REPLIES

JOHN PERRY

CHALMERS

It would have been pleasant if David Chalmers, to whose arguments a good deal of *KPC* is devoted, had simply said that the book won him over completely, vowed never to ignore identity or commit the subject matter fallacy again, and expressed unbridled enthusiasm for all forms of reflexive content. 'Twas not to be.

The comments he actually provided are next best, from my point of view. They show clearly that his reasoning stands, or falls, with the two pillars I identified: ignoring identity in favor of supervenience, and the subject matter fallacy. If I'm not right, at least I'm insightfully wrong.

CHALMERS ON THE MODAL ARGUMENT

I'll start with Chalmers' version of the modal argument---where I left off in the Précis. In his comments, he says he doesn't follow what I say in the book on the modal argument. No doubt I was trying to do too many things at once in that chapter. For one thing, I saw Chalmers' two-level system as a descendant of the modified Fregean view I developed in "Frege on Demonstratives," and the two-tiered views I defended in subsequent works, including the book I wrote with Jon Barwise, *Siutations and Attitudes*[1]. I think I explored the limitations of this approach in a useful way in a number of papers, and in the book I hoped to go beyond criticism in explaining how I saw Chalmers' system related to my own thinking about two-tiered systems. But here I'll just focus on what is wrong with Chalmers' argument. In particular, I'll eliminate all discussion of indexicality to eliminate potential red herrings.

My objections are these:

---

[1] Jon Barwise and John Perry, *Situations and Attitudes* (Stanford: CSLI Publications, 1999); this is a reprint, with additions, of Jon Barwise and John Perry *Situations and Attitudes* (Cambridge: Bradford-MIT, 1983).

(A) In developing the concept of primary intension Chalmers does not follow his own explanation of what primary intensions are (functions that model actual world mechanisms of extension/reference), but instead takes primary intensions to be what I'll "non-reflexive modes of presentation". These are basically Fregean-like senses, that are supposed to both determine reference and serve as our manner of thinking of the reference. But there is no reason to believe that such things determine the extension/reference of names of individuals or kinds, including kinds of sensations. By "reflexive" in this case, I means "type-reflexive". "'Chalmers'" refers to the person people use the name 'Chalmers' itself to communicate about" is a type-reflexive account of the reference of the name "Chalmers". "'Chalmers'" refers to the most energetic neo-dualist" is a *non type*-reflexive account.

(B) If we follow Chalmers' practice rather than his explanation, and take primary intensions to be *non type*-reflexive mediating modes of presentation, it is totally implausible to equate conceivability with contingency of the primary proposition.

(C) Even if we ignore problems (A) and (B), a special problem arises with Chalmers' treatment of sensation terms like "pain". He doesn't choose, for the primary intension of "pain," any mechanism that models the actual world mechanisms of reference. Nor does he find some *non type*-reflexive mode of presentation. He simply takes pain itself to be the primary intension of "pain". This choice is unmotivated, and in fact quite baffling. We don't think about pain by being in pain, (although we often think about pain when we are in pain). This step is quite incapable of supporting the weight his argument puts on it.

(D) Finally, even if we follow Chalmers' choice for the primary intension of sensation words like "pain," the rest of his argument doesn't work. It is based on ignoring the type of physicalism I advocate, that is, the identity theory.

The argument Chalmers provides in his "Replies" illustrates all of these points. Recall that for Chalmers' possibilities come in two kinds, regular possible

worlds and centered possible worlds. A center is an agent and a time; a centered world a pair of center and world. The centers are relevant only for indexicals, so we can ignore them along with indexicality. Chalmers argument:

> Let us say that S is *primarily possible* when its primary intension is true at some … world. Let P be the complete microphysical truth about the world. Let Q be a phenomenal truth. Then the anti-materialist argument can be put as follows:
>
> (1) P&~Q is conceivable
>
> (2) If S is conceivable, S is primarily possible.
>
> (3) If P&~Q is primarily possible, materialism is false.
>
> So
>
> (4) Materialism is false. (Chalmers, Comments, p. XX)

Here are my objections.

(A) ***In developing the concept of primary intension Chalmers does not follow his own explanation of what primary intensions are (functions that model actual world mechanisms of extension/reference), but instead takes primary intensions to be non type-reflexive modes of presentation***.

We clearly need to understand what Chalmers means by *primary possibility* and hence what he means by *primary intension*. Here is one way to understand Chalmers' system, based on one of the explanations he gives in his book. Let *T* be a term with the definition

$T =_{df}$ *D-that(the x such that C(x)).*

"*D-that"* is David Kaplan's word, which rigidifies any description. According to Kaplan's rule for *d-that*, the intension of *T* is a function whose value for any possible world *w* is the *x* that is the object that satisfies condition *C* in the actual world. Call this the *secondary* intension of *T*. Recognize, in addition, the *primary*

intension of *T*. This is the function that ignores the "*D-that*," so the value, for each world *w*, is the *x* that is the object that satisfies condition *C* in *w*.

So far, so good. In applying this framework, the crucial question is: what is the condition *C*? Here there seem to be two ways we can go. We can follow Chalmers' directions or we can follow his example.

Chalmers says that the primary intension will use at each possible world the mechanism used to determine reference in the actual world. He also says that his system is intended to capture the insights of Kripke and others. By "mechanism," here, we don't mean some physical machine, but something more like the rule that is applied or practice that is followed to determine the reference. Such "mechanisms" are modeled, in intensional semantics, with functions of various sorts.

