

Consciousness and Conceivability

Critical Notice

JOHN PERRY, *Knowledge, Possibility and Consciousness*. Cambridge, MA: The MIT Press 2001. Pp. xvi + 221.

The thesis that anything conceivable is possible plays a central role in philosophical debates about the self. Discussions about free will have focused, at least in the last hundred years, on whether a free yet determined action is conceivable. If it is, and if anything conceivable is possible, then a deterministic physics would by itself pose no obstacle to human freedom. Current debates about the nature and value of personal survival turn on whether it is conceivable for a person to move from one body to another. Discussions about the topic of John Perry's impressive book, the relation between a person's mental and physical states, has recently centered on whether it is conceivable for physically identical beings to differ in their conscious experiences. Some claim that zombies, beings physically just like us but without any conscious experiences at all, are conceivable. If they are, and if anything conceivable is possible, then it would seem to follow that facts about conscious experience are not physical. In short, some form of mind-body dualism would have to be right after all.

While conceivability claims are at the center of many debates about the self, evaluating them is not at all straightforward. For beliefs about what is conceivable or possible interact with beliefs about what is in fact the case in subtle ways.¹ For one thing, it is reasonable to think that our

1 For a very helpful discussion of this, see Robert Stalnaker, 'What is it Like to be a Zombie?' in *Conceivability and Possibility*, T.Z. Gendler and J. Hawthorne, eds. (New York: Oxford University Press 2002).

conceptual limits and powers are contingent and that had our capacities or environment differed in certain ways we might have formed or been able to form different conceptions from those we in fact have or can form. Causal accounts of reference and meaning suggest one way this could be so. If it is so, then assumptions about what these limits and powers in fact are, assumptions that we believe are contingent and empirical, might influence beliefs about what is conceivable for us. For another, beliefs about how things are may influence beliefs about what is possible. For on some views to believe that something is possible just is to believe that it is a way actual things and properties might have been. If this is right, then contingent beliefs about which things have which properties would influence beliefs about how these things might have been. What is more, possibilities are sometimes explicitly characterized in terms of the facts. A zombie world, for instance, is typically described as one that is physically just like the actual world but without conscious experiences. When possibilities are characterized this way, assumptions about how things in fact are influence beliefs about what such possibilities are like.

None of this shows that arguments that depend on conceivability claims are without merit. But it shows that claims about what is possible or conceivable might not always be neatly separable from claims about how things are as a matter of contingent, empirical fact. Debates that appear, on the surface, to be about which worlds are conceivable or possible may in fact turn on hidden empirical disagreements about which world is actual. Empirical investigations might reveal that what looked like a description of one possibility is in fact a misdescription of another. Conceivability claims and arguments that depend on them thus need to be evaluated very carefully.

The nature of conceivability arguments, especially in connection with conscious experience, is one of the main topics in John Perry's excellent book *Knowledge, Possibility and Consciousness*. His aim is to defend a view he calls 'Antecedent Physicalism,' according to which the qualitative aspects of human conscious experiences are identical with non-functional physical properties of the brain. Antecedent Physicalism entails that zombies are not possible and so Perry must deny either that they are conceivable or that anything conceivable is possible. He seems to deny the former, arguing that the apparent conceivability of a world inhabited by zombies is merely the real conceivability of a world where one's beliefs and concepts have a different content. To see this, he says, we must first recognize a 'broader conception of the content of thought' (113). More specifically, we must recognize that beliefs have reflexive contents, contents that are about the belief's own semantic properties.

Perry's discussion is remarkably rich and full of insight. His articulation of an identity thesis for the qualitative aspects of conscious experiences is sophisticated and subtle. And his account of reflexive contents

and the role they play in cognition is impressive. I am sympathetic with much of what he says. In particular, I too doubt whether the qualitative aspects of conscious experiences are functional properties, and I also question the conceivability of zombies. But I think that Antecedent Physicalism is further from common sense than Perry suggests. I will discuss this in section I. I am also unsure how far Perry's appeals to reflexive contents can help us to understand the illusion, assuming that it is one, that zombies are conceivable. I will discuss this in section II.

I

Antecedent Physicalism concerns the qualitative aspects of such experiences as a painful toothache or a color perception. Perry concedes that these experiences might have a functionalist component: it might be that for an experience to be, say, a toothache, it must play some functional role. But he denies that its playing such a role exhausts its nature. There is, in addition to any functional role it might play, its qualitative aspect, the elusive 'what it is like' to have the experience. A physicalist about these aspects, about 'qualia,' has few options. One is to identify the qualitative aspects of human experience with specific physical properties. Another is to deny identity but maintain that qualia supervene, either logically or as a matter of contingent physical law, on such physical properties. Perry rejects the second option. As I understand his reasons, he believes that qualia could supervene on physical properties only if they were functional properties, and he denies that they are functional (34; cf. 44). Thus committed to Physicalism, Perry then embraces an identity thesis, one that identifies (human) qualia with non-functional neural properties.

