PHILOSOPHICAL PERSPECTIVES

Philosophical Perspectives, 31, Philosophy of Mind, 2017
doi: 10.1111/phpe.12104

THE SIGNIFICANCE ARGUMENT FOR THE IRREDUCIBILITY OF CONSCIOUSNESS

Adam Pautz
Brown University

> Should [reductionism about consciousness] convince us that the line between the conscious and the non-conscious is of no great metaphysical significance?
>
> ——John Hawthorne (2006)

> What might a philosophy might be like that began to give up all reductionist dreams?
>
> ——Hilary Putnam (1992)

One of the most striking features of consciousness is the way it presents us with real or ostensible items, such as objects, properties, states of affairs, and events. Here I will focus on the presentation of properties. For example, suppose you look at a humble tomato. You are *conscious of* a round shape. You are also conscious of a color quality as filling a round region. These properties are present to your mind. As Russell (1912) put it, you are acquainted with them.

This conscious-of relation is in various ways significant. To begin with, it has dissimilarity-grounding significance. There is nothing else like it. When you are conscious of the color red, your relation to it is necessarily totally unlike your relation to any quality that you are *not* conscious of. It also has reason-grounding significance. For instance, if you are conscious of the color qualities *red* and *reddish-orange*, you are in a position to know that these qualities are similar. And you may have a reason to believe that there are items before you having those colors. Finally, it has a thought-grounding significance. It is a source of determinate intentionality. When you are conscious of a quality, you are able to easily and determinately think about that quality. In these basic cases, radical Quinean indeterminacy worries just do not get a grip.

In this essay, I use these facts to develop a new argument for a nonreductive view of consciousness. This argument will undermine the kind of reductive materialism defended by Lewis (1994), Sider (2011) and Dorr (2007), among many others. I call it the *Significance Argument*. The argument differs from the *Knowledge Argument* and the *Conceivability Argument*. In particular, unlike

these arguments, the Significance Argument does not depend on controversial *gap reasoning*. Rather, it is based on a series of interesting puzzle-cases that I call *multiple candidate cases*. In these cases, there is a multiplicity of physical-functional properties or relations that are candidates to be identified with the sensible qualities and our consciousness of them, where those candidates are not significantly different. I will argue that these cases show that reductive materialists cannot accommodate the various ways in which consciousness is significant. I also will argue that a nonreductive theory of the conscious-of relation can easily provide a very satisfying, unified explanation of the ways in which this relation is significant. It is nonreductive in the sense that it holds that there is no interesting "metaphysical analysis" (Sider 2011) or "identification" (Dorr 2016) of the conscious-of relation in physical-functional terms. Still, we will see that it does not require traditional dualism; it is compatible with a ground-theoretic version of materialism (Schaffer 2017). The particular nonreductive view I shall suggest can be viewed as a way of implementing the new *phenomenal intentionality program*. It also has similarities to Russell's older view of the foundational role of conscious acquaintance.

My plan is as follows. First, I describe what I consider to be the most promising approach to reductively explaining the conscious-of relation in physical-functional terms (§1). Then I argue that it does not work: because of multiple candidate cases, no approach of this kind can accommodate the ways in which the conscious-of relation is significant (§§2-4). Finally, I show how a nonreductive theory of the conscious-of relation can easily explain the ways in which this relation is significant while remaining compatible with a form of materialism (§5).

1. How Might Presence be Reduced?

Consider an experience of a purple oval between a blue sphere and a green cube (see Figure 1). Call this *the trio experience*.

A starting assumption of my essay is *external directedness*. By this, I mean a number of pretheoretical ideas. Necessarily, if anyone has the trio experience, she has an experience *as of* variously shaped and colored items in space. Furthermore, if anyone has an experience of this type, she is in a position to know certain things. For instance, she is in a position to know the timeless necessary truth that blue is more like purple than green. And she is in a position to know the

Figure 1. The trio experience

timeless necessary truth that the shape *round* is more like the shape *oval* than the shape *square*. Now we have strong theoretical reasons to think that these are truths about properties (Yi 2017). So external directedness implies that experience necessarily puts one in a position to know things about these properties.

These points apply equally to hallucination. For instance, suppose that Mary comes out of her black and white room, only to have the trio experience during a hallucination. Still, she counts as having an experience as of a purple oval, a blue sphere, and a green cube. And she is thereby in a position to know some things. For instance, she is in a position to know the timeless necessary truth that blue is more like purple than green. She is in a position to know what blueness is like. And, to repeat, this is a truth about color *properties*. True, she doesn't know that there are physical objects before her that have these properties. There are not such objects. But this doesn't prevent her from knowing something about the properties themselves. Strange as this may seem, hallucination can be a source of knowledge about non-mental reality, namely the nature of colors and shapes (Russell 1912).

This suggests that having the trio experience, whatever else it may involve, involves standing in a relation to certain color and shape *properties*. This relation has the following characteristics. Generally, two people have different experiences if, and only if, they bear this relation to different clusters of properties. Call this the *character-presentation connection.* (I say "generally" because there may be some few exceptions about blur, attention differences, and so on.) Also, when you bear this relation to some properties, it generally seems to you that there are items having those properties. Finally, when you bear this relation to a property, you are typically easily able to think about that property and predicate it of things in thought. Let us call this hypothesized relation the *conscious-of relation*.[1] (There may be different modes of conscious relations, corresponding to different degrees of attention, memory and imagery. I am operating with a toy model.) Thus conscious experience involves being *intentionally directed* at properties.

How is it that we are able to be conscious of properties in experience? In particular, can this relation be identified with some physical-functional relation? Or is it a primitive relation between minds and properties?

Sense datum theorists like Russell (1912) held that we are conscious of properties by being conscious of items that have the properties. Because of illusion and hallucination, these items cannot be ordinary physical objects. Russell (1912) held that they are "sense data" that are created by the brain but that are not located *in* the brain: they are located in a kind of private mental space and literally have certain shapes and stand in certain spatial relations. (For a somewhat similar view, see Peacocke 2008.) The conscious-of relation ("acquaintance relation") that we bear to such non-physical items is an *irreducible*, non-physical relation, according to Russell. There is something seductive about the sense datum view. If Mary has a super-vivid hallucination of some things, can't she be certain that there are some items for her to scan and explore?

However, the sense datum view provides a complex, non-reductive account of consciousness. It goes most naturally with dualism. True, it could also be combined with a strange form of materialism. For one might say that, necessarily, if Mary is in the right brain state, then this *grounds* the coming-into-existence of a new reddish and round thing, and grounds her acquaintance with it. However, this form of "materialism" would be just as complicated as dualism. Despite the recent enthusiasm for grounding (e. g. Schaffer 2017), brute psychophysical "grounding laws" add to the complexity of our theory of the world no less than brute, contingent psychological laws.

It is considerations like these that support a reductive approach to the conscious-of relation. By this, I mean an approach holding that that there is "identification" (Dorr 2016) or "metaphysical analysis" (Sider 2011) of this relation in physical-functional terms. As Levine says, "whatever acquaintance is, it can't be a basic relation; it must be constructible out of other, non-mental relations" (2006, 161). Following Cian Dorr, I believe that we have a good grip on the locution "to be *F* is to be *G*". A statement of this form can be called an *identification.* Here the "is" expresses a kind of identity, only it is flanked by two predicates rather than by two singular terms. As I use "reduction", to be a reductionist about *F* is to hold that there is a true identification of the form "to be *F* is to ... " where the right-hand is filled entirely by "physicalistically acceptable words". So a reduction of account of the conscious-of relation is one that identifies the dyadic conscious-of relation with a dyadic *physical-functional* relation.

