclined to suppose that experiences *are* physical states, until exposed to some good argument to show otherwise. Thus the book is directed at two groups: those that are sympathetic to physicalism but do not know what to make of these arguments, and those that deny physicalism because of one or more of these arguments. Readers outside these two groups are, however, most welcome.

The point about knowledge and possibility is this. There are two ways we can think about possibility, and hence about the content of states of knowledge, thought of as the possibilities the knowledge permits. First there are the

¹ John Perry, *Knowledge, Possibility and Consciousness* (Cambridge, MIT Press, 2001); hereafter, *KPC*.

² David Chalmers, *The Conscious Mind* (New York, Oxford, 1996).

³ Frank Jackson, . What Mary Didn't Know. *The Journal of Philosophy* 83 (1986): 291-295

possibilities for the *subject matter*: the extra-linguistic properties and things in the world the possibilities are possibilities *for*, or the knowledge is knowledge *of*. Thought of in this way, there is *no* possibility that Mark Twain is Charles Dickens, and *no* possibility that Mark Twain is not Samuel Clemens. Once we fix the subject matter of our thoughts (what the ideas are *about* or *of*) and the bits of language we use to express them (what the words *refer to*) all we have to deal with are the persons Clemens and Dickens. And there is no way Clemens and Dickens can be one and the same, and no way Clemens and Twain can be two instead of one.⁴ If we think about possibilities this way, it seems no possibilities are eliminated when we learn that Twain is Clemens, or that Twain is not Dickens.

Russell's theory of descriptions and Frege's concept of a *mode of presentation* are often deployed to retreat from things to properties while staying in the realm of extra-linguistic subject matter. Maybe the possibility we hadn't eliminated, when we didn't know that Twain and Clemens were one, was something like the possibility that Clemens was *the author of Huckleberry Finn*.

But clearly, in addition to these subject matter possibilities, there are additional possibilities that take into account how our thought and language might fit onto or into the rest of the world. There are coherent possibilities in which my Dickens-thoughts and my Clemens-thoughts end up being of one and the same person. These are possibilities in which there is one person, to whom both thoughts are traced back---in which, say, Dickens became a sailor rather than a novelist, and Clemens had two pseudonyms and wrote novels like the *Tale of Two Cities* under the pseudonym "Charles Dickens" and novels like *Huckleberry Finn* under the "Mark Twain" pseudonym. There are also coherent possibilities in which my Twain-thoughts and my Clemens-thoughts are of different people. It is

⁴ At least there is no way that is particularly relevant here. Clemens could split amoeba-like into two people; but of course then so would Twain; it's not clear how Clemens-Twain could split in such a way that one of the resulting people was Clemens and the other was Twain. Clemens and Dickens could somehow merge, although it is not at all clear, and perhaps not even plausible, that the result would be either of them. For thoughts about such situations, see my *Identity, Personal Identity and the Self* (Indianapolis, Hackett Publishing Company, 2002).

these possibilities that I eliminate when I learn that Dickens and Clemens are two, and Twain and Clemens are one and the same. I call these possibilities, and the knowledge that eliminates them, *reflexive*, for no more profound reason than that they are possibilities about the pertinent thoughts and language *themselves*.

All of this would be straightforward, and hardly the basis for philosophical confusion, were it not for the fact that our ordinary thoughts have both kinds of content; they pertain to both sorts of possibilities. My thought that Twain wrote *Huckleberry Finn* cannot be true unless the subject, Clemens, Twain, Sam, Mark--- whatever you call him--- wrote that great novel. That is the subject matter content of the thought. That is what *further* conditions the things have to meet for my thought to be true, *given* which person and which novel the concepts involved in my thought are *of*. But we can also assign that thought truth-conditions without taking those things as given. *Whoever* my Mark Twain concept is of, and *whichever* novel my *Huckleberry Finn* concept is of, my thought cannot be true unless that person wrote that book. That is the reflexive content of my thought (or rather one of them, for there are many reflexive contents, depending on which thought-world connections we keep fixed, and which we allow to vary). Tradition identifies *the* content of a thought with its subject matter content, conceived either referentially or descriptively. But sometimes the subject matter content of a thought is trivial (Clemens = Twain) or impossible (Clemens = Dickens) when the thought itself is not, and there is no viable extra-linguistic mode of presentation. It is then that we need to retreat to a reflexive content. If we do not, we will be inclined to postulate extra subject matter, to find the relevant contingencies, and that, the "subject matter fallacy", is at the heart of all three arguments.