After a brief introductory chapter, the second and third chapters of Perry's book are devoted to developing this view. He offers little in the way of positive reasons to accept it and nothing new is added to support its most controversial element, the claim that qualia are not functional properties. Instead, Perry directs doubters to the works of others, most notably Ned Block. I found this surprising since the typical arguments used to support this claim are conceivability arguments. The most well know one appeals to the alleged conceivability of an inverted spectrum. Given that one of Perry's topics is the nature and limits of conceivability, one might have expected him to explain why we should trust in the apparent conceivability of inverted spectrums but not that of zombies. How can we be so confident that we are avoiding conceptual confusion in the spectrum inversion case when, as Perry sees it, we so easily fall victim to it in the other? Disappointingly, Perry says nothing on this score.

Admittedly, Perry's primary goals are defensive. He wants to defend a certain physicalist conception of qualia against three now familiar anti-physicalist arguments: David Chalmers' Zombie Argument, Frank Jackson's Knowledge Argument, and Saul Kripke's Modal Argument.² He calls his view *Antecedent Physicalism* in part to stake out a dialectical position: it is, he thinks, a version of physicalism about qualia that shares enough with common sense to shift the onus onto the anti-physicalist. In the absence of convincing objections it is reasonable, he holds, to accept Antecedent Physicalism. His aim is to show that standard anti-physicalist objections are not convincing when leveled against Antecedent Physicalism.

In fact, though, it is not so clear just what Perry takes himself to be defending. For, as he admits, he is deliberately equivocal in his uses of the key words 'state' and 'aspect.' In one place, he imagines picking up a piece of hot charcoal and suffering a pain. He writes:

My having the pain, or being in pain, was a particular event, one that was caused by my picking up the charcoal and led to my dropping it quickly. The event involved my being in a certain state. Suppose that ϕ is the state brains are in when the possessors of the brains are in pain. Then we can say that ϕ was an aspect of the state of my brain, or that my brain was in state ϕ , or that I was in state ϕ' , where ϕ' is the state of having a brain that is in state ϕ . (36)

He complicates the terminological situation further by saying that 'Sometimes I'll say that the subjective characters are states of the person, sometimes that they are states of the brain, and sometimes that they are aspects of brain states, depending on what seems to fit in a certain context' (36-7). There are important differences among these formulations, differences that matter to the plausibility of the view he is trying to embrace.

Part of the problem is that both ordinary and philosophical discussions about states are rarely very clear, especially on the difference between state tokens and state types. Talk of state tokens, as opposed to types of states, is naturally understood as talk of property instantiations: to talk about the states of a thing is to talk about its having certain properties. Talk of state types is naturally understood as talk of the properties instantiated in tokens states of that type. Thinking of things in this way, we could represent a given state of a thing as an ordered set,

2 See David Chalmers, *The Conscious Mind* (New York: Oxford University Press 1996); Frank Jackson, 'What Mary Didn't Know,' *Journal of Philosophy* 83 (1986) 291-5; and Saul Kripke, *Naming and Necessity* (Cambridge, MA: Harvard University Press 1980).

with the thing and property as elements. So, for instance, a term of the following form could be used to denote the state of a's being P.

<a, being P>

Such a term would denote a state token, the instantiation of a certain property in a certain individual.³

Using these conventions, we could in the case of Perry's imagined pain identify the following states.

(1) <Perry, being in pain>

(2) <Perry's brain, being P>

(3) <n, being R>,

where n is some part of Perry's brain, and being P and being R are physical properties. (1) denotes the state of Perry's being in pain; (2) denotes a state of Perry's brain; and (3) denotes a state of some part of his brain.⁴ To say, as Perry sometimes does, that qualia are aspects of brain states suggests that they are properties of such states as are denoted by (2) and (3) not the instantiated property but a property of its instantiation. But Perry also says that qualitative states are states of the person, which suggests that qualia are properties instantiated in states like those denoted by (1). In one place he implies that qualia are properties of parts of the person's brain, which suggests that they are properties of brains or their parts, which would make them like the property instantiated in

3 I think another source of confusion, though one that is more difficult to avoid, is talk of a state's functional or causal role. I find it more natural to think of *individuals* as playing functional or causal roles. It is the electron that has a charge, and the piece of metal that plays the role of a valve, not some state of the electron or some state of the piece of metal. No doubt what roles an individual can play depends on its properties. But talk of a state's playing a certain functional role strikes me as a confusing way of talking about the functional roles of the subject of the state. In the case of mental states, this confusion is abetted, it seems to me, by thinking of beliefs and desires as individuals somehow constituted in the brain. Perry claims that it is simply commonsense to think of them this way (37). I disagree. But this is not the place to engage these issues.

4 I think it is implausible to suppose that any of these is identical with either of the others. Perry's having some property simply cannot be the same as his brain's having some property. I also think it is implausible to suppose that any of the properties involved in these states is identical with either in the other two. Perry himself has no neural properties, though his brain surely does.

the state denoted by (3) (67). This equivocating makes it difficult to get clear on his view.