Consider the name "Gödel." Assume there is a definition,

Gödel = d-that(the x such that $C_{Gödel}$(x))

What is $C_{Gödel}$? It is supposed to be the mechanism that determines the references of "Gödel" in the actual world. As I understand one of Kripke's insights, what determines the reference of a name *N* in the actual world is the facts about the beginning of the causal chain that led to the use of the name *N*. The referent is the individual that played a certain characteristic role at that point, suggested by the role of the child baptized at a baptism. $C_{Gödel}$ would be the condition of *x is a person the name "Gödel" was introduced to stand for, at the beginning of the causal chain that leads to the use of that name*.[2]

Note that the mechanism here is a general mechanism for for proper names, applied to the particular name "Gödel". Kripke sketches an account of a *relation* that obtains between names and their referents. We get a *condition on*

[2] Given that more than one person has been named "Gödel," it would be better to say "our use". Alternatively, we can follow the advice of sages like David Kaplan and Ken Taylor and take each of the Gödels to have a different name. I discuss such issues in *Reference and Reflexivity* (CSLI Publications, 2002). I ignore them here in the spirit of keeping to the essential issues involved in my objections to Chalmers.

referents by instantiating that relation on the name in question. This means the condition for any particular name will be *type-reflexive.* That is, the condition of being the referent of a particular proper name *N* is being related in a certain way to *N* itself. The term "type-reflexive" isn't as common as the term "token-reflexive". But the phenomenon is as important. Associating type-reflexive conditions with expressions is an important aspect of linguistic knowledge. It is a necessary part of learning any expression. The person who first hears the term "Gödel" and recognizes it as a proper name will realize that it stands for a person connected in a certain way to *it*; that is, to that very type. As the person learns more and more about Gödel, this new knowledge will supplement, but not displace, the knowledge that "Godel" refers to the person connected in a certain way with to it.

Suppose, however, that we adhered to the sort of descriptive theory Kripke criticizes. In that case our candidate for $C_{Gödel}$ would be something like: *x proved the incompleteness of arithmetic*. This is *non type*-reflexive mode of presentation. It seems that this would be an unlikely candidate for $C_{Gödel}$, given what Chalmers tell us about primary intensions. It is a condition that Gödel was known to satisfy, by some people for some of his life. It's not a mechanism, or a function modeling a mechanism, by which "Gödel" came to stand for, or continues to stand for, Gödel. N one thinks that "Gödel" refers to Gödel because he proved the incompleteness of arithmetic. The name "Gödel" would have referred to him even if his proof had been faulty, or he had left mathematics before investigating the issue at all.

So, there are some good reasons for thinking that if we follow Chalmers directions (look for the mechanism, appreciate Kripke's insights) we will be led to a type-reflexive account of primary intensions. After all, it seems to me that there are at least three of Kripke's insights relevant here--- expressed here in my own way:

1. That proper and common names have rigid intensions. The value of the intensions of such names is the thing they designate in the real

world, even relative to other possible worlds that arise in ordinary counterfactual thinking and the semantics of modal sentences.

2. That there is often no non type-reflexive mode of presentation associated with the name by individual or community that picks out the actual world referent. This is one argument, or part of the argument, in favor of 1----not only do our intuitions about what is said and counterfactual truth-conditions argue in favor of rigid designation, but also in there is often no non-rigid alternative subject matter condition that fixes the reference.

3. The connection between name and referent does not become mysterious in virtue of point 2. In fact the connection between name and object can be a causal-informational one, rather than a non type-reflexive mode of presentation associated with the object.

In practice, Chalmers seems to recognize only point 1. He assumes that there is a non-reflexive mode of presentation associated with each term, and that is the mechanism that determines its actual world reference. But he gives us no reason to think that this is so.

Now let's look at "water". Again, we have a choice between the a causal-informational type-reflexive account of the sort I would favor,

> C_{water} *is being the stuff that the word "water" was introduced to stand for and keep track of at the beginning of the causal-historical-informational process that led to its use.*

and an account that is based on a non-reflexive mode of presentation the individual or community is supposed to associate with the name:

> C_{water} *is being the clear drinkable liquid that is found in lakes and rivers.*

There is a large assumption seems to slip in here, which leads Chalmers to favor the second sort of account:

> for any name N of category C (proper name, common name, etc.) there is a non-type-reflexive condition ϕ on the appropriate domain of objects (individuals, sets, etc.), so that the extension/reference of N is the object x such that $\phi(x)$, and ϕ is the condition a person grasps who understands the meaning of N.

This is a much stronger assumption that the following:

> for any name *N* of category *C* (proper name, common name, etc.) there is a relation ***P*** between names of category *C* and the appropriate domain of objects (individuals, sets, etc.), so that the extension/reference of *N* is the appropriate object *x* such that ***P****(N,x)*, and users of names in category *C* are attuned to this fact, although seldom explicitly conscious of it..

I see no reason to make the stronger assumption. Indeed one lesson I would draw from Kripke is that we should not make it.

(B) ***If we follow Chalmers' practice rather than his explanation, and take primary intensions to be non-reflexive modes of presentation, it is totally implausible to equate conceivability with contingency of the primary proposition***.

Suppose, for the sake of argument, that we accept that names and concepts are associated with non-reflexive modes of presentation. We then surely want to object to principle (2) of Chalmers' argument.

Suppose a student of mid-eastern history and literature boldly (but falsely) hypothesizes,

(5) Jonah is King Tut.

Although false, (5) is conceivable, in the sense that it is not self-contradictory and its truth or falsity cannot be determined a priori. The secondary intension is surely impossible. What is the primary intension?

We are assuming that primary intensions are *non type*-reflexive modes of presentation. But how can we be sure that the requisite modes of presentation exist? That is, that there are *non type*-reflexive conditions associated with the names, either individual or community-wide, that determine the referents of "Jonah" and "King Tut", and are how we think of these fellows when we hear and use their names? If not, given our assumption, we have no primary intensions, and so, if we accept premise (2) of Chalmers' argument, nothing is really conceivable about Jonah and King Tut. But it seems to me that it is conceivable

that was no Jonah and no King Tut; it is conceivable that there was a Jonah and a King Tut, but that none of the things we believe about either of them are, in the main, true, and it is conceivable that King Tut and Jonah existed, and were the same person. These are conceivable, because they have to do not merely with the subject matter possibilities, but also with the way our language and thought fit onto the world.

Implicit in Chalmers's approach is the principle that conceivability only has to do with the ways the world our thoughts and words are about can be, and not the way our thoughts and language fit into and onto the world. That is, implicit in his approach is what I call the subject matter fallacy. Whether a fallacy or an insight, it is definitely a pillar of Chalmers' system.