Identity then is key to both points I make. The physicalism I defend holds that the *what-its-like* properties *are* physical properties; that is, it is a type-type identity theory about experiences and physical states. And cases of identity are those in which missing the distinction between reflexive and subject matter content is most likely to provoke the subject matter fallacy.

THE ZOMBIE ARGUMENT,

The Zombie Argument, due to Chalmers, holds that it is conceivable that there be a Zombie world; that is, a possible world i) physically indiscernible from the actual world, and ii) with no phenomenal states (referred to by various authors as qualia, raw feels, sensations, experiences, subjective characters of experiences, phenomenal states or properties, and so forth). Since the Zombie world is conceivable, the argument goes, it is possible. Since it is possible, phenomenal states cannot be physical states; if they were, then a candidate Zombie world would fail. If it were physically indiscernible it would contain the phenomenal states (since they are among physical states). If it lacked the phenomenal states it would not be physically indiscernible (since some of the physical states would be missing).

Consider possible worlds physically indiscernible from our own, i.e. worlds that satisfy condition i). I am to think that at least one such world, call it *w*, that also satisfies condition ii). For example, my current headache does not occur in *w*. This requires that the state my current headache thoughts are about is not any state that occurs in *w*, that the state to which I turn my attention rather than working on this article does not occur in *w*, and so forth. My Zombie-twin may be writhing and holding his head as his typing tapers off; but there is no pain; nothing has the property that I have, and attend to, when I think of what is going on in my head as *this damn headache*. I think such a world is conceivable at most in the reflexive sense; in the sense in which it is conceivable that Clemens is not Twain. But then the second step of the argument fails. The principle that if a world is conceivable, it is possible, clearly does not work when we go from a conceivability of a thought at the level of reflexive content, to its possibility at the level of subject matter content.

In the book, I leave the connection between the Zombie Argument and issues of reflexive content to the last chapter, on the modal argument. I use the discussion of the Zombie Argument highlight the importance of holding an identity theory as opposed to a supervenience theory of phenomenal states. I

argue that unless one already has been convinced that our experiences are mere epiphenomena, one will not be able to follow Chalmers' directions and conceive a world such as *w*. Suppose one holds the common sense view that our experiences cause all sorts of physical changes in the world. That is to say, more or less, that phenomenal states are necessary parts of conditions that causally suffice to make physical things happen. Suppose, for example, that a normal person is sitting in a dentist's chair, with no anesthetic. The dentist drills in such a way as to cause great pain. The pain, together with the connections between the brain and the muscles and such, and a lot of other stuff, causally suffice for the patient to scream, try to grab the drill, and curse the dentist as best he can with the dentist's drill and the dentist's hands in his mouth. The pain seems to cause the screaming and grasping. It seems if there were no pain, as there would not be had the dentist administered anesthetic, the screaming and grasping would not have occurred.

If we imagine a world without pains, there will be some physical changes as a result. There won't be screaming and grasping, even when there is no anesthetic, because we will have removed a cause. It seems then, that if experiences cause physical things, a world without experiences will not be physically indiscernible from one just like it. So it seems like a Zombie world isn't quite so easy to conceive, to consistently and carefully imagine, as we might have thought. If we give up a bit of common sense and suppose that experiences don't cause anything, then we won't have a problem conceiving the Zombie world. Then we will be epiphenomenalists: mental states are causes, but not effects. But why should we give up that bit of common sense? Why should we accept epiphenomenalism, for the privilege of accepting the Zombie argument?