Of course, given how little is known about human brains and about their role in conscious experience, Perry deserves some latitude. He is not to be expected to offer specific identities and it is fair for him to build some flexibility into his proposal. But having distinguished these versions, I think we can raise some reasons for doubt. More specifically, it is hard to think of any physical properties, on any version of the view, that are both non-functional and plausible candidates for being the qualitative aspects of experience. Consider the possibilities.

Suppose, first, that Perry's view is that qualia are properties of a state. In particular, suppose that the qualitative aspect of Perry's pain is supposed to be a property of the state denoted by (2).

(2) <Perry's brain, being P>

What non-functional physical properties does that state have? It has non-functional first-order properties, such as temporal and spatial ones. But these are not plausible candidates for being qualia, mostly because too many other states without qualia also have such properties. Though mass properties are not functional, states have no mass, though of course Perry's brain does. States also have second-order properties, such as the property of being a given kind of state. The state denoted by (2) has, for instance, the property of being a state of Perry's brain. But that property cannot be the qualia, again because so many other states have it without having those (or any) qualia. What about the second-order property of being a P state? We are supposing that (2) denotes a state of pain, which Perry takes to be a functional state. He maintains that different pains, even ones that play the same functional role, can have different qualia. So his view cannot be that the qualitative aspect of what (2) denotes just is its property of being a P state. Perry remarks that a brain state might involve some process in some part of the brain. Suppose it denotes a state that has the property of involving a given process in a given brain part. This is not a functional property. But how plausible is it to suppose that it is a quale? Could it really be that what-it-is-like for Perry to have that pain just is his pain's having the second order property of involving some brain process?

Suppose, instead, that the view is that qualia are properties of parts of the brain, and not properties of states of the brain (or of states of brain parts). Consider some group of neurons, *n*. What non-functional properties does it have? It is composed of certain chemicals. Is this second-order property of being composed of those chemicals a quale? I have no proof, or even any argument, that it is not. But it just strikes me as incredible that the qualitative aspect of my experience of, say, eating an

orange just is a second order property of being composed of certain chemicals. It just seems like the wrong kind of property. I have not shown that *no* non-functional property of a brain state or brain part is a plausible candidate for being qualia. But it seems to me that the fact that it is so difficult to come up with any examples of such properties, let alone plausible candidates, is a serious problem for Perry's view.

One source of the difficulty, I think, is that Antecedent Physicalism holds that qualia are not functional or causal properties: 'the what-it-is-like aspect of [an experience] is not a causal, historical or functional property' (86).⁵ The thesis that qualia are not functional or causal is, however, compatible with the thesis that an experience's quale is determined by, or supervenes on, functional properties. Though Perry, so far as I can see, does not directly address this issue, I take his view to be that qualia are not determined in this way. For he holds that it is possible for the color experiences of functionally identical agents to differ in qualia and he insists that while this is incompatible with a functionalist or causal account of qualia it poses no problem for Antecedent Physicalism, since functionally identical agents might differ physically (91). According to Antecedent Physicalism, functionally identical agents could differ in their qualitative experiences so long as they also differed physically. So Perry must hold, not only that qualia are not themselves functional or causal properties, but also that they do not supervene (at least not logically) on functional or causal properties.

But the idea that an agent's qualia do not supervene on her functional or causal properties faces a familiar and serious difficulty, one that Perry does not address. If qualia do not supervene on functional or causal properties, then changes in the qualia of an agent's experience need make no difference to the agent's functional or causal states. But then such a change need make no difference to what she would do, or say, to what she would believe or desire, or to any of her causal powers. In short, qualia would be epiphenomenal.

5 But earlier, he wrote as follows:

These subjective characters of brain states are probabilistically/nomically related to various other properties of brain states, have causal roles, and may have functions. But it is no part of common sense that they are *no more than* such causal roles or functions, and in fact it is a pretty firm postulate of reflective common sense that they are more than that' (39; italics in original).

I am not sure how to understand this passage. It might be suggesting that qualia play a causal role, but only contingently so. I address this suggestion in the main text. But it seems clear that Perry believes that qualia are, or at least have, properties that are not functional or causal properties.

Epiphenomenalism is the view that mental states are caused but have no effects, or at least no physical effects. They make no difference to the physical world. They may appear to do so, since the things that cause the mental events also cause other physical events that will appear to be caused by the mental event. But this is an illusion, Mental events ... are "nomological danglers." (67-8)⁶

As Perry notes, epiphenomenalism is not a common sense doctrine. The fact that my pain has the qualitative character it does is ordinarily taken to be part of the story of why I react as I do. Perry insists that Antecedent Physicalism is not committed to epiphenomenalism about qualia. But it is hard for me to see how this could be so if, as he says, functionally and causally identical agents could have qualitatively different experiences.

