Forthcoming In Byrne and Thomson (eds.)
Content and Modality: Themes from the Philosophy of Robert Stalnaker
Oxford University Press

Actors and Zombies*

Daniel Stoljar

1. Much of contemporary philosophy of mind is dominated by the intersection of three topics: physicalism, the conceivability argument, and the necessary a posteriori. I will be concerned here (i) to describe (what I take to be) the consensus view of these topics; (ii) to explain why I think this account is mistaken; and (iii) to briefly sketch an alternative.

2. The first of our trio, physicalism, is the thesis that, not necessarily but as a matter of fact, everything is physical. This thesis stands in need of clarification. For one thing, we need to be told what it is to be physical. This is a difficult and somewhat neglected question, but I want to set it aside. A rough and ready understanding will do for present purposes. Another aspect of the thesis requiring clarification is the sense in which it pertains to everything. There are a number of proposals about how to explain this, but here it is sufficient to identify physicalism as the thesis that the physical truths entail (in the sense of necessitate) *all* the truths, and so all the psychological truths. If this is physicalism, and if it is true, then it is contingent, i.e. true not necessarily but as a matter of fact. For it is contingent *which* truths are the physical and psychological truths at any given world. If the physical truths concern only extension in space and time, and the psychological truths concern ectoplasm, the physical truths will not entail the psychological. On the other hand, if the physical truths are as multifarious and complex as those that (we assume) obtain in our world, and if the psychological truths concern experiences more or less as they are construed by folk psychology, physicalism might be true.

Physicalism is if true contingent, but there is nevertheless a necessary truth lurking in the shadows that is important for our purposes to bring out. For suppose that

* Acknowledgments:

as a matter of fact physicalism is true, and thus the physical truths do entail the psychological truths. Then there must be a statement S which summarizes the complete physical truths including the physical laws and principles that obtain in our world, i.e. the truths that in fact obtain; likewise there must be a statement S* which summarizes all the psychological truths. Now consider the truth-functional conditional formed from these, 'if S then S*' and call this 'the psychophysical conditional'. If physicalism is true, this conditional is necessarily true. The reason is that S necessitates S* and it is *not* contingent which truths S summarizes even if it is contingent that the truths it summarizes are the complete physical truths of our world; mutatis mutandis for 'S*.' More generally, if physicalism is true, and if physicalism is the thesis that the physical truths entail the psychological truths, the psychophysical conditional is necessary.

3. Our second topic is the conceivability argument. The first premise of this argument is that it is conceivable that the antecedent of the psychophysical conditional is true and the consequent false. The second premise is that if this is conceivable then it is genuinely (i.e. metaphysically) possible. However, if there is a genuinely possible situation in which the antecedent of the psychophysical conditional is false, that conditional is, at best, contingent. But as we have seen, if physicalism is true, the conditional is necessary. Conclusion: if the premises of the conceivability argument are true, physicalism is false.

Why is it conceivable that there is a situation in which the antecedent of the psychophysical conditional is true and the consequent false? The usual way to develop this point is to consider the idea of a zombie, where, as Robert Stalnaker (2002, p.239) puts it, zombies are "creatures that are physically exactly like ordinary people, but have no phenomenal consciousness. A zombie world is a world physically exactly like ours, but with no phenomenal consciousness at all. The sun shines in such worlds, but the lights are out in the minds of the unfortunate creatures who live in them." The idea of a zombie in turn prompts a particular implementation of the conceivability argument. As Stalnaker (p.239) says: "…it is conceivable, or conceptually possible, that there be zombies. From this it is inferred that zombies, and zombie worlds, are metaphysically possible," and from this in turn it is inferred that physicalism is false.

The idea of a zombie makes the conceivability argument less abstract than it might otherwise be, but it also raises problems. Sydney Shoemaker (e.g., 1999), for example, argues that the idea of a zombie is incoherent, and so not conceivable, in view of the fact that there are constitutive connections between experience and beliefs about experience. What Shoemaker says may well be right, but it would be mistaken to go on to suppose (and in fact Shoemaker does not suppose) that considerations of this sort will undermine the conceivability argument. For these considerations attack at best an example. They do not attack the underlying argument. In this respect, the situation is akin to Putnam's famous (1981) attack on skepticism, in which it is argued that the causal theory of reference undermines various brain-in-a-vat examples. What Putnam says might (*might*) be right, but it will not undermine skepticism *tout court*, for the skeptic may mount his argument on the basis of a different example (cf. Campbell 2002). The same point applies to those suggestions that emphasize the constitutive connections between experience and belief.