If we want a reductive theory of the conscious-of relation, we must reject the traditional form of the sense datum view. What theory might we put in its place? One idea would be to try a reductive *internalist* account of the consciousness of properties. The idea is that the properties we are acquainted with, and that determine the character of experience, are always *neural properties of our own brains*. As Papineau puts it, "the only properties of conscious experience with which we can make introspective contact are properties*, intrinsic [neurocomputational] properties of subjects" (Papineau 2016; see also Block 2010, 24). But this violates the starting assumption of this essay: essential external directedness. For instance, having the trio experience surely essentially involves being acquainted with the properties *being round*, *being oval* and so on, which are just not neural properties instantiated in the head. Indeed, it was precisely for this reason that Russell (1912) thought it necessary to invent non-physical sense data distinct from brain-regions to be the bearers of perceptible properties (see Peacocke 2008 for a similar idea). But, again, the sense datum view is non-reductive.

So an internalist reduction of the conscious-of relation seems to be a non-starter. In my view, given the externally directed character of experience, any attempt to *reductively* explain presence in physical terms must be to some degree be *externalist*. Let's start with the consciousness of "primary qualities" like *shape*, and then we will turn to the tougher case of the consciousness of "secondary qualities" like colors.

Return to the case where Mary hallucinates a red and *round* item. Most philosophers are realists about space as we perceive it. Out there are things that are round-as-we-see-it. Normally, when Mary is conscious of roundness, she is in a brain state that is caused by an instance of this property. In the present case where Mary is hallucinating a round thing, she is in a state that is normally caused by an instance of roundness. This suggests an avenue for reduction: what it is for her to be conscious of this property is for her to be in a brain state that is normally caused by its instantiation. This is a real relation that she bears to the property as she is having her hallucination.

Of course, this is much too simple. For instance, thermometers undergo states that track temperatures, but they are not conscious of those temperatures. You also have unconscious states (e. g. states of the retina) that are caused by external features. For these reasons, a better idea is as follows

> The dyadic relation λxλy(subject x is conscious of property y) = the dyadic relation λxλy(x is in an internal (e. g. neural) state that is *poised for cognitive access* and that causally-covaries with the instantiation of y)

So the idea is that *the conscious-of relation is tracking plus cognitive access* (Dretske 1995, Tye 2000). An internal state is cognitively accessible iff it is apt to cause belief-like representations that predicate property y of something. Further refinements are possible. For instance, some theorists might appeal to the idea of function to indication rather than normal causation. But let's stick with the simple tracking account for now.

Let us call this the *reductive-externalist model* for reducing the conscious-of relation. It seems that, if reductive materialism is to be maintained, such an account of the conscious-of relation is almost inevitable. How *else* might we explain the presence of shapes in experience?

It is a consequence of this view that you can be conscious of a shape even if nothing before you has this shape. It retains Russell's idea that we are conscious of properties (properties that aren't properties of our experiences) but rejects his idea that we do this by being aware of sense data having the properties. It may seem odd that we can be conscious of a shape without being conscious of an instance of a shape, but it is an immediate consequence of the rejection of the sense datum view. Further, the oddness of the idea might to some extent follow from my choice of terminology. I have chosen to use "the conscious-of relation" as a name for the hypothesized relation to properties with the characteristics I've outlined. This name might suggest that the idea is that we literally see uninstantiated properties in hallucination. But this is no part of the view on the table. A better name for the relation might be the "appearing relation". The idea is that, when Mary hallucinates, she stands in the following relation to the property of being round: it *appears to her* that there is something having this property, even though nothing does have the property. Again, this is just an immediate consequence of the rejection of the sense datum view. When I say that she is "conscious of" it, this is all I mean. I just mean that she stands to

it in a relation with the features I've described. I don't say that she "sees" the uninstantiated property. Rather, the property characterizes how things appear to her.

Now let us generalize from the consciousness of "primary qualities" to the consciousness of "secondary qualities" like sensible colors. Although we motivated the present account by focusing on the experience of the primary quality *shape*, the account is itself general. It implies that in general what it is to be conscious of a property is to track it. Therefore, to be conscious of a *color* is to track it. But our visual systems track reflectance properties. So it follows that colors must be reflectances (or properties that are supervenient on reflectances). This might be called the *generalization argument*. So we get the result that colors, like shapes, are observer-independent physical properties of external items. And we are conscious of colors in the same way that we are conscious of shapes. In this way, the reductive-externalist model explains why colors and shapes appear as being in the same place.

We can take the generalization argument further. In having other types of experiences, we are *conscious of* bodily qualities arrayed in a bodily space, audible qualities in space, smell qualities, and so on. Reflection on the case of "primary qualities" suggests that the conscious-of relation is to be explained in terms of tracking. But the same conscious-of relation that we bear to primary qualities we also bear to secondary qualities. So the secondary qualities we are conscious of must be nothing but physical properties that our sensory systems track. In this way, we reach the conclusion that audible qualities, taste qualities, smell qualities, and so on are just objective physical properties of external items. Qualia just ain't in the head (or the brain). This view explains the evident fact that qualia appear arrayed in a kind of space, without requiring a private arena of sense data.

I have just argued that those wish to have a reductive view of the conscious-of relation must treat the "primary qualities" and the "secondary qualities" uniformly, treating them as objective features tracked by our sensory systems. Against this you might think that there is room for an alternative non-uniform reductive view. On the view I have in mind, shapes, positions and other "primary qualities" are response-independent features of objects, and we are conscious of them by tracking them. By contrast, the "secondary qualities" we are conscious of are response-dependent properties of external objects of the form: *normally causes neural state B* (Shoemaker 1994). But there are decisive problems with such a non-uniform view of the properties of which we're conscious. The chief problem is that there is no general theory of the conscious-of relation that's consistent with it. For, on this view, what physical-functional relation might the conscious-of relation be? Not the tracking relation defined above. For, although we bear this relation to shapes, positions and so on, we *don't* bear it to response-dependent properties of the form *normally causing neural state B*. Rather, we bear this relation to reflectances, chemical types, and so on. So my original claim stands: any reduction of the conscious-of relation along these lines requires

a uniform view: that the "secondary qualities" like the primary qualities are response-independent properties of external items. Therefore, the only general reductive theory of the conscious-of relation that has any promise requires that "qualia just ain't in the head." For this reason, I will primarily focus on this view in what follows.[2]

The reductive externalist view we have arrived at fits with a general reductive picture of the world promoted by Lewis, Sider, Dorr and others. To appreciate the beauty and simplicity of this view, consider an analogy due to David Lewis (1994). Imagine a grid of pixels each of which can be made light or dark. The arrangements of light and dark constitute interesting gestalt properties. They might, for instance, make a face. All properties instantiated on the grid are identical with complex properties—perhaps extremely complex disjunctive properties—built up from the basic properties and spatial relations among the pixels. The reductive materialist has a similar view of our world. Given some basic principles of property-formation, the pattern of instantiation of the fundamental properties and relations automatically brings with it the instantiation of an unimaginable number of complex properties, including *functional properties*. According to *reductive materialism* (or perhaps better, *identification materialism*), every property of the manifest image is identical with some of these complex properties. The idea is that all properties are reducible in this way. I have sketched a toy reduction of the *conscious-of relation*: it is a special kind of tracking relation. Likewise, maybe *being a hand* is identical with some complex *functional* property constructible from the fundamental physical and topic-neutral properties of the world - though, because of vagueness, it will be indeterminate which one. Notice that I understand "reductive materialism" (or "identification materialism") liberally so that functionalism is one form of it (after all, functionalists *identify* macro properties with functional properties definable in physicalistically-acceptable terms); this allows for multiple realizability.