Let's turn to water. For Chalmers C_{water} is being the watery stuff, which is short for being the clear drinkable liquid found in lakes and rivers. That's the *non type*-reflexive mode of presentation for water. For cases like this one, in which there are more or less plausible *non type*-reflexive modes of presentation, is (2) plausible? No.

It is conceivable that water is the dominant liquid found in rivers, but not the dominant liquid found in lakes. And it is conceivable that water is the dominant liquid found in lakes, but not the dominant liquid found in rivers. That is, it is conceivable that water is not the watery stuff, for there is no stuff that meets all the criteria or even the weighted sum in our stereotype or usual set of criteria for water. One might want to pass off this last conceivability as merely the (secondary) possibility that H_20 isn't what is found in our lakes and/or rivers. But it's not that. Someone might conceive that water is found in our lakes, but not in our rivers, and is not H_20. Some theorist might hold that the liquid in rivers is not water. It undergoes a transformation when it flows into lakes, or reservoirs, or is removed in small quantities, as when we drink from a river. And our theorist might think the whole H_20 business is based on a confusion of the relation of scientific kinds to the natural kinds of the human world. We could not refute such a bold theorist a priori. I agree with Chalmers that there are more

conceivabilities than may be found among the secondary intensions. But his own theory of conceivability is limited, by his rejection of type-reflexive primary conditions, to subject matter possibilities, and I think this is much too limited. Given this limitation, there is no reason to accept principle (2).

(C) Even if we ignore (A) and (B), there is a problem with Chalmers' treatment of sensation terms like "pain". He doesn't choose, for the primary intension of "pain," the actual world mechanism of reference. Nor does he find some non type-reflexive mediating mode of presentation. He simply takes pain itself to be the primary intension of pain. This choice is unmotivated, and in fact quite baffling.

The remainder of Chalmers's argument turns on his claim that with phenomenal terms the primary intension is the same as the secondary intension. If we think that "pain" is a rigid designator, then the secondary intensions will be a function from any world to the state of pain. So too for the primary intension, according to Chalmers. Basically, pain itself is the primary intension of "pain". Where does this come from? How does it fit Chalmers's explanation of primary intension, or his practice in other cases? Pain is not a mechanism that determines pain to be the referent of "pain". How about the idea that pain is its own mode of presentation? This may be momentarily appealing, but it will not bear much weight. Behind this idea is the truth that when we are in pain, we are aware of pain directly, not in virtue of some other sensation causally downstream from pain. Having a sensation of pain is not like having a taste of bourbon. The latter is something caused by the bourbon entering our mouth, engaging our tongue, and initiating processes in our nervous system. The sensation of pain is just the pain.

But this does not mean that the sensation of pain is its own mode of presentation. We must keep in mind that in the term "mode of presentation" is a philosophical term of art, rooted in translations of Frege, where the intended idea is not "the way things are presented in perception" but something more general, including, and in fact more focused upon, how things are presented in *thought*.

The mode of presentation of pain is what is involved in *thinking* about pain, and that is what the primary intension should be modeling. The sensation of pain is the sensation we *have* when we are in pain, but it is not how we *think* about pain. We need no mode of presentation of pain to *have* pain, and we need not have pain to have a mode of presentation of pain, for we can and often do think about pain when we do not have it. I have thoughts about pain when I am not in pain; I use the world "pain" when I am not in pain; I remember and anticipate pain, imagine it others, talk about it, read about it, and so on.

I *can* have a pain, attend to the pain, and I say and think "*that's pain"*. It seems in this case there is a very intimate connection. Later I use the term "pain," and I am entirely clear what counts as pain, for I remember what is was like to feel pain. Still a very intimate connection. Much more intimate than the connection between "bourbon" and bourbon, for example. There is no intermediary in the case of pain, that plays the role of the look and taste of bourbon. But, still, a concept is a concept, not a sensation, and a word is a word, not a sensation.

Here we have three things: a word "pain," involved in statements about pain, an idea or concept, involved in thoughts about pain, and pain itself, a type of sensation. What makes the word or idea stand for *that* sensation? We can just say, "the sensation is directly linked to the word". We can emphasize that "pain" is not only a rigid designator of pain; it is as directly associated with pain as a word can be with a property or state. But these remarks are ways of *excluding* certain patterns. Saying what the connection is not, is not to say what it is.

The directions for identifying the primary intensions were to look at what fixes the designation of the word or idea in the actual world. *Well what does*? Let's stipulate that the word "pain" and some idea, some component or aspect of thought, are associated, *as directly as can be*, with a type of sensation. What does that come to? Here is a view, consistent with the data, accommodating to all the directness just emphasized: In this case, as in all others, the word designates, and the idea is of, the type of sensation it is introduced, and used, to track—to store

information about, to communicate about, and to guide actions and plans concerning. The various degrees of directness and indirectness all come *within* that general pattern, and the pain-"pain" connection doesn't escape it.

But that connection, between "pain" and pain, or my idea of pain and pain, will be contingent. To be sure, if we *fix* the connection between the word or concept and what it stands for, we will have pain before us, and not any condition on or description of pain. But that's the *secondary* intension. The idea is of, and the word designates, the type of sensation they were introduced and are used to keep track of and exchange information about. This state, that the word "pain" is used to designate and communicate about, is one that we can be aware of and directly attend to. The concept the word expresses is normally tied to memories of being in that state, memories of the most intimate sort, with that uncanny resemblance of which Hume speaks to that of which they are memories. We use the concept to anticipate future pains; thoughts containing it provoke fear in us, when we contemplate our own pains, and sympathy, when we contemplate the pain of others.

We may have all sorts of false thoughts wrapped up in our concept of pain, just as we might have false thoughts wrapped up in our concept of Gödel, or Moses, or Aristotle, or heat, or gold. There may be no sensation that fits the totality, or even the weighted sum, of the stuff we believe about pain very well. But "pain" stands for that sensation it was introduced to stand for, namely, pain.