There is also a more general concern about the conceivability argument, whatever precisely the example is that lies in the background. This is that the notions in terms of which it is stated are notoriously unclear. The concern is serious, but I doubt those who discuss the conceivability argument against physicalism are under any special obligation to allay it; and indeed this fact will be important in what follows. For the conceivability argument we are concerned with is in important respects analogous to arguments that are used and accepted throughout philosophy, and in philosophy of mind in particular. For example, consider a very different argument of Putnam's (1965): the perfect actor objection to (philosophical) behaviorism. Perfect actors are people that behave actually and potentially exactly like ordinary people but have quite different phenomenal states. It seems conceivable, and so possible, that there are such people. And, if this is possible, behaviorism is false, for behaviorism entails that behavioral truths entail the psychological truths. It is standard practice in philosophy of mind to assume that this sort of argument is successful—a standard practice I assume is perfectly legitimate. But it is bad form to use a method of argument against theories you don't like, and then turn hypercritical when the same method is deployed against theories you do.

4. Turning now to our third topic, a truth is a priori—to put it roughly—just in case (fully) understanding it is sufficient for knowing that it is true; and a truth is necessary just in case it is true in all possible worlds. Traditionally, it was assumed that these two features are co-extensive: all and only priori truths are necessary. But what Kripke (1980) and others showed is that it is possible to have a truth that is both necessary and a posteriori. (It was also argued, more controversially, that it is possible to have truths that are contingent and a priori; but we will set aside this idea in what follows.) One of Kripke's examples is the identity statement 'heat = molecular motion.' This statement, he says, is true at all possible worlds (or at any rate is true at all possible worlds at which heat exists); and yet it is also a posteriori in the sense that mere understanding it does not entail knowing that it is true. Of course, every example is controversial in some sense, and this one is no different. But it simplifies matters greatly if we assume in what follows that Kripke is right on this point and that 'heat = molecular motion' is a necessary a posteriori truth. At any rate, that will be my procedure.

5. So far we have introduced our three topics; it remains to introduce the consensus view about them. The consensus view has two parts. The first points to the possibility of a version of physicalism I will call *a posteriori physicalism*. We have seen that if physicalism is true, the psychophysical conditional is necessary. But now let us ask: is the psychophysical conditional a priori or a posteriori? The answer to this is not determined by any assumption we have made so far. Physicalism itself is contingent, and presumably too it is a posteriori. But it does not follow that if physicalism is true, the psychophysical conditional is a posteriori. After all, the *modal* status of physicalism might diverge from that of the psychophysical conditional; why should the same not be true of its *epistemic* status? On the other hand, while our assumptions do not entail anything about the epistemic status of the conditional, they *do* make salient the possibility that it is a necessary a posteriori truth, and as such exhibits the same combination of modal and epistemic features that is exhibited by statements such as 'heat is motion of molecules.' Those who assert that this is the case are a posteriori physicalists; those who assert this is not, i.e. that the psychophysical conditional is a priori, are *a priori physicalists*.

So the first part of the consensus view is a posteriori physicalism; the second is the suggestion that the a posteriori physicalist *is*, while the a priori physicalist *is not*, in a position to answer the conceivability argument. The claim here is not simply that if a posteriori physicalism is true, the argument can be answered *somehow*. That point is obvious; the conceivability argument is an argument *against* physicalism, so the truth of physicalism entails it can be answered somehow. The claim of the consensus view is rather that, in explaining how exactly the argument goes wrong (assuming it does) one must draw on the distinctive claim of a posteriori physicalism, i.e. the claim that the psychophysical conditional is a posteriori. Of course different proponents of the consensus view may have different views about *just how* to respond to the argument in the light of this claim. But what is distinctive of the view, or at least the second part of the view, is the assertion that it is uniquely the a posteriori physicalist who can answer the argument. Turning this around, what is distinctive of the view is that physicalists must meet the conceivability argument by becoming a posteriori physicalists.

6. Is the consensus view correct? I don't think so. I don't disagree with the first part of the view, i.e., the claim that the psychophysical conditional is a posteriori. But I disagree that this fact, if it is a fact, bears on the conceivability argument. So my disagreement is with the second part.

Since my criticism focuses on the second part of the consensus view, it is different from a well-known criticism of the view that focuses on the first, due mainly to Jackson (1998) and Chalmers (1996). This criticism says that it is mistaken to suppose that the psychophysical conditional is a posteriori in the first place, or at any rate it is mistaken to suppose this if physicalism is true. According to proponents of this criticism, there are premises in philosophy of language (and perhaps epistemology) from which it follows that if physicalism is true, the psychophysical conditional is (not merely necessary but) a priori. Clearly in this case the question of what to say about the second part of the consensus view is moot. If the psychophysical conditional is *not* a posteriori, it cannot be this fact about the conditional that answers the conceivability argument. On the other hand, if the psychophysical conditional *is* a posteriori, or it if is an open

question whether it is, the question about whether this would answer the conceivability argument is likewise open.

The problem with the well-known criticism is that the premises from philosophy of language (and perhaps epistemology) from which it proceeds are extremely controversial. What is at issue here is what Stalnaker has called in a number of places (e.g. 2002 p 208) 'the generalized Kaplan paradigm.' Stalnaker himself rejects the generalized Kaplan paradigm; others defend it. My own view is that the matter is unclear. Take the highly complicated, nuanced and sophisticated version of the description theory advanced by, for example, Jackson (1998)—this is one version of what Stalnaker means by the generalized Kaplan paradigm; and now take the highly complicated, nuanced and sophisticated version of the anti-description theory advanced by, for example, Stalnaker. How is one to decide between them? I don't deny the issue might in principle be settled; it is rather that I myself don't see any clear way to settle it. So I will not engage this issue in what follows. Rather I will assume, as against the generalized Kaplan paradigm, that a posteriori physicalism is possible, and in fact is true. I think the consensus view is mistaken even given that assumption.