Someone might suggest in Perry's defense that although qualia are not individuated by a causal role, different qualia in fact do play different causal roles.⁷ On this view, although it is not essential to a quale that it play a given role, as a matter of fact each quale does play its own causal role. But there are two difficulties with this. First, it is hard to see how one could reasonably hold it. For to believe that it is *possible* for changes in qualia to make no causal or functional difference is to believe that such changes could occur without any change in the agent's beliefs. Such changes in qualia would, in this sense, be unnoticeable. But if one believed that unnoticeable changes in one's qualia *could* occur, then on what basis could one believe that they do *not* occur? For, by one's own admission, if they were occurring one would not notice it. How can one confidently believe that nothing unnoticeable is going on? It is difficult (at least for me) to see how one who believed that qualia need play no causal role could have any reason to believe that, as a matter of natural law, they in fact do. But this is what believing the present suggestion requires.

There is a second difficulty. One of the often advertised virtues of an identity thesis is that it reduces the explanatory burden. A physicalist who says that mental properties merely supervene on physical ones, and who takes this to be a contingent fact, needs to explain this supervenience.⁸ In virtue of what do these qualia supervene on those physical

6 I have corrected an amusing typographical error in this passage. In the original, epiphenomenal mental events are characterized as 'nomological dangers.'

7 A similar proposal would be that while qualia do not logically supervene on causal or functional states they do so as a matter of contingent natural law. What I say in the body of the text applies to this proposal too.

8 One might think that such a position should not count as a version of physicalism. But, putting the point picturesquely, the position entails that in creating the physical facts and establishing the laws of nature God had thereby created qualitative states. That sounds, to me, like a form of physicalism about qualia.

states? But a physicalist who says that mental properties are necessarily identical with physical ones, need not explain why this is so, since true identities need no explanation. On the suggestion under consideration, though qualia are identical with non-functional physical properties, they do, as a matter of contingent fact, play a functional role. But their playing such a role would need some explanation. In virtue of what does this quale play that role? And this is the very kind of question that adopting an identity thesis was supposed to help one avoid. So adopting this suggestion would remove one of the chief virtues of the identity thesis.

I said that Perry seems to hold that the qualia of an experience do not supervene on other functional or causal properties. So far as I can tell, he does not explicitly argue for this claim, though I have suggested that some of what he says seems to commit him to it. He does explicitly say that qualia are not themselves functional properties. Now, one might think that qualia could supervene on other properties only if qualia were functional. But it is not in general true that only functional properties supervene. Indeed, our experiences themselves have properties that are not functional and yet which supervene on functional ones.⁹ And what is more, to say which of these properties a given experience has is to say something about what it is like for the agent to have that experience. In sum, our experiences have non-functional properties that supervene on functional ones and which constitute, in one clear sense of the phrase, what it is like to have those experiences. I will close this section by discussing this.

Let me begin with an analogy. Being Mayor of New York City is no doubt very different from being Mayor of Buffalo. This might involve a difference in powers. Perhaps the Mayor of New York City has greater constitutional control of the city's school boards or more authority over its garbage collection. This would be a functional difference, broadly conceived. But the jobs might differ without differing in their powers. One Mayor might be better positioned to exercise a power they both have. Perhaps the Mayor of New York City has more money to spend on schools or on restoring his office. This need not be a functional difference, since it need not involve a difference in what each would do

9 The idea that mental states have non-functional properties that supervene on functional states is discussed in Robert Stalnaker, 'Comparing Qualia Across Persons,' *Philosophical Topics* 26 (2000) 385-405. Stalnaker's example is a desire's strength, which (in one sense, anyway) is a relational property, ordering the desire in relation to other desires. A desire's strength is not itself a functional property of the desire, even if the desire itself is a functional state. But he does not discuss what I will call a state's position within a system of states.

in various *possible* situations. Rather, it would be a difference in what each *will* do, given the actual situation.

This suggests that there is a difference between a state's functional role and its position in a system of other states. The former is a matter of how the state would interact with other states; the latter is a matter of which states there are for it to interact with. The example of the Mayors shows that functionally identical states can occupy different positions. It also helps to see that functionally different states can occupy the same position. Suppose that the city charter were changed giving the Mayor of New York City sole authority over a planned for, but not yet implemented, new branch of city government. And suppose that no changes were made to the powers of any other official. In that case, while the Mayor's powers have changed, his position has not. For he is in no position to exercise his new powers. What it was like to be Mayor after the charter's change is, in this sense, just like what it was like to be Mayor before the change.

A functional state's position within a system of states is thus not a functional property. But it is a relational one: a state's position is a matter of what else there is in the system. So it supervenes on other states: no change in a state's position is possible without a change in those other states. And if all the states in the system are functional states, then a state's position in that system supervenes on functional states: no change in a state's position without some functional change in the system. This shows that although a state's position is not itself a functional property, it might supervene on functional properties.

Consider now experience. Suppose that Mary and John are looking at a red square. Their experiences might differ in two ways. On one hand, they might play different functional roles. In order to focus attention, we need to do some idealizing. Suppose that the only difference is that Mary's experience disposes her to remember roses she has seen in the past, while John's disposes him to remember emeralds he has seen. In this case, the experiences have different powers. If an experience's functional role is essential to it, then, even if its functional role does not exhaust its nature, Mary and John's experiences would be of different kinds. I think Perry would agree that experiences can differ in this way.