Following Chalmers' directions for primary intensions, then, but not his example, we seem to find a condition that the state of pain only contingently fits: being the sensation "pain" is used by a community to communicate about.

If we follow Chalmers' directions for primary intensions, we won't pick pain as the primary intension of pain. As we saw, in practice Chalmers himself doesn't look for the reference-fixing mechanism, but for a *non type*-reflexive identifying condition. The sensation of pain doesn't fit this conception of primary intension either. Pain is a type of sensation. It is a phenomenal property of brain events, or of brains, or of individuals with brains, depending on how we want to

set things up. It is a type of sensation, but it is not a type of types of sensation. The sensation of pain is not a *non type*-reflexive (or reflexive) mode of presentation of the sensation of pain. Since it is neither the mechanism that determines that "pain" is of pain, nor a *non type*-reflexive mode of presentation of pain, Chalmers choice of pain as the primary intension of "pain" is baffling.

If we follow Chalmers in this choice, we will have more reason to give up (2). When we get away from the dramatic sensations, like pain and the color red, it is not hard to imagine getting confused about them. Recall the example about the sensation of smelling cinnamon (call it *c*) and the sensation of smelling nutmeg (call it *n*) in the Précis. Use "cinnamon" and "nutmeg" as names of the smells. Then, following the pattern of "pain," *c* will be the primary intension of "cinnamon" and *n* will be the primary intension of "nutmeg" on Chalmers' treatment. But is conceivable that (the sensation) cinnamon is (the sensation) nutmeg. Someone might hypothesize this, when his memories of the smells had grown dim. You couldn't refute him a priori. But, if *c* and *n* are the primary intensions, the primary proposition is *that* $c = n$, which is necessarily false. But then (2) is false, and conceivability cannot be equated with truth or contingency of the primary proposition.

(D) Suppose, however, we swallowed all of this too. The argument still doesn't work. That is, we make it to step (3).

(3) If $P\&\sim Q$ is primarily possible, materialism is false.

Here P is the complete microphysical truth about the world. *P* is something that has a secondary and primary intension, so it is a statement or sentences. If there are any problems about how long this sentence might be, and such things, we'll set them aside. If we thought, in a sort of old-fashioned way, that the microphysics of the world was a matter of basic particles in motion at times, then P would tell us where all of those particles are at each moment and what they are doing there. Every statement that logically supervenes on those facts about basic particles would be logically implied by *P* plus appropriate definitions.

The secondary intension of *P* comprise possible worlds that are microphysically the same as ours. What about the secondary intension of *P*? *P* is presumably a sentence whose vocabulary is drawn from microphysics. It isn't completely obvious to me what the primary intension should be on Chalmers' principles. I don't see, however, any reason to think that there are any terms of microphysics that work the way Chalmers' thinks phenomenal terms do; that is, I don't see why there would be any expressions in *P* that have the same primary and secondary intensions. Take a term like "charm". This stands for a property of quarks. Suppose that the secondary intension of "charm" is a rigid designator; that as a part of physical theory, "charm," it stands for a certain property rigidly, just like H_20 does. The primary intension will presumably be a matter of the criteria we use to apply "charm" in the actual world, and will not be rigid.

Q is a phenomenal truth. Let *Q* be "Pain occurs in brain *B* at *t*, "where "*B*" identifies the brain in terms of its spatial position at *t*. Admit that it is conceivable that *P* and ~*Q*. That is, agree that *Q* does not logically supervene on *P*. We grant (2) for the sake of argument. We grant for the sake of argument that the primary intension of "pain" is pain. How are we to squeeze (3) out of this?

(3) is equivalent to

(3A) If materialism is true, then P&~Q is not primarily possible.

Clearly, whether we accept 3A depends on what we take materialism to be. Let's confine ourselves to phenomenal properties. I distinguish between

The logical supervenience theory: Phenomenal properties logically supervene on physical properties

The identity theory: For each Phenomenal property P, there is a physical property Z, such that P=Z.

It is fairly easy to see why one who held the logical supervenience theory might have to accept (3A). Presumably logical supervenience is due to a connection

between the supervening and supervened upon concepts, and this would be reflected at the level of primary intension.

But I see no reason why an identity theorist should accept (3A). Given Chalmers' theory that the primary intension of "pain" is pain, plus materialism, the identity theorist can conclude that if *P* is true, *Q* will be true, and ~*Q* false. This would be because the *secondary* intension of *P* requires that a given physical state *T* occur, while the primary intension of ~*Q* requires that the very same state does not occur. But it does not follow that the *primary* intension of *P* requires that the physical state *T* occur. So it does not follow that *P&~Q* is not *primarily* possible. We would need an argument that the relevant physical term, "*B3*" say, has the very state it refers to as its *primary* intension. If not only the secondary intension, but also the primary intension, of "B_3", were a rigid designator, then the conceivability of *P&~Q* would pose a problem for the identity theorist. But as far as I know, Chalmers' doesn't think this is so, much less argue for it.

CHALMERS ON THE ZOMBIE ARGUMENT

Chalmers says that the Zombie argument rests on two premises, that Zombies are conceivable, and that what is conceivable is possible. He adds,

> The first premise rests partly on prima facie conceivability intuitions that many share, and partly on deeper considerations concerning the absence of any conceptual linkage between microphysical concepts (which are structural-functional in nature) and phenomenal concepts (which are not). In both cases, whether or not these premises are correct, their support presupposes nothing about epiphenomenalism.

I agree that the Zombie world is conceivable, but it does not follow that it is possible. The conceivability is due to the lack of conceptual linkage between microphysical concepts and phenomenal concepts which Chalmers notes. This lack of conceptual connection does not rule out identity of referent. One could work hard to construct a theory of conceivability that makes the argument work, but the antecedent physicalist need not accept such a theory of conceivability.

Suppose that *Es* are a nonempty subset of the *Bs*. If so, a world with all of the *Bs*, but none of the *Es*, is not possible. It will be conceivable, however, if there is nothing in our conception of *Es* that requires that they are *Bs*. The step from conceivability to possibility is not valid, in the absence of further assumptions. Chalmers assumption that *if* experiences are physical states, it is because they logically supervene on physical states, is such an assumption.