Chalmers' Zombie Argument is aimed at those who are already convinced that if experiences are physical, it is because the phenomenal properties at most *supervene* on physical properties. But then, Chalmers' argues, they either logically *supervene* or causally *supervene*. If the former, the Zombie world should be inconceivable, because there is a conceptual/logical connection between the

physical and phenomenal. If the latter, we have neo-dualism. But this overlooks the case of identity. One could say that identity is not supervenience, or one could say that it is the limiting case. Either way, the identity of physical states with experiences requires neither that that our phenomenal concepts be conceptually reducible to the physical, nor that the phenomenal states be additional, non-physical effects of our physical states.

THE KNOWLEDGE ARGUMENT,

The Knowledge Argument, due mainly to Jackson, asks us to imagine Mary, raised with no color experiences in a black and white room---a Jackson Room, as I call it. Let Mary have as much knowledge about the physical world as you want. Still, when she steps out of her Jackson Room, Mary will have new knowledge: what it is like to see red (for example). Since she can know any physical fact while in the room, but cannot know what it is like to see red, that latter must be a non-physical fact.

I agree that Mary acquires new knowledge when she steps out of the Jackson Room. This can be seen by the fact that her belief-states change in a way that changes the truth-conditions of those states. There are propositions that must be true for her post-Jackson Room beliefs to be true, that did not have to be true for the beliefs she had in the Jackson Room to be true. Although she may not explicitly formulate these truths in language, she is epistemically committed to these propositions. The truth of her explicit beliefs, as she mentally formulates them, requires these propositions to be true. But the truth of these further propositions does not require anything non-physical.

Suppose that Mary acquired an idea of the experience of seeing red while she was in the Jackson Room. She knows from her reading that there are colors and color experiences, and knows that various sorts of light entering the eye cause them. She knows what it is like to see black things and white things. She doesn't know what it is like to see red things, but she still has an idea, "by description" of this state.

Imagine that Mary, before entering the real outside world, is allowed in what I call the Nida-Rumelin Room⁵, where there are colors on the wall, but no objects, like tomatoes, apples and the like, that would allow her to figure out, on the basis of knowledge gained in the black and white room, which of the colors was red. In the Nida-Rumelin Room, Mary can form a second idea of what it's like to see red---that is, an idea that will *in fact* be of that state, although she won't know it. She sees a color on the wall, which in fact is red, that is so striking she gives it a name, "wow".

In the Nida-Rumelin room Mary has two ideas in her mind of the color red. One is associated with a name and some worldly knowledge about which things have the color, but is not associated with a kind of experience. The other idea is associated with a kind of experience, but not with the word "red", nor with apples, tomatoes, fireplugs and the like. The two ideas are not connected in her mind. So the truth of her system of ideas and beliefs, considered in abstraction from what these ideas stand for, doesn't require that the two ideas stand for the same thing. When Mary steps into the ordinary world, however, she sees fireplugs, apples and tomatoes, which she knows to be red. In seeing them she has the experience of wow. The two ideas, her old idea of red, which was associated with the name "red," fireplugs, tomatoes, and the like, but not with any experiences of seeing red, and the new one, associated with experiences of red, but not with anything else except her made-up name for it, "wow," become connected in her mind. She knows that red is wow, and that what it is like to see red is what it is like to see wow.

When these ideas become connected, the truth-conditions of her beliefs change. Now, for her beliefs to be true, the two ideas must refer to the same color, whereas before that wasn't so. That is, this is a condition of truth on her beliefs *even if we abstract from facts about what the ideas stand for*. As I explained

⁵ Martine Nida-Rümelin, What Mary couldn't know: Belief about phenomenal states. In Thomas Metzinger, editor, *Conscious Experience* (Schöningh: Imprint Academic): 219-241.

above, this is what I call a *reflexive* content of her belief, since it is a condition of truth on the belief *itself* rather than on the things the ideas in it stand for.