Someone might want to say that this would be a difference in what it is like for each of them to look at the red square. This would be to say that the what-it-is-like aspect of an experience is, at least in part, a functional property: that functionally different experiences must have different qualia. Like Perry, I am inclined to resist saying this, though my reasons are not his. He resists it because he thinks that functionally identical agents might differ in their qualia. This is, he thinks, a lesson we should draw from the inverted spectrum stories. But I want also to resist this lesson, for reasons I have already discussed. But like Perry I

would say that John and Mary's experiences might differ in a non-functional way, and that this difference might be what we aim to characterize when we say what it is like for each of them to have that experience.

The difference I have in mind is a difference in the position each experience occupies in the agent's cognitive life. We have seen that states that play different causal roles might nonetheless occupy the same position in different systems. They would do so if the states with which they can interact are the same. But (again idealizing in order to focus attention) suppose that Mary and John's cognitive lives, with the exception of this one experience, are the same. And suppose that Mary has no memories of seeing roses, and that John has no memories of seeing emeralds. Their experiences play different roles, but this difference will not manifest itself. For the experiences occupy the very same positions. From the inside, it is tempting to say, their experiences are the same. We might mark this sameness, while recognizing the functional difference, by saying that functionally different experiences have the same qualia. What it is like for John to look at the red square is what it is like for Mary to look at it, even though their experiences are of different functional kinds.

Suppose now that John and Mary's experiences do not differ functionally. Suppose that having that experience disposes each to recall the smell of roses and the taste of red wine. Mary has memories of smelling roses, but none of tasting red wine while John has memories of tasting red wine, but none of smelling roses. Mary looks at the red square and recalls the smell of roses; John looks at it, and recalls the taste of red wine. This difference is possible, even though the experiences are functionally the same, because the states occupy different positions. We might mark this difference in the experiences, while recognizing the functional sameness, by saying that what it is like for John to look at the red square is very different from what it is like for Mary to look at it, even though the experiences are of the same functional kind. (What it is like for a parent to look at baby pictures is very different from what it is like for non-parents, but not because of a difference in the experience's functional role: the difference is in the rest of the cognitive system.)

I have argued that there is a difference between an experience's functional role and its position within a cognitive economy, and have suggested that a state's position is part of what we aim to characterize when we say what it like for someone to have that experience. I foresee two objections. One concerns the idealizations I made in order to clarify the distinction. No experiences differ functionally in just one way and no cognitive economies differ in only one state. In the real world, there are complex functional differences among experiences and in the cognitive economies of any two agents. It was because I idealized away from this complexity that I could draw the distinction between a state's role and its position. But, and this is the objection, we could never draw this

distinction in practice. We could never tell whether a difference in the experiences of two agents was a difference in functional role or a difference in position.

I am actually quite sympathetic to this objection. The verificationist in me is tempted to go even further: perhaps there are cases where there is no fact as to which kind of difference it is. I think that such a verificationist would have common sense on his side. But this is a long story for another day. It is enough for now, I think, to have shown that there is, in theory anyway, a difference between an experience's functional role and its position, and that its position can supervene on functional states, and that characterizing an experience's position is, sometimes anyway, what we are trying to do when we say what it is like for someone to have that experience.

The second objection is this: even if we grant that an experience's position deserves the title of 'qualia,' this cannot be the whole story of qualia. For it is conceivable that experiences that play the same functional role and occupy the very same position should nonetheless differ in qualia. Suppose that Mary and John's experiences are functionally the same, and that the rest of their cognitive economies are the same too. The experiences thus occupy the same position. Still, it is conceivable that the square looks to John the way red squares look to me while it looks to Mary the way green squares look to me. This would be a difference in the experiences that is neither functional nor positional. And surely this would be a difference in what it is like for each of them to have the experience, a difference in qualia.

This is a more difficult objection to address. For I am inclined to deny that this situation really is conceivable. But I do not know how to argue for this. John Perry spends a good deal of his book trying to explain how we can be mistaken or confused about the limits of conception. I think he is right that we can be mistaken or confused about this. But I am not sure how far what he says goes in helping us to understand such mistakes and confusions. Let me now turn to this.

II

The bulk of Perry's book defends Antecedent Physicalism against three anti-physicalist arguments. David Chalmers' *Zombie Argument* holds that it is conceivable that beings physically just like us should have conscious experiences altogether lacking in qualitative feel.¹⁰ Saul

10 In chapter 4, Perry argues that the *Zombie Argument*, though advertised as an argument for dualism, is really an argument for epiphenomenalism. Perry argues

Kripke's Modal Argument, directed against the view that pains are identical with neurological states, holds that for any given neurological state one can conceive of it obtaining in the absence of a pain's qualitative feel. Frank Jackson's Knowledge Argument holds that one can know, and so conceive of, all the physical facts involved in color experience without having any adequate conception of, and so without knowing, the qualitative ones. All three arguments turn on claims about conceivability and Perry argues that they involve a similar kind of conceptual confusion.