I've changed my mind on one point, as a result of discussions with Murat Aydede and his colleagues and students in his seminar on consciousness at the University of Florida. I no longer think that the Zombie argument presupposes epiphenomenalism. It seemed to me when I wrote the book that if there was a world physically indiscernible from ours, but with no experiences, that would mean that the experiences in our world didn't have any physical effects. But that isn't quite right. It could be that the experiences in our world are redundant. They have effects, but for each and every experience in our world that has a physical effect there is some other state that would have brought about that effect if the experience hadn't. This hypothesis, inspired by recherché considerations in the literature on counterfactual analyses of causation, may be even less plausible than epiphenomenalism, however, so it doesn't change my argument very much. If someone does not think that experiences are *either* epiphenomenal *or* redundant, one will *not* accept that the lack of contradiction in a statement of the Zombie world shows that it is really possible.

CHALMERS ON THE KNOWLEDGE ARGUMENT

Chalmers' basic criticism of my treatment of the knowledge arguments is that I ignore the relevant sort of knowledge that Mary gains:

> …I distinguish three sorts of phenomenal concepts: pure phenomenal concepts (such as *R*), demonstrative phenomenal concepts (such as $this_i$) and relational phenomenal concepts (such as *the sort of experience typically caused by red things*). Perry's discussion seems to acknowledge only two

> sorts, the demonstrative and the relational; or at least, it seems to assimilate the pure phenomenal and the demonstrative with each other.

He also says that I analyze phenomenal knowledge as a sort of indexical knowledge. He seems to think that I take Mary's concept of the sensation of red, that is involved in her new knowledge when she steps from the room, as simply *being the sensation to which she is then attending*. This is certainly a property the sensation has at that point in her life, and it is this property of it that makes is suitably identified by the words, "this$_i$ sensation". It is certainly, on my view, an important property of the sensation which is part of the explanation of her learning something new about it. But it is not my candidate for Mary's concept.

> Let me begin by considering Chalmers' example of Jack and circles:
>
> Jack has never seen a circle before, and…on seeing a circle for the first time, he acquires the qualitative concept of circularity. He will then be in a position to think the qualitative thought, *Jill's favorite shape is a circle* and to think the substantive demonstrative thought, *this$_S$ is a circle*.

As I understand this example, we start out before Jack has a concept of a circle. Probably he has seen pennies and nickels and other circular objects before. But now he is being taught his shapes. He gets the hang of it, and after a while he says, when confronted with a circle, translating into the philosophical,

> (7) This shape property is the property of being a circle

He can also talk and think about circles when he is not seeing them. He can leaf through a picture book and find a circle if so instructed, for example. Now, according to Chalmers, he has the "qualitative" concept of a circle. Let's pause on "qualitative". Being circular is a quality in the sense of being a monadic property of objects. Jack's concept is qualitative in that he recognizes circles by looking at them, not by which books they are in, or who is pointing to them, or some relation they have to something else.

> He might at some point learn

(8) Jill's favorite property is being a circle

Note that we wouldn't ordinarily call this as a second concept of a circle, but as something he learns about Jill and circles.

On the view of concepts I developed in *KPC*, when Jack gains a concept of being a circle, there is a new structure in his mind, which we can think of as a bit like a file folder. It is assigned to the property of being a circle, because that is the shape to which he was attending, to which his teacher was attending and so forth. In other words, even for concepts of properties, I give a causal/informational account of what the concept is *of*. This is not to say I give an *indexical* account. I do give an account that deals with our ability to use indexicals in referring to objects, and our inability to do so, when we cannot---something any viable account must do. Associated with the concept---in the file folder, so to speak---will be various things. In Jack's case, this will include the word "circle", and the sort of idea that having an impression of a circle can give rise to (to use Hume's vocabulary). We might think of this as an image, or maybe a program for producing an image, or in a lot of other ways, which may prove to be more or less adequate, given what scientists of various persuasions have learned and will learn. This new structure is involved in a number of activities Jack can now do: thinking about circles, finding circles in a book, trying to draw circles, and understanding what it means for circles to be Jill's favorite shapes. Eventually, he will study circles in geometry, and learn that they are closed lines on a plane all points of which are the same distance from some point.

The fact that Jack learned what a circle was by ostension does not mean that subsequent thoughts of the form of (7) are trivial. He might have forgotten what circles look like, or he might have astigmatism so that circular objects no longer look circular, and so forth. It would be a bit odd to forget what circles look like. We expect a sighted person who has the concept of a circle to be able to recognize circular objects on sight, at least in favorable circumstances. Still, someone a bit slow, like me, may have to occasionally fiddle around with a pencil to remember what trapezoids or scalene triangles or regular pentagons look like.

I might forget what a scalene triangle looked like, even though I had the concept, and knew enough about them to figure out what they looked like.

Now consider Jack's learning about the experience of seeing red. One ordinarily sees colored things for a while, and then learns one's colors---about the same time one learns one's shapes. A bit later, perhaps in the late primary grades or in junior high, one may be ripe for learning the concept of the *experience* of seeing red. That is, one is induced to reflect on the fact that when we see red things, there is something characteristic going on in us; it is *like* something to see red things---and it sometimes happens when we aren't actually seeing red things. So Jack can think,

(9) This experience is the experience of seeing red.

He can not only have red experiences, and anticipate having them, but explicitly think about them. he is in a position to think such things as,

(10) The experience of seeing red is Jill's favorite experience

and the like.

The experience of seeing red seems to be something a person has or doesn't have; it is a quality, not a relation. Jack's concept of this experience seems to be qualitative, in this sense; he is thinking of the experience as something one has or doesn't, not as something one has relative to some other object. His concept of the experience of seeing red is qualitative in the same way his concept of circularity is. He just thinks of himself, or others, as having the experience or not, *not* as having it relative to one thing, but not relative to another, just as he thinks of being circular as something a plane figure has.