Above I said that a non-reductive view of the conscious-of relation would be complex, requiring special nomic psychological laws or special "grounding laws". The reductive picture of the manifest image is appealing because it would mean that we do not need any such special, complicated inter-level laws to explain the connection between arrangements of fundamental properties and manifest image properties. Instead, we just need identities. And intuitively identifications have a unique feature: unlike certain kinds of brute "grounding laws", *they don't add to the complexity of our theory*. Here is another quick consideration in favor reductive materialism. Presumably, reductive materialism was true at the start of the universe right after the big bang when the universe was extremely simple; that is, at this time, all properties instantiated in the universe were either fundamental physical or topic-neutral properties or complex properties constructible from them. But if it was true then, considerations of uniformity suggest that it stayed true. For if not, exactly when did properties start to "pop up" that are *not* constructible from the fundamental physical and topic-neutral properties and relations?

In the case of the conscious-of relation, if you want to have a reductive view, the externalist model seems inevitable. After all, when Mary hallucinates and is conscious of the property of being round, what other *physical* relation does she bear to this property, with which the conscious-of relation might be identified? That is why I will make the externalist model the focus of this essay. However, you might think that the externalist reductive view is a total non-starter that shouldn't be taken seriously. I disagree: I think that the view is false but I don't think the view cannot be so easily ruled out all. Let me briefly look at some standard quick objections, and say why they fall short.

First, there is the explanatory gap. For instance, Mark Johnston also recognizes that reductive materialists must go in for something like the externalist model I've sketched. But he says it is a non-starter. He writes, "Herein lies the deep inadequacy of reductive materialism: There is no reduction of a relation which essentially involves disclosure to any combination of relations which essentially do not" (Johnston 2011, 215–216, fn. 8). One of his main objections takes this form: "how can a tracking relation make the lights go on?" (2007, 241). However, Johnston does not say enough to distinguish this objection from the standard explanatory gap objection. In response, reductive externalists can just *accept a posteriori materialism*, with inscrutable identities that cannot be *a priori* deduced from the fundamental physical facts (Tye 2000). This is also the standard response to the Knowledge Argument.

Second, you might object to the externalist reductive model that it leads to phenomenal externalism. For instance, on this view, you and your twin on "Inverted Earth" are conscious of different sensible colors, and hence have different color experiences, despite being internal duplicates. And you might think this is just absurd *a priori* (Levine 2001, 113). However, anyone who says this must be unaware of the history of human thought about experience. Many pre-modern thinkers, including Euclid and Ptolemy, proposed "extromission theories" of visual experience, on which rays emanate from the eyes and we see what the rays fall upon. On such a view, two internal duplicates might differ phenomenally, provided that the rays emanating from the eyes hit different objects. Evidently, this view could not be ruled out *a priori*—otherwise such eminent thinkers would not have proposed it. But then we cannot *a priori* rule out modern-day phenomenal externalism either (Pautz 2014).

Third, you might object to the reductive externalist program on the basis of the "dismal history of failure". For example, the reductive externalist model described above is of course only a toy model. It is incomplete; there are enormous problems of detail. In fact, reductive materialists face this issue in more boring cases. For instance, Sider asks, what is the metaphysical analysis of (say) being a hand? He admits, "I have no clue" (2011, 294). But, as Sider notes, this is not a damning objection to reductive materialism. He makes a couple of points. First, it is enough that we can provide toy reductions (117). Second, "the array of definable relations is extremely rich" (130, 294). So there are bound to be definable properties or relations that match pretty well the possible-worlds

extension of, say, "is a hand". Here is another point: in the pixel-screen example mentioned earlier, we cannot supply a reduction of *being a pixel face*. But, clearly, this property reduces to pixel-arrangements. So the "dismal history of failure" is not strong evidence for irreducibility.

To sum up. In this essay, my starting assumption is essential external directedness. Given this, the only model for reducing consciousness is the reductive externalist one. And standard objections to this model fall short.

Elsewhere, I have developed an *Empirical Argument* against the reductive externalist model of consciousness (Pautz 2010, 2016, 2018). In this essay, I add a more *a priori* argument, the *Significance Argument*. The argument has two parts. First, the reductive externalist model cannot accommodate the various ways in which consciousness is significant because of "multiple candidate cases" (§§2-4). Second, a nonreductive theory can do so in an attractive and unified way (§5).

My discussion is relevant to a current debate raging between two warring programs in the philosophy of mind. Uriah Kriegel (2013) has well characterized the debate. On one side there is what he calls the *naturalist-externalist research program* (NERP). On the other side is what he calls the *phenomenal intentionality research program* (PIRP). My target in this essay is a version of naturalist research program and my alternative will represent a way of implementing the phenomenal intentionality program.

2. The Dissimilarity-Grounding Significance of Consciousness

I will begin by arguing that reductive materialists cannot accommodate the *dissimilarity-grounding significance* of consciousness. I will focus on a single "multiple candidate case", which I call *Black-and-White Earth*. I begin by laying out the example; then I use the example to show that reductive materialists cannot accommodate obviously true ideas about the dissimilarity-grounding significance of consciousness.

An Example: Mary on Black-and-White Earth. As explained above, reductive externalists must say that the colors we are conscious of are *response-independent physical properties* of external objects, such as reflectance properties, or perhaps the microstructural bases of reflectance properties. For the sake of argument, let us just suppose for now that is right.

Now here is the *Black-and-White Earth* case. On Black-and-White Earth, the following things are true. First, every object contains a smaller object as a part. Second, the outer objects always have the achromatic colors black, gray or white. Third, the black outer objects contain red inner objects, the gray outer objects contain reddish-orange inner objects, and the white outer objects contain green inner objects. Third, the color of inner object and that of the outer object are causally yoked together by way of a natural, super-fast chemical process. So,

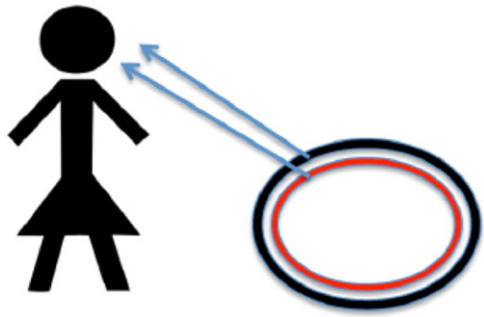

Figure 2. Both the red color of the inner rock and the black color of the outer rock cause Mary's brain state

for instance, if an inner object has a red reflectance, this causes its outer object to have a black reflectance. So if you could directly change the color of the inner object, this would change the achromatic color of the outer object by way of the super-fast chemical process (somewhat as when a chameleon changes color). And conversely if you change the color of the outer object (say from black to grey) you change the color of the inner object (from red to orange). In short, the colors of outer and inner objects are nomically yoked together.

It follows that when someone on Black-and-White Earth looks at an object, their visual system causally detects *two* colors at the same time: the color of outer object *and* the color of the inner object.

Now this hypothetical case poses an especially acute version of a well-known problem for reductive theories of representation in general. It is sometimes called the *depth problem* or the *distance problem*. In *LOT 2: The Language of Thought Revisited* (2008), Jerry Fodor called it the "which link problem". Suppose that Mary is an inhabitant of Black-and-White Earth, somewhat as Frank Jackson's famous Mary character is an inhabitant of a black and white room. Suppose that she views a black rock on a beach, which contains a red inner rock: if only she broke the outer shell (but she does not and maybe cannot), she would find a bright red, perfectly smooth rock on the inside. Her visual system equally causally detects the black color of the outer rock and the red color of the inner rock. Figure 2 illustrates the situation.

It would be absurd to suppose that Mary has conscious acquaintance with *both* the outer colors and inner colors. What would that be like? So reductive externalists might say that Mary is only conscious of the black color of the outer rock and in general that she is conscious of the colors of the outer objects. Or they might say that she is conscious of the colors of the inner objects—the most distal element of the causal chain. On this option, although the color of the outer rocks is part of the causal process, they are not perceived—no more than you perceive your retinas or the light between you and the object. But what could make it the case that one of these possibilities obtains to the exclusion of the other?

I do not think that appealing to biological function, asymmetric dependence or behavioral dispositions will help in this case.