I have said that I want to set aside the well-known criticism. But it bears emphasis that the debate surrounding this criticism contributes greatly to the consensus position being the consensus position, and in fact this is my excuse for using the label. The reason is that this debate encourages the thought that *if* the psychophysical conditional were a posteriori, this *would* have a major impact on the conceivability argument. In fact, both sides in the debate about the first part of the consensus view seem to proceed under the assumption that this last conditional claim is true. So, while many philosophers think outright that the a posteriori nature of the psychophysical conditional answers the conceivability argument, many more philosophers agree that it *would* answer the argument *if* it were a posteriori.

7. I have distinguished two parts of the consensus view and said I have no quarrel with the first. What then is my quarrel with the second? The basic criticism can be stated very simply. The conceivability argument is an argument to the conclusion that the psychophysical conditional is not necessary. But, shorn to essentials, the response on

behalf of the a posteriori physicalist is this: the psychophysical conditional is necessary *and* a posteriori. But now I ask you to forget your prejudices and look afresh at this answer. How can this alone possibly constitute a persuasive response? In general, if I have an argument from a set of premises Q1, Q2…QN, to a conclusion P, it is not a persuasive response to me to simply assert not-P. How then can it possibility be a persuasive response to me to assert not-P *and* R for any R apparently unrelated to the premises? The assertion that the psychophysical conditional is necessary *and* a posteriori is on its face no more of a response to the conceivability argument than the outright assertion that the conditional is necessary; or that it is necessary and is really very interesting; or that it is necessary and by the way that's a lovely shirt you're wearing. On the face of it, the consensus view is a spectacular non-sequitur.

Perhaps this way of putting matters makes my criticism of the consensus view sound a bit sophistical. So let me put things slightly differently. Nowhere in the conceivability argument is there any explicit mention of the a posteriori. Strictly and literally what we have been told is something about conceivability, and then something else about possibility. So it is a mystery—at least it is a mystery to *me*—how the notion of the a posteriori is supposed to enter the picture. At the very least we require a story in which the connection of the a posteriori to conceivability is explained. Unless a story is produced, we have no answer here to the conceivability argument.

8. No doubt proponents of the consensus view are at this point bursting to tell me the story. I will consider some proposals in a minute. But first I want to point out that the criticism I have just made of a posteriori physicalism—that, at least on the surface, it does not answer the conceivability argument—is closely related to a similar point made by Kripke in *Naming and Necessity*, at any rate as I understand him. (In what follows I will state Kripke's point in my own terms rather than his.)

The way in which the matter comes up for Kripke is via a comparison of a conceivability argument about experiences (his example is pain) with a conceivability argument about secondary qualities (his example is heat). We have seen that zombies are people who are physically just like us but who lack phenomenal consciousness. But imagine now a type of physical object physically just like the pokers that exist in our

world but which uniformly lack heat; call them zpokers. Offhand, it looks conceivable, and so possible, that there be zpokers. But then heat, or at any rate heat in pokers, must be something over and above the physical, i.e. must be something over and above motion of molecules. In short, there is a conceivability argument about heat—call it CA (heat)—that parallels the one we have been considering—call it CA (pain).

Now the line of thought suggested by the comparison between CA (pain) and CA (heat) may be summarized as follows. First premise: CA (heat) is unsound—after all, we *know*, or at any rate have assumed, that 'heat is motion of molecules' is necessary and a posteriori; so an argument to the conclusion that it is not necessary must be mistaken. Second premise: CA (heat) is analogous to CA (pain). Conclusion: CA (pain) is unsound too. Moreover, the reason that this line of thought is important for us is that it naturally suggests that the second part of the consensus view is true. After all, the most salient philosophical fact about CA (heat) is that it involves a necessary a posteriori truth, i.e. 'heat is motion of molecules.' Moreover, it is natural to assume that it is *this* fact that explains the failure of CA (heat). More generally, if we arrange things so that the connection between pain and the physical is in all respects like the connection between heat and the physical, we would have an answer to the conceivability argument against physicalism; in short, the second part of the consensus view is true.

But, as is of course well known, Kripke rejects this line of thought, on the ground that there is no relevant analogy between the two arguments. In the case of heat, we may distinguish heat itself from sensations thereof. And this distinction permits us to deny that it is conceivable there be zpokers. What is conceivable instead is that there be pokers that produce no sensations of heat; but this is a different matter. In the case of pain, however, there is no distinction between pain and sensations of pain. At least in the intended sense, pain just is a sensation of pain, and thus there is no possibility of producing a response to the argument that turns on a 'distinction' between them; there is none. (To be sure, there may be another sense in which pain is something in your toe. But this does not affect the substance of the issue. Kripke could have made his point by contrasting heat and sensations of heat directly.)