When Jack comes to have the concept of the experience of seeing red there is a new structure in Jack's mind, that we can think of a sort of like a file folder. It is *of* that experience of seeing red, because that is the experience the structure was formed to keep information about, help recognize, and the like. The structure contains various things, most importantly what I call a Humean idea of the

experience--- an idea which normally derives from having the experience, and is subsequently a central part of thinking about, remembering and anticipating such experiences. Thinking of having the experience of some kind in this way is not having the experience, but it is in some uncanny way like it. Usually the same kinds emotions attach to the thinking as to the having, although in a milder form. It is usually pleasant to anticipate or imagine having pleasant experiences, and unpleasant to anticipate or imagine having unpleasant ones, for example.

Having an experience of seeing red is a property Jack has himself. He is not aware of having the experience, in virtue of having some further experience caused by the experience, as he is aware of the redness of, say, a tomato by being caused by the tomato to have certain sensations.

I think both Chalmers and I think of the words "the experience of seeing red" as designating a property or a state that brains are in at times. The only difference, on that score, is that he does not think it is a physical property of brains, but a *non*-physical one that causally supervenes on the physical properties. We both think that Mary has an experience when she leaves the room. We both think this makes a big difference in her concept of red. She knows what it is like to see red; she knows what seeing-red experiences are like. In thinking of the experience in this new way, she is not thinking of it as "this experience" or as "So and so's favorite experience" or anything like that. There is nothing in my account that directly conflicts with the immaterialist view, except my working assumption, that experiences are physical.

Thus, I'm not sure why Chalmers thinks that I do not give an account of Mary's qualitative concept of the experience of seeing red. I was at some pains to do so. I distinguish between the reflexive contents of

This experience is this experience

This experience is the experience of seeing wow

This experience is the experience of seeing red

The experience of seeing wow is the experience of seeing red

just as I distinguish between the contents of

This shape is this shape

This shape is the shape (I'll call) squiggle

This shape is circular

The shape squiggle is the shape circular

It is quite conceivable that we could raise Jack in a situation in which he learned a lot about circles without ever seeing one. He could learn the geometric definition of circles, and learn that many coins of the sorts others are allowed to see and possess are circular, and still have no idea what a circle would look like. I don't mean it is impossible that he could figure it out, but he might not. Someone like Jack, coming out of our circle-less room, and seeing a quarter, might learn

That shape is circularity

What we learn, I claim, is to be found at the level of reflexive contents. Similarly, Mary coming out of the Jackson Room and seeing a tomato, learns

That color is red,

and what's more:

This experience is the experience of seeing red

In both of these cases, too, we need to appeal to reflexive content to get at what is learned.

ROSENTHAL ON THE KNOWLEDGE ARGUMENT

Rosenthal thinks that my formulation of Jackson's argument begs the question. I don't think it does. Moreover, it seems to me the formulation of the argument he ascribes to Jackson is totally unconvincing. I'll deal with this second point first.

The key question is whether Mary is allowed, in the Jackson Room, to know about the existence of color-qualia in other people. Jackson at one point writes as if she would not be allowed, for he says that she will be surprised to learn, when she leaves the room, that others have been having such experiences. I found this restriction puzzling, but Rosenthal thinks it is crucial to the argument:

> By hypothesis, Mary's textbook knowledge exhausts the physical nature of seeing red. But the antimaterialist requires also that this knowledge be exclusively physical, since Mary's learning about something nonphysical from her textbooks would obscure any new nonphysical knowledge she might get on first consciously seeing red. To ensure that her textbook knowledge is wholly physical, Perry stipulates that Mary's textbooks take no "position on whether *Qr* is a physical aspect of the brain or some other kind of property" (99).
>
> But that's not enough. If *Qr* is nonphysical, Mary's textbooks teach her about something nonphysical whether or not they describe it that way. Perry stipulates that Mary's textbooks teach her only about the incontrovertibly physical features of *Qr*, such as its causal interactions with physical stimuli and behavioral responses. But even that isn't enough; unless we've established independently that *Qr* itself is physical, Mary's learning about it may well be learning about something nonphysical.
>
> Arguably this is Jackson's conception of the knowledge argument. As Jackson describes things, before consciously seeing red Mary not only lacks knowledge of what it's like *for her* to see red; she doesn't even know that there's some subjective character *others* have that's special to consciously seeing red.... She knows only how the relevant neural states causally interact with behavior and stimuli.

Rosenthal's idea is that if it's an open question whether Q_R is physical, Mary can't be allowed to know about it. But then I can't see how the knowledge argument is supposed to work. It seems to be this:

Mary learns all the facts that are uncontroversially physical in the Jackson Room.

She learns a new fact when she emerges.

So it wasn't an uncontroversially physical fact.

So…?

I can't see that anything follows from this.

As I understand the knowledge argument, the key is the new knowledge Mary gets when she first sees something red. Nothing in the way I set up the argument in any way detracts from this key part of the argument. I do not deny that she has new knowledge, but emphasize that she does, and try to analyze it.

My project in the book can be seen as a conditional proof. Assume physicalism. Provide an account of subjective characters on this assumption. The successful result of this exercise does not, of course, prove physicalism. It simply shows that on the assumption of physicalism, we can account for qualia; the physicalist need not deny qualia, and hence the existence of qualia is not an argument against physicalism.

Whether this convinces someone of physicalism or not will depend. If the neo-dualist arguments were the only reasons for doubting physicalism, it should. But there are many other reasons people might have for disbelieving physicalism. They might have religious beliefs that are incompatible with it. They might have a policy against holding any metaphysical beliefs. They might think that no coherent positive account of basic physical dimensions such as space and time has been given. They might believe that facts about extra-sensory perception show that physicalism is false.

My own bias is in favor of physicalism, for not quite the same reasons David Lewis gave in his "An Argument for the Identity Theory." He said that mental states played some of causal roles, as a matter of definition, and as far as he could see the evidence was overwhelming that physical states played all the

causal roles, so the mental states must be a subset of the physical. I differ on the basis for the first premise. I think mental states play some of the causal roles, but I doubt very much that it is a matter of definition.