Perry's aim is not to prove that what these arguments claim to be conceivable is in fact inconceivable. Rather, it is to offer an account an Antecedent Physicalist could give of how someone can mistakenly think that they are conceivable, of how, more generally, one can be mistaken or confused about the limits of conception. His strategy turns on the fact that so long as there is a difference between what an agent believes or says about the world and the means they use to believe or say it, it will be possible for the agent to mistake or confuse a contingency in those means for a contingency in the world.

Perry's account depends on a specific view of the nature of beliefs and thoughts, one which models them on natural language sentences. According to this view, which he sketches in chapter 3, thoughts and beliefs are like sentences in being composed of parts. Where sentences are composed of words, beliefs and thoughts are composed of 'ideas.' Ideas that stand for individuals are called 'notions' and those that stand for properties or relations are called 'concepts.' Ideas are, on this view, themselves individuals, particular objects somehow constituted in the brain. They become 'associated' with one another to form thoughts, beliefs, desires and other mental states.

Ideas, Perry says, may have two kinds of semantic properties. On one hand, they may have a descriptive content which may (but need not) be true of, or 'denote,' some individual or property. On the other hand,

persuasively that the possibility of a zombie world would show that qualia make no causal difference to the actual world. But, he notes, a dualist need not accept this epiphenomenalism and a physicalist need not deny it. But I am not sure how this is supposed to show that the Zombie Argument is really an argument for epiphenomenalism and not for dualism. For the issue is whether a physicalist could accept the possibility of a zombie world, and it seems clear that she could not. In fact, Perry seems to agree. What he says the physicalist can accept is the possibility of a world that lacks qualia and is physically just like the actual world but for those physical facts (whatever they are) that constitute the qualia. But this is not a zombie world, and is not the world Chalmers argues is possible. So I do not see how Perry's argument is supposed to succeed.

ideas may (but need not) refer to some individual or property. The referent of an idea need not be the object or property it denotes. Denotation depends on the idea's descriptive content, while reference depends on its 'circumstances,' which include causal relations to the outside world and facts about how the concepts gets 'applied' by its owners. Crucial to this picture is that an idea's semantic properties are contingent. An idea could have had some other descriptive content, its descriptive content can change over time, and it could have had a different referent even with no change in its descriptive content.

When ideas associate to forms thoughts and beliefs, the result has truth conditions. In the case where the thought involves a notion that refers to some individual, *a*, and a concept that refers to some property, *P*, the thought is true just in case *a* is *P*. Picturesquely, we can represent this by a set of possible worlds, those where *a* is *P*. Perry calls this the thought's 'subject matter content,' since it concerns the referents of the thought's ideas. But, Perry says, there are other truth conditions associated with that thought. It is true just in case the referent of its notion is *P*. It is true just in case *a* is in the extension of the referent of its concept. It is true just in case the referent of its notion is in the extension of the referent of its concept. And no doubt there are more. We can represent these truth conditions by sets of worlds too: those where the truth conditions are satisfied. Perry calls these the thought's 'reflexive contents,' since they concern the thought's own constituents.

A thought's subject matter content always has the same truth value as its reflexive contents. They are materially equivalent. But they are not strictly equivalent. This is clearest in the case where the subject matter content is a necessary truth. The reflexive content is never a necessary truth since it is contingent that an idea has the semantic properties it does and so no claim about its semantic properties is necessarily true. But even when the subject matter content is contingent, it is not strictly equivalent to any of the thought's reflexive contents, again because the truth of the reflexive contents always depends on contingent facts about ideas whereas the subject matter content does not. It is this divergence between the subject matter content and the reflexive contents that Perry appeals to in accounting for mistakes about the limits of conception. In such cases, on this view, one mistakes or confuses the possible falsity (truth) of a contingently true (false) reflexive content for the falsity (truth) of a necessarily true (false) subject matter content.

To see how this is supposed to go, consider an example. Suppose that *S* is looking at a red square and that *Q* is the quale of her experience. *S* has a concept, *N*₁, that refers to *Q* and which has descriptive content *D*₁ and a second concept, *N*₂, that refers to a neural property, *P*, and which has descriptive content *D*₂. Suppose, as Perry holds might be the case, that *D*₁ is a purely demonstrative content while *D*₂ is a highly theoretical

description of a neurological state or property and that Q, the quale, just is P, that neurological property (as per Antecedent Physicalism). Suppose, now, that S doubts whether P is Q: it is, she says, conceivable to her that P is not Q. According to Perry, S is misdescribing what she is conceiving of. For since there is no possible world where P is not Q, she is not conceiving of such a world. Rather, she is conceiving of a world where one of the reflexive contents of her thought that P is not Q is true. Plausibly, it is a world where N1 and N2 do not co-refer. In one such world, N1 refers to George Bush and N2 to Bill Clinton. This is a world where one of her thought's reflexive contents, namely that its constituent concepts co-refer, is false. But, intuitively, since this is not a world S would confuse for one where P is not Q it is not this reflexive content that captures what she is really conceiving of. A better candidate, and the one Perry proposes, is a world where N1 and N2 have the descriptive contents they in fact have but where one of them fails to refer to P. Since D1 is, we are supposing, a purely demonstrative content, it is hard to see how N1 could have had D1 and yet failed to refer to P. But D2 is, we are supposing, a highly theoretical description of P and it is not impossible, Perry says, that it might have been the descriptive content of N2 even though N2 referred to something other than P.