What is the relation between Kripke's discussion and our own? Well, we started from the question: what is the connection between the fact (assuming it to be fact) that

the psychophysical conditional is necessary and a posteriori, on the one hand, and the conceivability argument on the other? We also noted that it is at least unobvious how this question is to be answered. Kripke's discussion can be usefully thought of as starting in the same place. It is just that he goes on to consider and dismiss a suggestion about how the connection might be explained. In short, Kripke's discussion is further evidence that our basic criticism of the consensus view is correct.

9. Unless there is some way to connect a posteriority with the conceivability argument, the consensus view is a non-sequitur. Kripke in effect discusses one way in which this connection might be explained, but the suggestion runs aground on the difference between heat and pain. But of course, even if this *particular* suggestion is unsuccessful, it scarcely follows that *nothing* similar is. So I want next to examine a related suggestion due to Stalnaker. Stalnaker makes the suggestion I want to focus on through the voice of a character he calls Anne; but I will take the liberty in what follows of assuming that the position is his. Of course, whoever in fact holds the position, it is important and needs to be discussed.

10. Stalnaker begins by considering a philosopher Thales who asserts that water is, not a compound like H20, but some sort of basic element. Stalnaker himself refers to this element, following Putnam, as 'XYZ', but I will call it 'Thalium.' Surely it is an empirical fact that the stuff we call 'water', the stuff we use to fill bathtubs and water the garden, is H20 rather than Thalium. Similarly, surely it is an empirical fact that we live in an H20 world rather than a Thalium world. This suggests that, properly understood, the word 'water' is, as Stalnaker puts it, "theoretically innocent" (p247). In using it, we refer to something, but we don't prejudice its nature. To put the point slightly differently, the fact that 'water' refers to H20 is to be explained, not merely by the way in which we use the word, but by the way in which we are embedded in our environment. If we lived in a world that Thales thinks is the actual world, and we used the word rather as we use it actually, our word would in that case have referred to Thalium.

Now just as it is an empirical fact that we live in the H20 world rather than the Thalium world, Stalnaker says, it is an empirical fact that we live in a materialist world

rather than a dualist world. And this suggests that properly understood words such as 'experience', 'pain' and so on are theoretically innocent too. In using them, we refer to something without prejudicing its nature. The fact—assuming it to be a fact—that 'pain' refers to some neural or physical condition is to be explained, not merely by the way in which we use the word, but the way in which we are embedded in our environment. If we lived in a world that the dualist thinks is the actual world, we would use the word rather as we use it actually, but our word 'pain would in that case have referred to a non-physical property.

These considerations prompt an account of what has gone wrong in the conceivability argument that is different from, but related to, the suggestion considered by Kripke. In effect, the suggestion considered by Kripke was that, in advancing a conceivability argument about heat, we are confusing the conceivability of (1) with that of (2):

(1) There is molecular motion in the poker but no heat in the poker.

(2) There is molecular motion in the poker but nothing in it causing heat sensations.

Or, if a similar argument were to be advanced by Thales against the hypothesis that water is H20—call such an argument CA (water)—the suggestion considered by Kripke would be that Thales is confusing the conceivability of (3) with that of (4):

(3) There is a H20 in the bathtub but no water in the bathtub.

(4) There is a H20 in the bathtub but nothing in it causing perceptions as of water.

However, Kripke went on to say, these points are no help at all in the case of the conceivability argument against physicalism, i.e., CA (pain). For here the parallel suggestion would be that a proponent of the argument is confusing the conceivability of (5) with that of (6):

(5) There are people physically like us but which lack pain.

(6) There are people physically like us but which lack states that cause sensations of pain.

And this parallel suggestion fails, Kripke argues, since there is no way to make sense of the idea that (5) has been confused with (6).

Stalnaker's alternate proposal is that in mounting CA (water), Thales is confusing (3) not with (4) but with:

(7) In a Thalium world considered as actual, there is H20 in the bathtub but no water in the bathtub.

Moreover, this point *does* have application to CA (pain). For it is now available to us to say similarly that here we are confusing (5) not with (6) but with:

(8) In a dualist world considered as actual, there are c-fibres firing in me but I am not in pain.

The phrase 'world considered as actual' is due to an important paper by Davies and Humberstone 1982, and has a technical meaning within two-dimensional modal logic, a topic to which Stalnaker has made seminal contributions. The details of these ideas are difficult, but I think there is no harm in the present context to interpret what is intended as follows:

(7*) There is H20 in the bathtub and there is no water-as-Thales-understands-water in the bathtub (i.e., there is no Thalium in the bathtub).

(8*) There are c-fibers firing in me and I am in not in pain-as-the-dualist-understands-pain.