The depth problem is an instance of the problem of the "dismal history of failure" that I mentioned in the previous section. However, as I said, I will be setting this problem aside in this essay. Here I will assume that the depth problem has some solution. My main point in this section is that, even if there is a solution to this problem of detail, the reductive program faces a deeper problem: it cannot accommodate the *significance* of the conscious-of relation.[3]

Before getting to that, we need to see how a solution to the depth problem might go. Given an "abundant" theory of properties and relations, we know that there is *bound* to be a "solution" to the depth problem, because there is bound to be a physical relation that Mary bears to the black color of the outer rock but not the red color of the inner rock. We can name such a relation as follows:

> *The tracking-17 relation*: λxλy (x is an internal state that is normally caused-17 by the occurrence of property y and that is "cognitively accessible", that is, that is apt to produce a belief-box representation R of y, that is, a belief representation that is normally caused-17 by y)

Here by "caused-17" I mean some kind of super-discriminating causal-informational-teleological relation that Mary's brain state bears to the black color of the outer rock only—not the redness of the inner rock. (Even though I call this relation "caused-17", it may not be a purely causal relation; for instance, it may bring in teleological conditions.) Again, we cannot specify this relation, but we know that there is such a relation.

Now, of course, Mary causally detects the inner colors just as much as the outer colors. Since relations are abundant, if there is a relation, tracking-17, which Mary bears to the black color of the outer rock and not the red of the inner rock, there is also bound to be a very similar variant tracking relation, tracking-18, with the reverse extension in this case: one that Mary bears to the red color of the inner rock, not the black color of the outer rock:

> The tracking-18 relation: λxλy(x is an internal state *s* that is normally caused-18 by the occurrence of a property *y* and that is "cognitively accessible", that is, that is apt to produce a belief-box representation *of y*, that is, a belief representation that is *normally caused-18 by y*)

Think of it this way: on viewing the rock, Mary's visual system equally causally detects both the *black* on the outside and the *red* on the inside. They are just different steps in the causal chain. In particular, Mary bears the tracking-17 relation to one step, and the tracking-18 relation to another step. In fact, there are more elements to the causal chain: for instance, Mary bears another causal relation, tracking-16, to a *complex property of light* (composed of photons) in the space between the rock and herself. But let's just focus on tracking-17 and tracking-18.

What might the cause-18 relation be? Maybe it is the following relation definable in terms of the cause-17 relation:

> $\lambda s \lambda y$(state s is caused-17 by some property z and z is immediately causally dependent on property y)

So, as Mary looks at the rock, Mary's brain state (and Mary's downstream belief-like representation) is caused-17 by the black color of the outer rock and it is also caused-18 by the red color of the inner rock. Since the cause-18 relation is nearly identical with the cause-17 relation, the *tracking-18 relation is nearly identical with the tracking-17 relation.* The difference between them is insignificant.

We can now finally say how reductive externalists might solve the depth problem. They can say that it is just a fact that the conscious-of relation is identical with the *tracking-17* relation, and not the *tracking-18* relation. So Mary is only conscious of the black color of the outer rock, not the red color of the inner rock. In general, she is only conscious of the achromatic colors black, white and gray that characterize the outer objects, not the chromatic colors that are lying just under them. Although there is a world of vibrant color qualia just below the surface, she has no acquaintance with them. Analogy: even though photons in the air are part of the causal process when you perceive something, you do not perceive those photons. In the same way, reductive externalists will say that, even though the vibrant colors of the inner objects are part of the causal process, they are not themselves perceived.

The dissimilarity-grounding significance of consciousness. Now we can get to the main point of the present section. Even if reductive externalists can solve the depth problem in some such way, their view faces a deeper problem about the dissimilarity-grounding significance of the relation of conscious acquaintance. To lead to the problem, let me start with a traditional idea about sensible properties, the *relata* of this relation.

Berkeley said that "nothing a colour or figure can be like nothing but another colour or figure", and this was part of his argument against Lockean materialism (Berkeley 1710, section 8). A more refined thesis version of Berkeley's thesis might say that nothing can be *intrinsically very similar* to a color but another color. For instance, it would be absurd to say that a sound quality is *intrinsically very similar* to a certain color in the way that one color is intrinsically very similar to a nearby color (though of course, they might be alike in some other sense, for instance, in both being enjoyable, or in both being properties). We just know this idea to be absurd *a priori*.

Now imagine that the reductive externalist model implied this absurdity. Of course, as a matter of fact, it doesn't. As a matter of fact, the experience of pure red and the experience of middle-C have very different external correlates: namely, a reflectance-type and a sound-wave property. But imagine that it had turned out to actually be the case that the external physical property P that is the external correlate of the experience of pure red were intrinsically very similar

to the external physical correlate *P** of the experience of middle-C. Again, this is not the case. Still, it is conceivable that it should have been the case. And this is enough for my present point: if it were actually the case, no one would then accept that pure red = P and middle C = P*. The reason is that this view would imply that the color pure red turns out to be intrinsically very similar to middle-C (indeed it might imply that pure red is intrinsically *just as* intrinsically similar to middle-C as it is to a nearby reddish shade on the color wheel)! It would imply a *ludicrous similarity*.

My first problem with the reductive externalist model takes this form. As I said, the only difference is that my problem concerns, not the sensible qualities themselves, but the *conscious-of relation* that we bear to sensible qualities. I will argue that the reductive externalist model implies *a handful of ludicrous similarities* involving this relation—similarity claims that are just as absurd as the claim that red is intrinsically very similar to middle-C.

To begin with, note that the following is uncontroversially true in the situation I have described: [#] *When Mary stands in the tracking-17 relation to the black color of the outer rock, she also stands in the nearly identical tracking-18 relation to the red color of the inner rock.* These relations are just evidently intrinsically nearly identical. Just look at their definitions above: they differ only minutely. So [#] is true in the situation.

Now, if the reductive externalist model is true, then *the conscious-of relation is identical with the tracking-17 relation*. And this identity statement, together with [#], implies the following:

1. When Mary is *conscious of* the color black, she stands in a nearly identical relation to the color red, *even though at the time Mary is not at all conscious of the color red*. Even though she is fully conscious of black and not at all conscious of red, there is *barely any difference* in her relation to these qualities. In fact, even though is fully acquainted with the color black and not at all with red, the difference between Mary's relation to black and her relation to red is *no greater than* the minute difference between tracking-17 and tracking-18 (for, on the reductive externalist model, this difference *just is* the difference between tracking-17 and tracking-18).

The problem for the reductive model is that (1) is just not true. Imagine being Mary. Mary is conscious of the color black, which seems to her to pervade a round surface. She is *acquainted with* it. The color black is *present to* Mary. *There is something it is like* for her to be acquainted with the color black. By contrast, by hypothesis, Mary is not at all acquainted with the color red. The color red is not at all present to Mary. *There is nothing it is like* for her to be related to the color red - any more than there is anything it is like to stand in a mere spatial relation to a color while you are asleep (for instance, if you should sleep just below a giant red orb). Therefore, at this time, Mary's relation to the color black

is *nothing like* Mary's relation to the color red, contrary to (1). If Mary said "While I am conscious of the color black, I am in an intrinsically very similar relation to the color red", Mary would be saying something straightforwardly false—and something she knows to be false by introspection and reflection. It is just as false as the claim that pure black is intrinsically very similar to middle-C. Since this is so, the reductive externalist model, which implies it, is false as well.