Reference is not a matter of denotation or fit but of the circumstantial facts that determine [a concept's] source and applicandum. There are many possible ways these facts might be different. There are worlds in which scientists, perhaps because of lack of proper funding, manage to come up with a concept that denotes stimulated C-fibers but for which some other state X is both source and applicandum. This world would be strange but not impossible. (185; cf. 148, 199)

Such a world is not one where Q is not P. It is, rather, a world where S's system of concepts is different. Confusing these worlds, she describes herself as conceiving of a world where Q is not P. But this is not what she is really doing.¹¹

11 I might be misrepresenting Perry's position here. For, on his view, the content that Q is not P could be the subject matter content of S's belief, even though it is necessarily false. One might think this is enough to show that a world where Q is not P is conceivable. Perry does not clarify this point. If this is his view, then in response to the claim that zombies are conceivable he has to deny that anything conceivable is possible, and his aim in the book would then be to explain how one can be mistaken about what is possible and not, as I have represented it, to explain how one can be mistaken about what is conceivable. It is not clear to me from the text which of these aims is Perry's. But whatever his aim, his strategy is clear: the mistake is to be explained as deriving from confusing a contingency in one's means of representing the world for a contingency in the world itself.

Perry uses this account of conceptual confusion to respond to the three dualist arguments. As I have said, Perry is not aiming to prove that it is impossible to conceive of a zombie world, or of any given neurological state's obtaining in the absence of pain, or of knowing all of the physical facts without knowing any of the qualitative ones. Perry's aim is just to show how an Antecedent Physicalist could explain how one might be mistaken in thinking that he is conceiving of such things, how one might be confused about the contents of his conceptions. His discussion of each argument, and in particular his discussion of Chalmers' appeals to primary intentions, is far richer and more detailed than I can do credit to here and repays careful study. I am sympathetic to his general strategy, and I think his account is ingenious and sophisticated. But I am not sure how much the picture he develops really helps us to understand this kind of confusion.

Perry's reliance on a linguistic model of thought is, in my view, distracting. He talks as if this model is common sense about the mind. It might have become, in recent times, common *philosophical* sense about the mind. But it is, as Perry knows, a highly speculative and justly controversial picture. Reliance on it raises questions that are not easily answered and which distract from Perry's general strategy in trying to understand conceivability. For one thing, Perry does not explain how ideas are individuated. It is not by their semantic properties, since they have these only contingently. As a result, it is not clear what it means to say that my idea of pain could have had a different semantics. Nor does Perry explain how ideas 'form' into beliefs and thoughts through 'association.' If a belief is an association of ideas, and ideas have their semantic properties only contingently, do beliefs have their semantics contingently too? Is the idea in my belief that pain is unpleasant the same as the idea in my desire to avoid pain, or do I have different ideas of it? If the former, then what is the relevant notion of association; if the latter, then how do these ideas differ? Perry also does not explain how a belief gets its reflexive contents, or even how any idea comes by any of its semantic properties. Finally, since ideas are particulars, no two agents have the same ideas, and so no two agents' beliefs have the same reflexive contents, even if they have the same subject matter content. This means that no agents have the same confusions about the limits of conception. It might also mean that since no agent is acquainted with the ideas of any other agent, no agent is in a position to formulate the reflexive contents of another's thoughts and beliefs. All of these, it seems to me, are controversial consequences of a picture that is not an essential part of the broad strategy Perry is trying to pursue.

The heart of Perry's strategy for explaining how one can be mistaken about the limits of conception is that it is possible to confuse what one says or thinks about the world and the means one employs to talk and

think about it. So long as an agent employs some means such confusions are possible. This is an old strategy, with roots in Wittgenstein's early work and in Carnap's distinction between statements made within a framework and those made about the framework. Perry's distinction between subject matter content and reflexive content parallels Carnap's. To make good on the strategy, though, one must make clear what role reflexive contents play in an agent's cognitive life. Unfortunately, Perry is not very clear on this. He suggests that it would be a mistake to think that an agent believes the reflexive contents of her beliefs.