On this interpretation, Stalnaker's proposal is that in CA (water) we confuse (3) with (7*) and CA (pain) we confuse (5) with (8*). And the significance of this suggestion is

that both (7*) and (8*) is in the context unobjectionable. It is not impossible that what Thales says is true, so it is not impossible that there is Thalium in the bathtub. But this does not undermine the hypothesis that water is H20. Similarly, it is not impossible that what the dualist says is true, so it is not impossible that there are c-fibers firing in me and I am not in pain-as-the-dualist-understands-it. But this does not undermine the hypothesis that physicalism is true.

11. Stalnaker's suggestion is ingenious, but I have two objections. To see the first, consider again the perfect actor argument against behaviourism. We have seen that this argument proceeds from the premises, first, that it is conceivable that there are perfect actors, i.e., people psychologically distinct from us but behaviourally identical, and second, that what is conceivable is possible. The conclusion of the argument is that behaviourism is false, for behaviourism entails that behavioural truths entail the psychological truths. As I have said, I take it to be quite obvious that this argument is successful, and that what we have here is a good argument against behaviourism.

But unfortunately Stalnaker is in no position to say this. For there is no reason at all why the behaviorist might not respond to these arguments in precisely the way that he recommends we respond to the conceivability argument. In particular, there is nothing in Stalnaker's account to prevent a behaviorist from responding as follows. "The perfect actor argument fails because it confuses pain with pain-as-the-anti-behaviorist-understands-it. Everyone agrees that pain understood *that* way could come apart from behavior, but if you assume that you have begged the question against me. The question is whether pain as we ordinarily conceive of it can come apart from behavior, and this the argument does not show." I take it that there is something seriously wrong with the idea that a behaviorist might respond to the perfect actor argument in this way, and so there is likewise something seriously wrong with Stalnaker's proposal.

One might reply by pointing out that there are many *other* reasons to resist behaviorism—empirical reasons, say. True enough, but irrelevant: I am not denying that there might be other arguments against behaviorism; of course there are. Nor am I saying that Stalnaker's position commits him to behaviorism; of course it doesn't. What I am saying is that Stalnaker's response to the conceivability argument has the bad

consequence that a good argument against behaviorism turns out to be a bad argument. His response provides the materials to respond to Putnam's perfect actor argument; but since we *know* that the latter argument is a good one, there must be something mistaken about his response.

Alternatively, one might reply by gritting one's teeth. Stalnaker has prescribed a drug to rid us of the conceivability argument. The drug has a side effect, but perhaps this is something we should learn to tolerate, a bad consequence outweighed by good. I think this response forgets just how plausible the perfect actor argument is as a refutation of behaviorism. In the standard philosophy of mind class you begin with dualism and show that it is implausible, and then you turn to behaviorism and shows that it is implausible, and then you move onto other things. But how did you persuade the students behaviorism is implausible? At least a large part of this case is provided by the perfect actor argument (and similar conceivability arguments such as Block's (1981) blockhead argument.) These arguments are completely compelling to undergraduates, and I think the reason for that is that they *are* completely compelling. So casting the consequence of Stalnaker's proposal that I have pointed out as tolerable is not an option.

12. In any case, there is a further reason why gritting one's teeth is no response to the problem about perfect actors. This is that it is plausible to suppose that the technique for defeating the conceivability argument that Stalnaker advances would defeat *any* conceivability argument at all, or at least any conceivability argument of the sort we are considering.

To illustrate, take any two distinct truths A and B. Suppose someone argues that it is conceivable that A is true and B is not, and concludes that it is possible that A is true and B is not, and that in consequence the truth of B is something 'over and above' the truth of A. Someone who adopted Stalnaker's strategy as I understand it (and put in schematic form) might respond as follows: "Distinguish B from B-as-understood-as-over-and-above-A; for short, distinguish B from over-and-above-B. When you claim that it is conceivable that A is true and B is not, all that is genuinely conceivable is that A is true and over-and-above-B is not. But from this nothing follows: everyone agrees that it is possible that A is true and over-and-above B is not." The problem for Stalnaker is

that, if this strategy worked, it could be used against any conceivability argument of this form. So either no conceivability argument like this is sound, or the strategy is unsound. I assume that some conceivability arguments are sound; for example I assume that the perfect actor argument is sound. So the strategy is mistaken.

Stalnaker's defense of the a posteriori physicalism runs into a problem that in my view is endemic to many contemporary attempts to respond to the conceivability argument: it overgenerates. As we have noted, the conceivability argument against physicalism is in structure identical to arguments that are used throughout philosophy. This fact suggests the following condition of adequacy on any candidate response to that argument: if you think you have isolated a factor that constitutes the mistake in the conceivability argument against physicalism, check to see if that factor is present in parallel arguments you accept; if so, consign your proposal to the flames. The problem for Stalnaker, I am suggesting, is that his proposal fails to meet this condition of adequacy.