I don't expect to be able to persuade many philosophers of physicalism who are not already pretty sympathetic to it. Some philosophers find it easy to imagine that mental states don't play some of the causal roles, and others don't think that physical states plays any of them either, since they take the whole notion of causation to be problematical. There is not much in my book to persuade these folks to change their view. What I thought I could do, and I must admit, think I did, is show that certain arguments *against* physicalism simply aren't any good.

ROSENTHAL'S *HOT* THEORY

The title of *KPC*, of Chalmers book *The Conscious Mind*, and of many other books and articles in the literature suggest that there is one phenomenon, consciousness, at issue. But that's not actually the position I take in *KPC*. There I hold that there are some states *it is like something to be in.* To be in such a state is to be sensate, to have experience. It is indeed mysterious how there can be states it is like something for their possessor to be in. But I don't call that particular mystery "consciousness". I reserve that term for awareness of these states. It seems fairly clear to me that creatures can have experiences---that is, be in states it is like something to be in---without being *conscious* of those experiences. Humans can be and regularly conscious of the states they are in; they notice them, classify them, name them, remember them, anticipate them, try to avoid them, and, in general *think* about them.

So far, this sounds similar to Rosenthal's Higher Order Thought (HOT) theory. He thinks consciousness involves higher order thought, and I agree. What I worry about in his theory is not the consciousness but the sentience. I think consciousness involves being aware of sentient states. But it seems like

Rosenthal may think that there is no sentience, there is no "what it is like," until there is consciousness. That doesn't seem right.

Consider a mouse caught in a trap, with a broken leg, trying to get out. I think the mouse is in pain. It is in a state it is like something to be in. I suppose it is a pain of the sort I feel, or similar to it. I feel sorry for the mouse. I doubt that it is conscious of the pain. I don't think it is thinking about it. I see no reason to suppose that in addition to having the pain it is aware of having the pain. It is *in* pain, and that's bad enough.

Now, as I understand Rosenthal, I can say quite a bit of what I want to say about the mouse. I can say that he is not conscious of the pain, and in that sense the mouse's pain, unlike a lot of human pain, is not conscious: the mouse has the pain, but he does not think about the pain, conceptualize it, have concepts of it, attend to it. He has the pain, but is not consciously aware of the pain. Rosenthal could let me use the term "sentient" for states that would be conscious, the sort of states of which one could be consciously aware.

So far so good. But on Rosenthal's theory, can I say that there is something it is like for the mouse to be *in* pain, even though the mouse is not thinking about the pain? Or does the *what it is like* property come with the consciousness? This is the key question I have about Rosenthal's theory. If the former, I think I can agree with it. But it won't provide an analysis of the key property that makes for sentience: being a state it's something like to be in. If the latter interpretation is correct, then clearly Rosenthal has something I don't, an account of "what it's like" properties, in terms of higher order thoughts. I have no account of them at all. I acknowledge that there are such states, and I argue that that we have been given no good reason to suppose they are not physical. So philosophically, Rosenthal's view, on the second interpretation, has an advantage over mine. But so interpreted view doesn't fit what I believe about sensations and thoughts.

It seems to me that having sensations, being in states that it is like something to be in, fits into an evolutionary intelligible story. Nature uses states it is painful or pleasant to be in to control behavior, so its creatures will stay alive

long enough to reproduce, and be inclined to do so. A system where dangerous situations hurt and ones that enhance reproduction are fun has a lot to offer nature, even if it does not include any awareness of or thoughts about these states.

Human life is something else again. We think incessantly about our sentient states, and work to be in some and stay out of others. All of this awareness and thinking presumably provides, or at least at some time provided, some further evolutionary advantage over merely being in such states, or was an offshoot of something that did. It seems very natural to me to divide sentience and consciousness, then, using the terms as I do. If the Higher Order Thought theory allows me to do this in such a way that sentience is as robust a phenomenon, as central in the life of animals including humans as I think it is, then I would be very drawn to it for its philosophical advantages, mentioned above.

CHURCHLAND

I appreciate Paul Churchland's comments. I am happy with either agreement or understanding; getting a good helping of both from the same author is very gratifying.

I can't say much here to compare our different but possibly complementary takes on the "two ways" response to the knowledge argument. But I will take up Churchland's invitation to say a bit about modality. There seem to be two aspects of contemporary thinking about possibility that annoy Churchland. One I'll call *modal realism* and the other *de re possibilities.* I'll explain the philosophical grounds for my relative lack of annoyance.

Modal realism is often associated with David Lewis's ideas[3]. Lewis's view involves the following: i) There are infinitely many concrete universes other than our own; ii) Everything that happens in any of them is possible, iii) Everything that is possible happens in at least one of them iv) to say that it is possible that *P*

[3] David Lewis, *On the Plurality of Worlds*.Oxford, Blackwells, 1986.

is simply to say that there is a world such that *P*, and v) Each such universe is actual relative to itself, and possible relative to the others, which is the same status our own universe has.

I see no reason to believe i). If I did believe i), I would accept ii). But I would still see no reason to accept iii). And I don't see any reason to accept iv). That is, even if I accepted i)-iii) I would think it to be a sort of miracle, and easily conceivable that some possibility might be missing. If I accepted i)-iv), I would be inclined to accept v).

On most of this I sense agreement with Churchland. But in his dismissal of Lewis's bold hypotheses, he also dismisses other more plausible forms of modal realism---that is, other theories that take there to be real possibilities. Robert Stalnaker's view does not postulate other concrete universes, but merely "other ways the world might be"[4]. This view is quite different than Lewis's metaphysically. By calling Stalnaker's ways the world might be "possible worlds" we can think about possible worlds analyses of various things, such as counterfactual and conditionals, and incorporate many of Lewis's insights about these matters, without adopting his views. A way the world might be can simply be a particular kind of property our world might have or might not have had. These properties determine answers to every more specific issue that arises about the world. Such properties are very helpful ways of thinking about possibility. I have some skepticism whether this is the best we can do --- see "From Worlds to Situations,"[5]---but at any rate I don't see anything metaphysically objectionable in the way that Lewis's alternative concrete universes are. So modal realism, of this sort, doesn't annoy me as it does Churchland.