Let us turn to a second ludicrous similarity implied by the reductive externalist model. Consider the actual world. Suppose that, in the actual world, some person, Martha, has a vivid color experience of a smooth, red-colored rock. Then the reductive externalist model implies the following:

2. Martha has a reddish experience; call it P17. Mary on Black-and-White Earth has *intrinsically extremely similar property P18*, but P18 is *not* a color experience, and in fact *it is not an experience property at all!*

Here is why the reductive externalist model implies (2). On the reductive model, Martha's having the reddish experience consists in her *tracking-17 the color red*. Call this relational property P17. At the same time, Mary has a distinct but very similar property, namely *tracking-18 the color red* (look back at Figure 2). Call this relational property P18. Now P17 and P18 consist in standing *nearly identical* relations to the *same* property. But, on the reductive externalist model, Mary's having *P18 doesn't* constitute her being acquainted with the color red. In fact, *P18 is not itself an experience property at all*, on this view. For, on this view, Mary of course doesn't have two distinct experiences—just one, namely, the experience of the black color of the outer rock. So, on the reductive model P17 constitutes experiencing red, but P18, even though it is intrinsically very similar to P17 (hence intrinsically very similar to experiencing red, on the reductive externalist model), doesn't constitute an experience at all. That is why the reductive model implies (1).

But, again, (1) is just ludicrous. It is just as ludicrous as saying "there is a property that is intrinsically just like the color *pure red*—but it's not a color at all!"

Or here is another analogy. Suppose you had a headache in the morning. Someone asks you how you are doing now. Suppose you replied "well, I am in a state that is intrinsically nearly identical with my morning headache", your friend says, "that's too bad", and you reply, "oh, don't worry, my new state is not a headache—in fact, it's not an experience at all, *even though it is intrinsically identical to my morning headache.*" You would find this speech absurd. (2) is equally absurd. Since the reductive model implies (2), we must reject it.

Let us now turn to a more complex case that will allow us to derive a third and final absurd similarity claim from reductive externalists. It is a twist on the Black-and-White Earth case. Let us now suppose that the black rock that Mary is looking at has *two* "inner rocks": the bright red inner rock in turn contains a *yellow* inner rock. That is, just add to Figure 2 another circle within the circle. By way of the super-fast chemical process, the yellow of the innermost rock causes

the red color of the middle rock, which in turn causes the black color of the outer rock. Then Mary bears the tracking-17 relation to black color of the outer rock, the tracking-18 relation to the red of the middle rock, and the *tracking-19* relation to the yellow of the innermost rock. Here tracking-19 is defined along the same lines as tracking-18, but with one step in the causal chain added.

The following is clearly true claim in this case: [&] *The difference between Mary's tracking-17 the color black and her tracking-18 the color red is the same as the difference between Mary's tracking-18 the color red and her tracking-19 the color yellow*. Compare: the difference between having 1000 hairs and 2000 hairs is the same as the difference between having 2000 hairs and 3000 hairs. Now, on the reductive model, the conscious-of relation is identical with the tracking-17 relation, and both tracking-18 and tracking-19 are non-experiential relations. This, together with [&], entails:

3. The difference between Mary's experiential relation to the color black and Mary's entirely non-experiential relation to the color red is *the same as* the minute difference between the non-experiential relations tracking-18 and tracking-19.

The problem is that (3) is just false. There is an *enormous difference* between Mary's relation to the color black and Mary's relation to the color red. Think of it from the first person. By hypothesis, Mary is *conscious of* the color black: the color black is *present to* Mary. There is something that it is like. By contrast, Mary is not at all conscious of the color red: it is not at all present to Mary. There is nothing that it is like for her to be related to this color (any more than there is anything it is like to stand in mere spatial relation to a color while you are asleep). So, in this situation, the difference between Mary's experiential relation to the color black and Mary's entirely non-experiential relation to the color red is *enormous.* By contrast, the difference between Mary's non-experiential relation to the color red and her non-experiential relation to the color yellow is *minute*. For it is just the difference between the non-experiential relations tracking-18 and tracking-19. And that difference is minute. From these points it follows that, contrary to (3), the difference Mary's relation to the color black and her relation to the color red is *much greater* than the difference between Mary's relation to the color red and Mary's elation to the color black.

Let me sum up the problem. The reductive model of the conscious-of relation violates the *dissimilarity-grounding significance* of consciousness. By that, I just mean it implies the ludicrous similarities 1–3. If you are honest with yourself, you will admit that these similarity verdicts are plain false—just as false as the idea that pure red could be intrinsically very similar to middle-C (just as similar as it is to reddish-orange). And this is just the beginning. As we shall see in the following sections, the reductive model has other absurd consequences about the reason-grounding and determinacy-grounding significance of consciousness. And we will see afterwards that a nonreductive model avoids such ludicrous

consequences while allowing us to retain allegiance to a materialist theory of consciousness. Taken together, these points provide a strong reason to prefer the nonreductive model to the reductive model.

Let me make a few clarifications. First, the source of the present problem is [#]: tracking-17 is intrinsically very similar to tracking-18, and [&]: the difference between tracking-17 and tracking-18 is the same as the difference between tracking-18 and tracking-3. I want to guard against a wrongheaded reaction to [#]: "Against [#], proponents of the reductive model can say that there is a huge difference between tracking-17 and tracking-18, namely, that tracking-17, but not tracking-18, *constitutes the distinct, unique conscious-of relation*." This is wrongheaded because proponents of the reductive model don't accept the view that the externally-determined tracking-17 grounds a *distinct* conscious-of relation whereas tracking-18 does not. That view is a nonreductive or dualist view that is quite contrary to the reductive model. (It would be an externalist but non-reductive view of the conscious-of relation: this relation reaches out in the world and is irreducible, but what external properties we bear it to depends on what properties we tracking-17. See footnote 20.) Rather, their picture is that the conscious-of relation *just is* the tracking-17 relation. And that doesn't mean that tracking-17 somehow "glows" while tracking-18 doesn't: there is no more to the nature of tracking-17 (the conscious-of relation) than its physicalist definition above. And, since tracking-17 and tracking-18 have nearly identical definitions, they are intrinsically extremely similar. Proponents of the reductive model just have no way of getting around [#] and [&]—and so no way of getting around the ludicrous similarities 1–3.[4]

Second, it would be wrong to say that the present problem about dissimilarity-grounding significance is just an instance of the familiar explanatory gap objection applied to the reductive externalist model of the conscious-of relation. To see this, suppose that there can be opaque identities that are not deducible from the physical ground floor, for instance the identity of the conscious-of relation with tracking-17 rather than with tracking-18. Then the explanatory gap objection fails because it relies on the mistaken assumption that identities must be *a priori* deducible from the physical ground floor. But my problem still applies. For the reductive model still implies 1–3. There is just no way of getting around that. So my problem is quite separate from the explanatory gap problem.

Third, you might think that my problem is based on the alleged intuition that the conscious-of relation is "perfectly natural" in the sense of Lewis (1994) and Sider (2011). And you might object that this too theoretical, so that my problem has no teethe: reductive materialists will just deny it. But this is a serious misunderstanding of the problem. In fact, not only does my problem not depend on such a theoretical claim: it doesn't even employ the idea of "naturalness" *anywhere*. This is all to the good because I am quite skeptical of the idea of naturalness and especially of relative naturalness. My problem only depends on the pretheoretical idea that 1–3 are false similarity claims in the case described.

That is, the conscious-of relation is a dissimilarity-maker. This claim is not at all the same as controversial theoretical claim that the conscious-of relation is "completely natural".

Let me explain why these claims are not the same. Suppose that the conscious-of relation is a perfectly natural relation R. This doesn't by itself mean that it is a potent dissimilarity-maker (it doesn't rule out ludicrous similarity claims like 1–3), for it compatible with the hypothesis that there is a distinct *but very intrinsically similar* perfectly relation R* in nature. (Compare: the determinate masses are typically considered perfectly natural but each determinate mass is very similar to nearby masses.) Conversely, there are coherent views on which the conscious-of relation should have turned out to be a very unnatural relation (that is, to have an extremely long definition in fundamental terms), and yet be a potent dissimilarity-maker in every world (so that absurd similarity claims like 1–3 are never true). For instance, it is conceivable that it should have been constituted by a ray emanating from our eyes to objects, where the ray consists of a huge number of types of particles in a special configuration and the ray can only hit one thing at a time—a ray of a kind that occurs nowhere else in nature. Then it would have been a very special relation unlike any other; but, if degree of naturalness is understood in terms of length of definition (Sider 2011), it still would have been extremely unnatural, since it has a very long definition in a fundamental language. As a matter of empirical fact, this view is incorrect. In fact, there is an *abundance* of *very similar* physical relations to constitute conscious acquaintance (like tracking-17 and tracking-18). Given this empirical fact, reductive proposals have the unfortunate consequence of entailing ludicrous similarity claims like 1–3.