13. I said earlier that I had two objections to Stalnaker's account. The first, which we have just been discussing, is that it mistakenly gives the behaviorist the materials to respond to Putnam's perfect actor objection, a point that generalizes to other conceivability arguments as well. The second is that what Stalnaker says has nothing to do with the epistemic status of the psychophysical conditional. For suppose—perhaps impossible—that a priori physicalism is right and the psychophysical conditional is necessary and a priori. Of course a priori physicalists face the conceivability argument too. How are they to respond? There is nothing to prevent them from arguing as Stalnaker does, or—what I assume to be the same thing—as his proxy Anne does. Anne is a B-type materialist, or what I am calling here an a posteriori physicalist, and what she says about the conceivability argument she is perfectly entitled to say. Still, there is no reason why an A-type or a priori physicalist might not say the same. In fact some a priori physicalists *do* say the same or at least very similar things. One is David Braddon Mitchell (see 2003; see also Hawthorne 1997.)

So Stalnaker's strategy for responding to the conceivability argument is open to both versions of physicalism. How serious is this as a criticism of the strategy? In one

sense it's not serious at all. Its availability to both positions does not render the view implausible. Indeed, in this respect, Stalnaker's proposal is similar to the one discussed in connection with Kripke. Kripke suggested that the way around the CA (heat) was to distinguish heat from heat sensations and then pointed out that such a response is unavailable to someone seeking a response to CA (pain). What Kripke says is plausible, but the epistemic status of physicalism plays no role in it. Suppose I were an a priori physicalist, not only about pain but about heat as well. I would *still* need an answer both the CA (heat) and to CA (pain). And if what Kripke says is right, I would have an answer to CA (heat) but would have no answer to CA (pain).

So in one sense it is no criticism of what Stalnaker says that it is available to both versions of physicalism. On the other hand, it is very natural, on reading of Anne's intervention into the debate about the conceivability argument, to suppose that it is somehow her being an *a posteriori* physicalist that permits her to make the response that she does. After all, the *only* thing we know about Anne is that she is an a posteriori physicalist: "Don't look for a real-world analogue for this character", we are told, "at least not one with this name" (284). If what I have been saying is right, Anne's being a certain kind of physicalist is irrelevant: her being an a posteriori physicalist is one thing, and her advancing the strategy she does is quite another.

Furthermore, the observation that Stalnaker's proposal is available to both the a priori and the a posteriori physicalist lends additional weight to our criticism to the consensus view. The a posteriori physicalist obviously has to say *something* to the proponent of the conceivability argument; *every* physicalist has to say something to the proponent of the conceivability argument. But when we look in detail at what Stalnaker suggests qua defender of a posteriori physicalism, we find that what is being said has nothing to do with the physicalism in question being of the a posteriori variety. So, contrary to the consensus view, the fact that distinguishes a posteriori physicalist from other sorts of physicalist is not the fact that answers the argument.

14. I have suggested that the claim that the psychophysical conditional is necessary and a posteriori by itself does nothing to answer the conceivability argument, and that two initially promising suggestions (one discussed by Kripke, one advanced by

Stalnaker) about how to develop a posteriori physicalism lead nowhere. At this point you might object that I have simply been being dense.

"Surely," you might say, "the construction 'it is conceivable that p' *just means* 'it is not a priori that not-p'. And, since 'it is not a priori that p' just means 'it is a posteriori that p,' the connection you are looking for is very short indeed. In particular 'it is conceivable that not-p' is logically equivalent to 'it is not a priori that not not-p', and by the definition of the a posteriori, and double negation elimination, this in turn is equivalent to 'it is a posteriori that not-p'."

I think, as against this, that there is no point denying that 'it is conceivable that p' has a reading according to which it means 'it is not a priori that not p'. The notion of conceivability can be legitimately spelled out in a number of different ways; this is one of those ways. But the idea that, in the *specific* context of the conceivability argument, this is what 'it is conceivable that p' means is quite another matter. When Putnam tells us about perfect actors, I don't think he means to be saying merely that it is not a priori false that there are perfect actors. I think he means to be saying that a certain case appears to be possible or (if this is different) is imaginable. On the other hand, talk of what seems to be possible or of what is imaginable is prima facie different from talk of what is or is not a priori.

It might be replied that while this is true prima facie, it is not true all things considered, and in particular, it is imaginable that p'—to focus on this notion for the moment—itself just means 'it is not a priori that not p'. However, I think this last equivalence is decidedly implausible (cf. Yablo 1993). For consider the negation of any of the standard examples of necessary a posteriori truth—say, 'water is not H20.' It is clear that it is a posteriori that water is not H20 and so of course it is not a priori that that water is not H20. Is it likewise imaginable that water is not H20? I think not. As Kripke argued, it is not at all clear that we can imagine water not being H20. Of course we can imagine related things. For example, we can imagine water not producing perceptions as of water; perhaps also we can imagine water that isn't Thalium. But none of this is strictly speaking imagining that water is not H20. More generally, therefore, the idea that 'it is imaginable that p' just means 'it is not a priori that not p' is open to counterexample. More generally still, this very shortest way to connect the notion of the necessity a

posteriori with the notion of conceivability, and so defend the consensus view, is implausible.