Now let's turn to *de-re possibilities and necessities*. Suppose that by noon of Friday of creation week God had decided which properties and relations should be instantiated and co-instantiated. He wanted simply to say "Let it be:" followed

---

[4] Robert C. Stalnaker, *Inquiry* (Cambridge, MIT Press, 1984).
[5] See "From Worlds to Situations," *Journal of Philosophical Logic* 15 (1986): 83--107. Reprinted in John Perry, *The Problem of the Essential Indexical*, (Stanford: CSLI Publications, 2000).

by a rather long Ramsey sentence, be done with it, and take Saturday off. But he had a worry (this account does not follow the Bible in every detail). The worry was whether, in addition to deciding everything he had decided, he needed also to decide *which objects* were to do the instantiating. Upon reflection, it seems to him that he doesn't need to decide this, and indeed it makes no sense. He says "Let it be…" and as we can all see it worked.

This picture suggests the concept of what I'll call "pure possibilities". Each alternative way that God could have chosen for properties and relations to be instantiated and co-instantiated constitutes a way the world could be, or at least could have been. There are no other possibilities.

Now consider Harry, fourth grade teacher figuring out a seating chart. He has twenty-two desks and twenty-two pupils. He is given the desks by the school staff members, who bring them and put them in the room. He is given the pupils by the principal, who gives him a list. Each of the seating charts is a possible way the room might be. Each of the seating charts is a way of arranging the desks and pupils that are *given*. Each of the charts is a possibility *for a given domain of objects*.

When philosophers think about grand issues like freedom and determinism, good, evil and the best of all possible worlds, and the nature of basic laws, they usually have in mind pure possibilities. These possibilities are not individuated in terms of the objects that instantiate them or would instantiate them. God didn't have two choices, one with me a twenty-first century philosopher and Butch Casssidy a ninetheenth century outlaw, and the other with me as a nineteenth century outlaw named "Butch Cassidy" and with all of the other properties Butch had in the real world, and Butch as a twenty-first century philosopher with my name and all of my properties. Or so it seems to me.

In everyday thinking about possibility, however, we are much more likely to think about possibilities for a population. What would the room be like if the sofa were there and the credenza here? What would happen if I move Elwood to Charley's desk, Charley to Myrtle's, and Myrtle to Elwood's? What would happen if Jim Carrey replaced Barry Bonds for the next series with the Dodgers?

Would our department be better if we recruited Quine or Sartre? And so forth. In reasoning about these possibilities, or using these possibilities to reason, we keep some properties of the individuals fixed (the baseball talents of Bonds and Carrey, the philosophical abilities of Quine and Sartre, the personalities of the pupils) and allow others to vary.

The necessities Churchland doesn't like involve possibilities, or lack thereof, for a population. Can the teacher suppose that Elwood *is* Charley?. Then he'd have an extra desk. But that's not a possible arrangement of his twenty-two students and twenty-two desks. Elwood is *necessarily* not Charley. But isn't that annoying? Where did that necessity come from?

I think the reflexive-referential semantics I used in *KPC* works well to get at the relations between pure possibilities and possibilities for a population, and this will lessen our annoyance and these necessities. Reasoning about possibilities for a population is something we do as occupants of the world, not transcendent creators of it. We stand in various relations to other occupants. Various relations are involved in various types of things we do with regard to these co-occupants. To think about Elwood, Harry uses the name "Elwood" which was assigned to him, memories he has of seeing Elwood on the playground, information and misinformation he has gotten about Elwood from students and other teachers, and so forth. He is taking all of that as fixed, when he thinks about changing Elwood's seat. Given that all that is fixed, he represents what would change if he moved Elwood with a singular proposition.

Thus the sort of thinking about Elwood that we represent theoretically with singular propositions, have to do with possibilities for domains. In this case, they are possibilities for Elwood rather than complicated pure possibilities. The reflexive-referential theory finds plenty of contents for Harry's thought, "If I move Elwood nearer to my desk, he will quiet down". The one most likely to be relevant takes all the things that Harry implicitly takes as fixed about the term "Elwood,": for example that it refers to *that kid*, the one who is talking now just as he usually does. This content gets at what Harry is taking as fixed, things he has

no control over, and things that he hopes to change. If he makes the changes, he will have changed the way properties and relations are instantiated and co-instantiated. He will have changed which pure possibilities will be instantiated. Making the de re possibility that *Elwood* sit nearer his desk into an actuality isn't an additional thing he will do, over and above effecting the way relations and properties are instantiated or co-instantiated. It is a way of getting at the changes that exploits what is taken as fixed to describe what is taken as up for grabs.

This method of getting at possibilities might even be useful for getting at God's creation, if say it took him several hours to get it all done. Suppose in the morning he figures out what happens up until Martin Luther is ten years old, and in the afternoon figures out the rest of the story, up through the end of everything in, say, 2009, when I retire. Observing this process, and describing God's thinking shortly after lunch, we might say, "Luther can become a carpenter, or a veterinarian for oxen, or a priest...." We are describing the options left open to God by his decisions before lunch.

We mere mortals also create. We design, describe and build actual things, and design and describe fictional situations. And we mix things up, as with Tolstoy's *War and Peace* which, it seems to me, is about the real Napoleon and the real Moscow, but also involves a number of pure fictions, who Tolstoy couldn't exploit his real relations to, to think about and get into his story. Use of singular propositions and the attendant necessities that come with them is useful in describing all of this, and leads to all sorts of interesting language games and odd ways of speaking and mental crams and in general provides a playground for philosophers---hardly a ground for annoyance for philosophers, although we can doubtless annoy many non-philosophers if we play too loudly. [6]

[6] I thank Murat Aydede and Gene Witmer for helpful comments on the penultimate draft of this essay.