There is no reason that the physicalist should deny that it is possible to have two ideas of the same color, or that there is a kind of idea that only occurs when one has the appropriate sensation, or a memory of what it is like to have it, that won't normally occur without having had the experience, and is quite unlike other ways of thinking of the experience. So there is no reason for the physicalist to deny that there is new knowledge. The new proposition that is known is that the two ideas are about the same color.

To repeat, this new proposition is what I call a *reflexive* content of the new belief, formed by connecting the two ideas. The truth of Mary's new belief requires that the ideas involved in the belief *itself* be about the same color. The proposition is about Mary's ideas, not the colors they are of. Mary's new knowledge eliminates the possibility that the state she is in, and to which she is attending, or remembering, is anything other than the state of having the sensation of red.

We usually describe beliefs in terms of their *subject matter* contents: the requirement the truth of the belief puts on the things the ideas in the belief are of. However, in some cases, as with the Clemens-Twain case above, it is not possible to get at how our beliefs change in this way. These are cases in which one connects two ideas that are in fact about the same thing. Reflexive contents allow us to isolate the facts in question. It allows us to get at what possibilities the new knowledge eliminates.

Given the distinction between subject matter contents and reflexive content, one can attack the knowledge argument in two ways, depending on what is required for Mary to know all the physical facts about color and color vision when she is in the black and white room. If this requires her to know all of the subject matter propositions about colors and the like that would be contained in ideal textbooks of the subject, then the physicalist must allow that this could be so. But this does not preclude her from coming to know the same subject matter

facts via different, experience-based ideas, and hence coming to have new reflexive knowledge. The physicalist need not suppose it possible for Mary to have any experience-based idea of red before she leaves the room, and so need not admit that Mary could know all the physical facts about color and color vision in the sense of having both subject matter and reflexive knowledge.

THE MODAL ARGUMENT

The Modal Argument is due to Kripke; Chalmers has a somewhat different version of it. I'll consider Kripke's version and my treatment of it in this Précis, and discuss Chalmers' version in my reply to his comments.

Consider an identity statement of the sort that might be true, if physicalism were true:

$$A = B$$

where A is a word like "pain" and B is some scientific description of a brain-state. Terms like "pain" are names of states; as such they are rigid designators; they stand for the same thing relative to every possible world; or, as one might say, they contribute the thing to which they refer to the propositions they express.⁶ We assume that scientific terms like B get at the states to which they refer by essential properties. They are also rigid designators, referring to the same thing in each possible world. Thus both A and B, if they refer to the same thing in the actual world, refer to the same thing in all possible worlds. So our identity statement, if true, is necessarily true.

How then are we to explain that the statement is not a priori? Kripke⁷ points out that it would be a mistake to consider this case analogous to something like

Water is H₂O.

⁶ I take "rigid designator" to mean "has the same referent relative to each possible world," and "directly referential" to mean "contributes the object designated, rather than some identifying condition of it, to the proposition expressed by statements of which it is a part." Ordinarily understood, an expression like "7 + 5" would be a rigid designator, but not directly referential.

⁷Saul Kripke. *Naming and Necessity* (Cambridge: Harvard University Press).

Water is something whose essential properties --- that is, its chemical makeup --- are not disclosed to us in our ordinary experience with it. We know it as the colorless liquid found in lakes and rivers that we drink. It is easy to mistake the contingency of

The colorless stuff we find in lakes and rivers and drink is H₂O with the contingency of "Water is H₂O". This explains the illusion that the latter is contingent.

But pain is something whose nature is disclosed to us, in our ordinary experience with it. We know what pain is like, and that's what pain is. Pain isn't known to us as the stuff that feels like *this* in our world; rather, it's just *this* that is the pain. So there is no ordinary conception of pain or stereotype of pain, whose contingent connection with the brain state would explain the illusion of contingency. So the safest bet is that the contingency is not an illusion. But if there is contingency, then there is no necessity, and then there is no identity.