16. I noted earlier that, while the particular method that Kripke discusses for connecting the topic of the a posteriori with the topic of the conceivability argument breaks down, nothing we have said *proves* that no proposal along these lines could work. Obviously, it remains true that nothing has been proved. Still, I think our previous reflections make very plausible the hypothesis that there is in fact no connection here, and hence the second part of the consensus view is mistaken.

More generally, there would appear to be two topics: first, the necessary a posteriori and associated matters; second, what if anything has gone wrong in the conceivability argument and associated matters. The interesting suggestion of the consensus view is that these two topics are intimately connected. However, in light of what we have said, a more plausible view is that they are not connected in any obvious way.

Of course, that leaves us with at least two daunting projects. One is to fit the necessary a posteriori into a smooth picture of our thought and talk and the way in which that thought and talk relates to the world. This is something I have already indicated I will not do, for the simple reason I have no idea how. The other is to say something sensible about where and how the conceivability argument goes wrong (assuming it does). This too is a long story, but I think here I have something to say. I will devote the final sections of the paper to very short account of what this is.

17. Summarizing his interpretation of Kripke's achievement, Stalnaker (1997; 168) writes:

> The positive case for the theses that Kripke defends is not novel philosophical insight and argument, but naïve common sense. The philosophical work is done by diagnosing equivocations in the philosophical arguments for theses that conflict with naïve common sense, by making the distinctions that remove the obstacles to believing what it seems intuitively most natural to believe.

Viewed from a sufficiently high-level of abstraction, something similar is true in the case of the conceivability argument. There is a response to the conceivability argument that is intuitively very natural to believe. The case for this response is, if not naïve common sense, then at least scientifically and historically informed common sense. And the work in defending this response is mainly in identifying and undermining the philosophical reasons for dismissing or ignoring it.

What then is the response to the argument that is intuitively so natural to believe? The natural response is—wait for it!—that we are missing a piece of the puzzle; that is, we are ignorant of a type of truth or fact which (a) is either physical or entailed by the physical; and (b) is itself relevant to the nature of experience. To say this is not to say that we will remain forever ignorant of this type of truth, nor that our ignorance must concern basic physics—it may concern a fact that supervenes on basic physics but which is nevertheless not psychological. (Remember there are *many* such facts.) The positive case for this response can certainly be made, but it is largely consists in reminding ourselves of our epistemic position. It is an obvious empirical fact that we are ignorant of the nature of consciousness—there is no reason why a response to the conceivability argument may not draw on that fact along with anyone else. It is also true that historically we have been in similar situations before (cf. Stoljar Forthcoming). These facts provide good, but not demonstrative, evidence that this is our situation here too.

How does the hypothesis of ignorance answer the argument? Well, consider the claim there are people like us in all physical respects but who lack phenomenal consciousness. The phrase 'all physical respects' contains a quantifier, and so we may ask about its domain, and so about the interpretation of the central claim of the conceivability argument. Suppose the domain is construed broadly, so as to include absolutely all respects; in particular, so it includes respects relevant to experience but of which we are ignorant. (The hypothesis of ignorance in effect says there are such respects.) Then the conceivability claim would put pressure on physicalism, but it is doubtful that we can genuinely conceive of the relevant situation. How am I supposed to conceive various respects about which I have no knowledge? On the other hand, suppose the quantifier is construed narrowly, to include only those respects or types of respects of which are not ignorant. Then the conceivability claim is plausible, but it will

not put any pressure on physicalism. For the physicalist will be on good ground responding that the possibility claim at issue only seems possible because it is driven by a conceivability claim that does not take all relevant respects into account.

Not only does the proposal answer the argument, it does so in a way that speaks to the concerns that emerged in the course our previous discussion. For one thing, on this view, it remains an open question whether the psychophysical conditional is a priori or not, so in that sense we not being offered a version of the consensus view. But more important the account satisfies the condition of adequacy we formulated when thinking about Stalnaker's proposal. According to this condition of adequacy, any proposal about where the mistake is in the conceivability argument must be checked against conceivability arguments we accept. In particular, we should ask: does the epistemic response have the effect of granting to the behaviourist the materials to respond to the perfect actor objection? The answer to this question is 'no', and the reason is that there is a major discrepancy in the way in which a behaviourist appeals to behavioural truths, on the one hand, and the way in which the physicalist appeals to physical truths on the other. Behavioral truths are, and are intended to be by the behaviorist, truths that we can be established on the basis of direct perception: behavioral dispositions, or any rate their manifestations, are supposed to be available to perception. That was the basic rationale of the behaviorist program. And it is very plausible that no truth of that *sort* will be of any help in thinking about the perfect actor objection to behaviorism; a fortiori, not unknown truth of that sort will be of any help in thinking about behaviorism. On the other hand, physical truths meet no such epistemological condition; in fact, it is far from obvious that they meet any positive condition at all apart from being non-experiential. Hence there is room here for an ignorance-based or epistemic response to the conceivability argument.