In general, the propositions we believe, the ones referred to by the phrase "what he believes," are not ones about our own perceptions and ideas but ones about their *subject matter*: the things, people, places, and events that our perceptions and ideas are of. That is, when we describe our beliefs and perceptions we standardly do so in terms of their subject matter content. We do not say that I believe, of my Dretske notion, that it is of someone who wrote a certain book, but that I believe that Dretske wrote the book. But those very same beliefs have *other* contents... They are not alternative things the agent believes; they are less loaded contents of his belief. (133; italics in original)

But if the agent does not believe the reflexive contents, then just what is her relation to them? The question is not whether there are such contents. Thought of a sets of possible worlds, their existence is no more problematic than that of subject matter contents. And the question is not whether the agent bears some relation or other to them, for that too seems unproblematic. The question is whether an agent bears a *cognitively significant* relation to them, and if so, what that relation is? Perry seems sensitive to this question. For immediately after the passage quoted above he suggests that an agent can be 'attuned to' facts or situations the reflexive contents of her beliefs are about. Apart from some unhelpful metaphors about 'information games,' Perry does not say much more about the nature of attunement. It is not naturally counted a propositional attitude ('attunes that' sounds odd). It is most naturally taken as a species of knowledge ('is attuned to the fact that' sounds alright). But attunement is not a species of knowledge, since reflexive contents can be false.

This issue can be approached from another direction. To characterize an agent's beliefs is to say something about how she is disposed to act, how she will treat the world. We do this by identifying a way the world might be, a content, and characterizing her dispositions in terms of it. To say that she believes a certain content is to say (in part anyway) that she is disposed to act as if the world is the way it would be were that content to be true. To say that she desires it is to say (in part anyway) that she is disposed to act to make it that way. We describe her dispositions by relating her to a content, just as we identify an object's mass by relating

it to a number. But it is not enough simply to identify the content. If our aim is to say something about the agent, we have to say what the relation is, whether it is belief or desire or what have you. Noting that an agent's belief is true just in case the referent of her belief's notion is P specifies a way the world might be, it specifies a content. But identifying this reflexive content of her belief does not by itself characterize the agent's dispositions or cognitive life. It would be like trying to characterize an object by identifying a number: until the object's relation to the number is specified we have not made clear what we are trying to say about the object. Perry might be right that an agent need not believe the reflexive contents of her states. He might even be right that we can use reflexive contents to characterize important aspects of an agent's cognitive life. But until he says what cognitive attitude she bears to them, it is not clear what he thinks that aspect is.

I think a case can be made that the attitude is naturally thought of as belief, or at least as a species of acceptance. Suppose that S says 'Q is not P,' and that we believe that that sentence expresses something necessarily false. How should we understand what S has said? Assuming that she was sincere, she expressed belief. So to understand what she was trying to say, we might try to figure out what she believes. What would someone who uttered those words believe? For one thing, she must believe that that sentence says something true. At least, she has acted as if she believed that. Assuming that she means what we do by 'is not,' she must further believe that 'Q' and 'P' refer to different things.¹² We would fail to understand her if we took her to believe, as we do, that they refer to the same thing. So one thing we know about her beliefs is that we disagree with her about some semantic facts. This is a disagreement about how our words hook up with the world.

But is our disagreement with her nothing more than a semantic one? Doesn't it also concern what properties there are: she thinks there are two where we think there is but one? It does seem that the actual world, as she takes it to be, differs more than just linguistically from the actual

12 Perry resists saying that the agent has such a belief, in part I think, because he believes that claims about an agent's beliefs are claims about ideas composed in her brain, and not just claims about her dispositions: 'When we ascribe a belief to a person about a certain individual and involving a certain property or relation, we suppose that the agent has a notion of that individual and an idea of that property or relation' (134-5). It seems to me that this is building too much theoretical content into claims about an agent's beliefs. One could, more cautiously, agree that such claims are made true by facts about the ideas inside an agent's head while denying that what they claim is that such facts obtain. But I am doubtful about the truth of even this more cautious claim.

world, as we take it to be. But it is not easy for us to say what that further difference is. For *we* cannot use our words 'P' and 'Q' to state this difference. We cannot, for instance, say that it is a world where P is not Q, since we believe there is no such world. We can say that it is a world where the words 'P' and 'Q' refer to different properties. But can we say what those properties are? Is either of them P? It is not clear to me how best to approach these questions. I am inclined to think that there might be no way to answer such questions and that this kind of conceptual incommensurability is a fact to be understood and not a problem to be solved.

All of this helps us to see how disagreements about what is conceivable or possible can depend on disagreements about how things in fact are. For it is clear that if S is right about the semantic facts then the sentence 'Q is not P' expresses something necessarily true, but that if we are right then it expresses something necessarily false. If all sides further agree that anything conceivable is possible, then we could also agree that if S is right about our language then the sentence expresses something conceivable, because necessarily true, whereas if we are right then it expresses something inconceivable, because necessarily false. Which of us is right about what worlds are possible (and so conceivable) depends on which of us is right about which world is actual and that disagreement concerns not just which things have which properties but also how our words relate to whatever things there are and properties they have. One of the chief virtues of Perry's book is the extent to which it brings to the surface the puzzling way debates about possibility and conceivability intermingle these semantic and metaphysical questions.¹³

DAVID A. HUNTER
Buffalo State College
1300 Elmwood Avenue
Buffalo, NY 14222
USA

13 I am indebted to Andrew Hunter and to two anonymous reviewers for very helpful comments on an earlier draft of this paper. Part of section I was read at the 2002 Conference for the Scientific Study of Consciousness, in Barcelona, where I benefited from comments by Ned Block and Joe Lau.