The same account of concepts that was needed for the knowledge argument allows us to provide an alternative account of the contingency, however. Consider my experience-based concept of pain---the sort that people who have had pains and remember them have. At the core of this concept are my memories of pain, acquired from earlier experiences of pain ---my Humean ideas of pain, that in some uncanny way resemble the sensation itself. There is a lot more to this concept, including the word "pain," and lots of beliefs about what causes pain of various types, what one can do to relieve pain, and so forth. But at the core are these memories of experiences of pain.

If the word "pain" is associated with a concept like that, it will be a rigid designator, designating the state that gave rise to the concept, the state we were *in*, the state of pain, whatever contingent facts about that state, or misinformation about it, for that matter, are wrapped up in the concept.

Indeed, one can have not merely false beliefs, but even necessarily false beliefs wrapped up in such concepts. It may be helpful to look at a variety of examples.

First, an example involving smells. For most of us, our memories of smells are much less distinctive, and more easily confused one with the other, than our ideas of colors, or of different kinds of severe pains. Suppose over the years of a spice-deprived life, I have come to think that cinnamon and nutmeg are the same spice, and the smell of cinnamon and nutmeg are the same, and the sensation of smelling cinnamon and the sensation of smelling nutmeg are the same. I experienced both in childhood but not since. I have not had another chance to smell them. When I am given a chance to experience these smells again, I at once recognize my mistake. The new experiences enrich the old ideas; I know at a sniff, which is nutmeg and which is cinnamon, and how different they are. But before that I conceived that they were the same --- the spices, the smells, and the sensations.

Don't worry about the words "cinnamon" and "nutmeg" or even about cinnamon and nutmeg. Think about the concepts I had of the sensations involved in smelling each. I used to smell all the spices on my mother's spice rack. I liked the experience of smelling them a number of them. I longed to experience it again, but sniffing spices seemed undignified for a professor. But I had come to think these were the same---maybe different names had been given to the same spice by different manufacturers. My concepts of the sensations are rigid designators. They are of the smells that gave rise to the memories I have of them. So, given what my concepts are concepts *of*, there is no possible world in which my belief is true. The sensation of smelling cinnamon simply isn't the sensation of smelling nutmeg. What then was the content of my belief? Why did I abandon the belief when I smelled the spices? What worlds was I excluding as unreal, if there were no worlds in which my belief would have been true?

The worlds in question are the ones that we find when we consider reflexive contents of my beliefs. My belief that the sensation of smelling nutmeg

was the same as the sensation of smelling cinnamon was a structure in my mind involving my memories of nutmeg and of cinnamon. For the belief to be true, these ideas would have had to derive from the same experience. This is the possibility that is eliminated when I have a chance to sniff the spices again.

This is the basic idea of my reply to the modal argument. We can find the contingent possibilities that explain our sense that a given sensation need not be identical with any given brain state at the level of reflexive content. So we need not accept that the only explanation of sense of contingency is that the states are in fact not identical.

Suppose that in fact pain is perfectly correlated with brain state B_3 , so that a physicalist will say that pain just is B_3 . But surely the physicalist must admit that it was not determinable *a priori* that Pain = B_3 . Other possibilities could not have been excluded, and this one settled on, by mere reflection on our concepts and words.

The reason is that our physicalist, scientific concept B_3 , and our experiential concept of pain are so different. There is nothing in the concepts themselves that allows us to figure out that they are concepts of the same state. This is so, even though both concepts are rigid designators, and even though our experiential concept of pain is direct, in that we do not experience pain by experiencing some other sensation *caused* by pain; pain does not have *appearances* from which we infer it as the most likely or typical cause.

In such a case, as with the cases above, we find the possibilities that discovery eliminates, and the sense of contingency at the reflexive level.

This is the basic answer given in *KPC* to the modal argument. The chapter goes into a number of issues to deal with various turns of argument, and to deal with Chalmers' version of the argument, but in so far as these issues are relevant, I'll consider them below, in the Replies section.