And there is an even worse problem.

3. The Reason-Grounding Significance of Consciousness

My next problem is that reductive materialists cannot plausibly accommodate the *reason-grounding* significance of consciousness. As we shall see, this is intimately related to the problem I just developed, that it cannot accommodate the dissimilarity-grounding significance of consciousness.

I begin by saying what the reason-grounding significance of consciousness is. Then I use a twist on the Black-and-White Earth case to develop a destructive dilemma for reductive externalists, showing that they cannot plausibly accept this idea.

What is the reason-grounding significance of consciousness? Here is a very plausible and popular idea: having an experience with a certain phenomenology is *sufficient for* having a reason to believe certain things. Notice that this is just a claim of sufficiency: it doesn't require the strong thesis that experience is necessary for having reason to believe certain things, so that a Zombie or super-blindsighter

doesn't have reasons to believe things. I will argue there is a hitherto unnoticed tension between this idea and reductive materialism.

One currently popular version of this idea is what Pryor (2000) calls "dogmatism". On this view, necessarily, if you are conscious of the ostensible state of affairs *that object o has property P* (as Pryor 2000, footnote 37 puts it, if you are presented with this state of affairs with "phenomenal force"), then you have a *prima facie* reason to believe that o has property P. For instance, if Mary on Black-and-White Earth is conscious of the ostensible state of affairs of there being a black item (a rock) on a grey background, she has a prima facie justification for believing that there is such a thing. You have such a reason even in an illusory or hallucinatory case in which this state of affairs *doesn't really obtain*.

I accept this idea. However, it is controversial. For instance, some think that only "success states" (e. g. genuinely seeing *that o is P*) can play a justifying role in experience. This goes against dogmatism in cases of illusion and hallucination.

For this reason, in illustrating my epistemological problem for the reductive model, I will focus on an idea I think should be less controversial. As I said at the start of §1, experiences can necessarily give us reasons believe truths about resemblances among colors or shapes. If this is right, it should apply to Mary on Black-and-White Earth too. For example, suppose that on Black-and-White Earth Mary experiences a black thing, a dark-grey thing, and a white thing right next to each other. Intuitively, necessarily, if Mary is thus conscious-of these qualities, then Mary thereby have a reason to believe the necessary truth that black is more like dark-grey than white. (In fact, some would say she can get "certain" or "demon-proof" reason to believe this necessary, timeless truth.) As Russell said in *The Problems of Philosophy*, "between universals, as between particulars, there are relations of which we may be immediately aware" (1912, 102–103). This specific claim doesn't face the same problems as dogmatism. For instance, unlike dogmatism, it is compatible with the thought that only *success states* can be a justifying role. For in both veridical and non-veridical cases you are successfully aware that black is more like dark gray than white even if you are not successfully aware of objects with these colors.

A dilemma for reductive externalists. Suppose, then, that proponents of the reductive model accept that, by virtue of being *conscious of* the qualities black, dark grey and white, Mary has a reason to believe that black is more like dark-grey than white. Now, on their view, the conscious-of relation is the tracking-17 relation, and these qualities are reflectance properties. If the conscious-of relation has such a reason-grounding significance, and if the conscious-of relation is the tracking-17 relation, then it follows that the tracking-17 relation has this round-grounding significance: simply by virtue of standing in the *tracking-17 relation* to these properties, she has a reason to believe that black is more like dark-grey than white.

However, this view faces a dilemma. On Black-and-White Earth, Mary also bears the very similar *tracking-18 relation* to the qualities red, reddish-orange

and green possessed by the inner objects (see Figure 2). *What should proponents of the reductive model say about the reason-grounding significance of this relation?* They have two options; but neither is at all plausible.

The first option is *restrictive*: while tracking-17 has enormous reason-grounding significance, tracking-18 *has none at all*:

4. Mary's bearing the *tracking-17 relation* (= the conscious-of relation, on this view) to the qualities black, gray and white gives her a reason to believe that black is more like grey than white; but her bearing the *nearly identical* tracking-18 relation to the qualities red, reddish-orange and green does *not* give her any reason to believe *anything* about those qualities.

The second option is *pluralist*: tracking-18 *as well as* tracking-17 has reason-grounding significance:

5. Mary's bearing the *tracking-17 relation* (= the conscious-of relation, on this view) to the qualities black, gray and white gives her a reason to believe that black is more like grey than white; *and* her bearing the tracking-18 relation to qualities red, reddish-orange and green gives her a reason to believe that red is more like reddish-orange than green.

However, neither horn is plausible.

Take (4) first. (4) is just absurd. To bring this out, let's start with some examples. First, let us suppose that the friendship relation comes in degrees. Let friendship-17 and friendship-18 be two such degrees. Now friendship-18 has normative significance. If Mary stand in this relation to someone, Mary has certain *pro tanto* duties with respect to them. Further, friendship-17 is a good thing: if Mary bears this relation to someone, then that adds value to the world. It would be absurd to accept these claims but then go onto to say: "But you know what?—although friendship-18 is nearly identical with friendship-17, friendship-18 is totally different; it just has absolutely *no* such normative significance ... When you bear this relation to someone, you have no duties to help them under any circumstances, and it is not a good thing at all." Here is another absurd view. Let pain-17 and pain-18 be two similar degrees of pain. It would be absurd to say: "Pain-17 has a normative significance: if you have pain-17, you have a *pro tanto* reason to desire that it stop, and you have a *pro tanto* duty to see to it that others do not have pain-17. But pain-18 has absolutely no normative significance: it is false that if you have pain-18, you have any reason at all to desire that it stop, and it is false that you have a *pro tanto* duty to see to it that others have pain-18. In fact, if someone next to you has pain-18, and you could easily make it stop, you have no reason at all to do so." Now if you think that these views are absurd, you should think that (4) is absurd in the same way. For (4) says that tracking-17 has a certain normative (viz. epistemic) significance, but that the nearly identical relation tracking-18 has no such

significance. Likewise for all the other possible variants: tracking-16, tracking-19, and so on.

My main case against (4), then, is a kind of argument from analogy. But we can also argue against it on the basis of a plausible general principle: roughly, if relation or property P necessarily possesses a certain normative significance, and if P* is intrinsically very similar to P*, then P* must have *a similar* normative significance. Call this the *small difference principle.* This principle is rough, and further refinements are possible.[5] Clearly, (4) violates this principle, as do the bizarre normative views mentioned above about friendship and pain. These views require *normative discontinuities* or *normative singularities*.