18. So in briefest outline is the epistemic response to the conceivability argument. Why have so many missed it? No doubt part of the story is our tendency to discount our own ignorance. But another, and perhaps ultimately more interesting, reason derives from a powerful view of what philosophical problems are and what contributions to them should be.

A statement of the view I have in mind can be found in the famous passage from the *Investigations* in which Wittgenstein says:

> We must do away with all explanation, and description alone must take its place. And this description gets its power of illumination—i.e. it purpose—from the philosophical problems. These are, of course, not empirical problems; they are solved rather by looking into the workings of our language, and that in such a way as to make us recognize those workings: in despite of an urge to misunderstand them. The problems are solved, not by giving new information, but by arranging what we have always known. Philosophy is a battle against the bewitchment of our intelligence by means of language. (1954, p.47)

The most famous line in this passage is probably the last one, but for me the penultimate one is the most important. At least in philosophy of mind, this idea about philosophical problems is remarkably influential and persistent, much more influential and persistent than the Wittgensteinian apparatus within which it first appeared—so, at any rate, it seems to me. Frank Jackson (1998), to take one modern example, says of what he calls serious metaphysics is "discriminatory at the same time as being complete or complete with respect to some subject matter" (p.5). Similarly, John Perry (2001) describes his approach by saying that it "won't be physiological or neurological, nor even….very phenomenological. [It] will be logical, semantical and philosophical." (p.118). As I read it, the suggestion implicit in Perry's remark is that in an important sense all the relevant empirical facts are in; we just need a way to think through those facts. Wittgenstein, Jackson and Perry are remarkably different in other respects, but on this matter they speak with a single voice, or so it seems to me. All three are united in the idea that solving the problem presented by the conceivability argument does not involve any new information; it involves rather rethinking the information already in our possession. On the other hand, this idea precisely is in conflict with informed common sense, for, when confronting the conceivability argument, the view of informed common sense is precisely that new information is required.

19. I have my own views about how to respond to this conflict, but this is not the place to pursue them. I certainly don't mean in these sketchy remarks to recommend a blanket rejection of this account of what philosophical problems consist in. For one thing, it is quite clear that *some* philosophical problems *do* conform to this general description. But the idea that philosophical problems *as a class* do, and that the problems represented by the conceivability argument do in particular, seems to me to be something of a dogma. One consequence of the dropping the dogma is that a more particularist approach to philosophical problems comes into view—perhaps there is *nothing* much to say *in general* about what a philosophical problem is like. But another more immediate effect is the removal of one of the main impediments to informed common sense when it comes to the conceivability argument against physicalism.

References:

Braddon Mitchell, David 2003 'Qualia and Analytic Conditionals' *Journal of Philosophy* 100:111-35

Block, N. 1981. 'Psychologism and Behaviorism' *Philosophical Review,* 90, 5-43.

Campbell, J. 2002. ' Berkeley's Puzzle' In John Hawthorne and Tamar Szabó Gendler (eds.) 2002 *Conceivability and Possibility* Oxford University Press.

Chalmers, D. 1996. *The Conscious Mind* Oxford University Press

Davies, M and Humberstone, L. 1980. 'Two Notions of Necessity' *Philosophical Studies*, 38, pp. 1-30.

Hawthorne. J. 2002 'Advice for Physicalists' *Philosophical Studies*, 109, pp.17-52.

Jackson, F. 1998. *From Metaphysics to Ethics.* Oxford University Press.

Perry. J. 2001. *Knowledge Possibility and Consciousness*. MIT Press.

Putnam, H. 1965. 'Brains and Behaviour'. In R.J.Butler (ed). *Analytical Philosophy, Vol.2,* Blackwell.

Putnam, H. 1981. *Reason Truth and History.* Cambridge University Press

Shoemaker 1999. 'On David Chalmers' *The Conscious Mind'*, *Philosophy and Phenomenological Research*, 59, pp. 439-444.

Stalnaker 1997. 'Reference and Necessity'. In Crispin Wright and Bob Hale (eds.) *Blackwell Companion To the Philosophy of Language* Blackwell Reprinted in Stalnaker 2003b; references to the reprinted version.
Stalnaker 2002. What is it like to be zombie? In John Hawthorne and Tamar Szabó Gendler (eds.) 2002. *Conceivability and Possibility* Oxford University Press Reprinted in Stalnaker 2003b; references to the reprinted version.
Stalnaker 2003. 'Conceptual Truth and Metaphysical Necessity'. In Stalnaker 2003b.
Stalnaker, R. 2003b. *Ways A World Might Be: Metaphysical and Anti-Metaphysical Essays.* Oxford University Press
Stoljar, D. Forthcoming. 'Physicalism and Phenomenal Concepts' *Mind and Language*
Wittgenstein, L. 1954. *Philosophical Investigations*. Macmillan.
Yablo, S. 1993. 'Is Conceivability a Guide to Possibility?' *Philosophy and Phenomenological Research*, Vol. 53, I, 1-42.

Philosophy Program
Research School of Social Sciences
Australian National University
Canberra 0200 ACT Australia
dstoljar@coombs.anu.edu.au