The reductive theorist cannot reply: "Mary is only acquainted with the achromatic surface colors and doesn't bear anything like this relation to the vibrant chromatic inner colors". To bring this out, let acquaintance-17 be tracking-17, acquaintance-18 be tracking-18, acquaintance-19 be acquaintance-19, and so on. Then it is undeniable that Mary is acquainted-17 with certain qualities, acquainted-18 with certain other qualities, and so on. And there is no big difference between acquaintance-17 and acquaintance-18: they are very similar and in every way on a par. To say that acquaintance-17 has reason-grounding significance but that acquaintance-17, acquaintance-18, and so on, have none, requires a bizarre normative singularity that goes against the small-difference principle.[6]

Now let me turn to the second horn, (5). On this option, just as both friendsip-17 and friendship-18 have similar normative significance, so tracking-17 and tracking-18 have similar epistemic significance. So, unlike the restrictive option, this pluralist option fits with the small differences principle and doesn't require bizarre normative discontinuities. Still, this option is absurd too. For, given the reductive model, it implies the following. As Mary looks at the objects, Mary is only acquainted with black, gray and white. Mary is not at all acquainted with red, orange and green (the colors of the inner objects). In fact, Mary has *never* been acquainted with those colors, and indeed has never been told anything about them. Still, (5) implies that, as Mary views the objects, Mary does not only have a reason to believe that black is more like gray than white: Mary also has nearly equal reason to believe that red is more like orange than green! That is, even though Mary is not acquainted with these qualities, and never has been, Mary is in a position to know *what these qualities are like*. This is just absurd! Just imagine being Mary in this situation. There you are: you are fully acquainted with black, gray and white *right there before you*, whereas you have never been acquainted with red, orange and green. Nor have you been told anything about red, orange and grey. If this is the situation, do you have *just as* much reason to believe that red is more like orange than green as you have to believe black is more like gray than white. Evidently not! You have no such reason. (5) over-generates reasons.

To drive this point home, suppose that Mary is released from her white-black predicament: finally, the outer shell of objects is broken, so that she can finally be conscious of the vibrant inner chromatic colors red, orange and green.

Given (5), she doesn't gain any *new* reasons to have beliefs about those colors. Yes, she now has a reason to believe, for instance, that red is more like orange than green—but according to the pluralist option (5) she *already* had this reason even *before* she was conscious of these qualities. This is just an absurd idea.[7]

In short: to accommodate the reason-grounding significance of consciousness, the reductive model requires either (4) or (5). Both result in absurd results. This provides a strong reason to reject this reductive approach.

Let me conclude by making two points.

(i) To illustrate the difficulty of the reductive externalist model in accommodating the reason-grounding significance of consciousness, I have focused on the plausible idea that consciousness gives us reason to believe necessary truths about the resemblances among qualities, such as colors or shapes. However, to illustrate the problem, you could use just about any thesis about the reason-grounding significance of consciousness. For instance, consider the popular theory of dogmatism mentioned above: bearing the conscious-of relation to an ostensible state of affairs (e. g. that a thing has a certain shape or sensible color) gives one *prima facie* reason to believe this state of affairs really obtains. This plausible claim generates the very same dilemma. On this view, if you are *conscious of* a state of affairs (or, using Pryor's terminology, if you are *presented with it with phenomenal force*), then you have immediate *prima facie* reason to believe that it obtains. To illustrate how this view creates a problem for the reductive externalist, let us first define some terms. Let us say that Mary is forcefully-presented-with* a state of affairs iff she bears the tracking-17 relation to it; and let us say that she forcefully-presented-with** a state of affairs if she bears the tracking-18 relation to it. Given these stipulations, Mary is both forcefully-presented-with* the outer object's being *black* and she is forcefully-presented with** the inner object's being *red*. And these relations are *barely different*. On the reductive externalist model, which one of these relations "really is" the epistemically powerful relation of being-forcefully-presented-with? Suppose it "really is" the forceful-presentation* relation (that is, the tracking-17 relation). Then Mary has immediate reason to believe that a *black* thing is there. Now, does the barely different forceful-presentation** relation, which Mary bears to the inner object's being *red*, have a similar epistemic significance? If not, we get a *normative singularity*. If yes, we get an absurd *over-generation of reasons*. Therefore, dogmatism about perceptual justification cannot be plausibly combined with the reductive externalist model. These two popular ideas are in tension.

(ii) To illustrate the difficulty of the reductive model in accommodating both the dissimilarity-grounding and reason-grounding significance of consciousness, I have assumed that reductionists about the conscious-of relation must hold that colors are reflectance properties (or their microstructural bases). However, you might think that this was unfair. Couldn't reductionists hold that sensible colors are Shoemaker-like dispositional properties, for instance? The idea there would be that although the primary qualities we are conscious of are response-independent properties so things, the secondary qualities are dispositions to cause brain states

in us. I have two points about this non-uniform position. First, I argued in §1 that reductionists cannot accept this non-uniform view. Second, even if they can, the problems I have developed here still arise. For instance, consider a world where every round object contains within it an oval object, where the shape of the inner object and that of the outer object are nomically yoked together. Then you bear the tracking-17 relation to the outer shape and the tracking-18 relation in the inner shape. And the very same problems arise. It would be absurd to say that one but not the other has epistemic significance; and it would also be absurd to suppose that both have.[8]

The conclusion I draw from this section and the previous one is that reductionists can accommodate neither the dissimilarity-grounding nor the reason-grounding significance of consciousness.

4. The Determinacy-Grounding Significance of Consciousness

I have just argued that reductionists cannot plausibly accommodate the reason-grounding significance of consciousness. But consciousness doesn't just give us a reason to have certain belief. It explains our capacity to have those beliefs in the first place. That is, consciousness has a belief-grounding significance. More than that, consciousness makes it easy to have beliefs with very *determinate* contents about qualities and other ostensible items. For instance, if you are conscious of a color quality, you are easily able to more or less determinately think about it (and this is so even in cases illusion or hallucination where it does not in fact quality anything). Of course, although I focus on belief, consciousness also enables us to have other determinate propositional attitudes about qualities: for instance, if you experience a bad smell, this enables you to have a determinate desire that *it* go away. However, I will argue that reductive materialists about the conscious-of relation also cannot plausibly accommodate such obvious facts.

Once again, the problem derives from multiple candidate cases. In the Black-and-White Earth case, I assumed a modicum of determinacy; in particular, I assumed that, it is determinate that Mary *is conscious of and thinks about* the *black* color of the outer rock, not the *red* color of the inner rock. But I will now introduce a series of additional multiple candidate cases that call into question this assumption.

In a nutshell, I will be showing that the traditional Quine-Kripkenstein-Putnam problem of indeterminacy applies to our beliefs and experience about basic qualities. The problem starts at the source. In fact, I will suggest that, if anything, the Quinean problem may be *harder* in such basic cases, due to unique features of our thought about qualities. Radical indeterminacy may be hardest to avoid exactly where it is least plausible.

As I said, I will illustrate the problem of the determinacy-grounding significance for reductive materialists with a series of new multiple candidate cases. The first two are real-world cases and the third is hypothetical. I will begin by

simply laying out the cases and describing the *prima facie* problem. Only after laying them all out will I turn to my official argument that reductionists cannot plausibly account for the determinacy-grounding significance of consciousness.

First Example: Simple Mary and the Strong Smell. My first example concerns the experience of smell. Imagine that a jar of perfume spills. Mary is conscious of a distinctive quality. It is present to her mind. It grabs her attention. She thinks *it* is present. So, she thinks about *it*. There may not be perfect determinacy here; but at least there is not *radical* indeterminacy here. There are not many very different properties, where it is indeterminate which one she is thinking about. Rather, there is *one* property or quality and it is determinate that she is thinking about *it*. And this can be so even if Mary is a conceptually unsophisticated individual, such as a child or even an animal. This is an example of the determinacy-grounding significance of consciousness.

But it is very hard to see how reductive materialists can accommodate this point. To appreciate why, the first thing we need to understand is that this case is in fact a *multiple candidate case*. There are multiple physical properties that are candidates to be the unique smell quale that Mary is thinking about.

For instance, on the reductive externalist model I have been focusing on, the smell quale Mary is thinking about is a chemical property, call it *P1*, of the cloud of molecules. Qualia aren't in the head, according to this view. Mary does not and indeed cannot attend to and think about the intrinsic neural properties of the brain state that realizes her experience; she can only attend to the external chemical property tracked by this brain state. This fits with "transparency observation". In §6, I said that the generalization argument puts pressure on reductive materialists to accept such a view. If shapes are response-independent features of things that we are conscious of because of sensory systems detect them, then considerations of uniformity suggest that the same is true of all sensible properties.

For instance, when you see a tomato, the reddish quality that seems to pervade a round region is not instantiated *in* your brain. It is instantiated in the world along with the shape. It then becomes compulsory to extend the same externalist reductive model to the experience of smell qualities.

However, there are other candidates to be what Mary is thinking of and referring to. For instance, Ned Block (2010, 24), contrary to reductive externalist, thinks that this olfactory quality is identical with an *internal neural property instantiated in Mary's own brain*. Call it P2. His motto is "qualia are in the head". He rejects transparency. David Papineau (2016) takes a similar view.

So, even in this basic case, there are *multiple candidates* to be what Mary is thinking about (P1 *vs* P2), just as in more standard illustrations of the indeterminacy problem (plus *vs* quus, rabbits *vs* undetached-rabbit-parts). (See figure 3.)

To illustrate the problem for reductive materialists, let's assume a language of thought view. Then, on reductive materialism, all that is going on in this

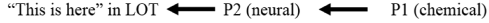

Figure 3. The physical facts in the case of Mary and the strong smell

case is that the sentence "this is here" (a neural pattern with a syntax) enters Mary's belief-box in her brain, where this is caused both by the occurrence of P1 in the vicinity of her nose and by the occurrence of the neural property P2 further downstream. On reductive externalism (Dretske and Tye), the smell quale is determinately identical with some or other chemical type P1. If this is so, it is determinate that Mary attends to and thinks about P1, not P2; that is, "this is here" is Mary's belief-box determinately refers to P1, not P2. Therefore, they face the following question: how could the purely physical facts make this the case?

In fact, there is a problem here facing all reductive materialists. All materialists should want to avoid the following claim:

6. When Mary smells the perfume, there are multiple, radically different properties that are candidates to be the smell quale that Mary is attending to and thinking about (e. g. P1 and P2), and it is *indeterminate* which of them *is* the smell quale that Mary is attending to and thinking about.

All materialists should want to avoid (6) because it is absurd. It implies that it is indeterminate whether the smell qualia that we are conscious of in experience are "in the head" or "in the world". To get a fix on what (6) would mean, consider an analogy. Imagine that there is clairvoyant who spontaneously forms beliefs about the whereabouts of two conjoined non-identical twins, Bob and Jane (beliefs like "that person is in China", etc.): clearly, here it would be indeterminate whether his use of "that person" refers to Bob or Jane. According to (6), what is going on in the Mary case is similar. (6) implies that it is indeterminate *whether internalism or externalism about qualia is correct*. It is indeterminate which side is correct in the debate between internalists like Block and Papineau and externalists like Tye and Dretske. For instance, it would mean that it is indeterminate whether an *accidental BIV internal duplicate* of Mary *experiences the same smell qualia as Mary* (Block and Papineau) *or no smell qualia at all* (Tye and Dretske). (Analogy: since it is indeterminate whether "my house" refers to my house proper or my house plus the garage, it is indeterminate whether the truths about my house supervene only on what happens in the house proper or whether they also partly depend on character of the garage.) And it would mean that it is indeterminate whether or not the "transparency observation" is correct. But it is just obvious that these implications of (6) cannot be right. The Mary case is *totally different from* the clairvoyant case. In the clairvoyant case, the subject is *not* conscious of what he is thinking about. By contrast, there is *one* specific quality that Mary is *conscious of*. *It* grabs her attention. And she is thinking about the one quality that she is conscious of. Therefore, it is not radically indeterminate what she is thinking about.[9] Moreover, it is not indeterminate whether an *accidental BIV*

internal duplicate of Mary *experiences the same smell qualia as Mary or no smell qualia at all*. That idea just makes no sense.

But what could make it the case that Mary is thinking of one or the other, contrary to (6)? All reductive materialists face this question. So they all face a present challenge about the determinacy-grounding significance of consciousness: the way in which it makes possible more or less determinate thought about qualities.

To underscore the problem, consider an analogy. Consider a *Simple System* that has no experiences whatever. Maybe it is an insect or a robot. Suppose that, like Mary, the Simple System can detect perfume. It detects the perfume by detecting its chemical signature P1. This is a "sign" of the perfume for the Simple System. Another "sign" of the perfume is its own neural pattern P2. When the perfume is present, and the mentalese sentence "this is here" goes into the Simple System's belief-box, this is caused both by the occurrence of P1 in the air and (further downstream in the causal chain) by P2 in smell system. Here the right verdict would seem to be radical indeterminacy. We can take its mentalese sentence "this is here" to be about either one; it's a matter of interpretation. But if indeterminacy is the right verdict in thesis case, then by parity of reasoning reductive materialists are committed to (6) in the case of Mary. Yet in the case of Mary (6) is absurd from the first person.

This of course is an instance of a familiar type of problem of Quinean indeterminacy. But I think that there are two features of our intentional directedness at qualities that make the problem special here: *Determinacy* and *Easiness*. By Determinacy, I mean that when we have experiences it is not radically indeterminate what qualities we are conscious of and think about; and it is not radically indeterminate what our experiences are like. Determinacy is more obvious here than it is standard cases; while accepting indeterminacy in other cases (plus/quus, rabbit/undetached-rabbit-part) may be an option, here (6) is just not an option.[10] By Easiness, I mean that it is extremely *easy* to think about qualities—much easier than it is to refer to rabbits, the plus function, the property of being a chair, and so on. Therefore, it can take place under very *minimal conditions*. In fact, when it comes to thought about basic qualities, Determinacy and Easiness are at their maximum. Because of this, standard solutions to the Quinean problem do not carry over here. For reductive materialists, it is very hard to come up with an account that satisfies both Determinacy and Easiness simultaneously.[11]

For instance, some have suggested that descriptive fit plays a role in minimizing Quinean indeterminacy. So you might hope that this could be applied to the present case. On one elaboration, the idea would be that, when "this is here" goes into Mary's belief-box as a result of P1 and P2, "this is a feature of an external odorant in space" *also* goes into Mary's belief-box (or would tend to go in there). This externalist "proto-theory" of smell quality is satisfied by external property P1, not the internal property P2. So Mary's belief counts as being about P1, and not P2, in accordance with the reductive externalist model.

But, precisely because thought about qualities is so easy and basic, the descriptive fit gambit doesn't get off the ground here. We have assumed that Mary is a cognitively unsophisticated child or animal who doesn't yet have any kind of externalist (or internalist) descriptive beliefs or proto-theory about olfactory qualities. Still, (6) is false: it is not radically indeterminate what quality she is referring to. Evidently, Mary's ability to determinately think about qualities is explanatorily prior to her having a proto-theory about them.[12]

Another answer to Quinean indeterminacy appeals to rich dispositions. For instance, to think of the plus function, you must learn the right dispositions. To think of the property of being a game, you must learn to use "is a game" in the right way. (And, in the case of "is a game", even after you've acquired the right use-disposition, it is not determinate what property you are thinking of.) Or again: if I am in the dark and form a representation "D" that then causes me to move in direction D, then maybe "D" refers to direction D (Hawthorne 2007). More controversially, Fodor (1994) has suggested that, while informational-causal relations between the mentalese word "rabbit" and the world aren't enough to answer the Quinean challenge about how it refers to rabbits rather than undetached-rabbit-parts, our inferential dispositions settle the matter in favor of rabbits. Likewise, Putnam (1992, 30) and Prinz (2008) have suggested that inferential dispositions are the key to solving the "depth problem"—of which the present problem about Mary is an example.

But, clearly, dispositions cannot save the day in the case of our thought about qualities. Again, the reason is that it is extremely *easy* for Mary to more or less determinately think of the smell quality—much easier than it is to think of the plus function, and etc. Simple Mary can do it even if she lacks any behavioral or inferential dispositions that could select one of P1 or P2 as the referent. (In fact, it is not even clear what dispositions *could* do the job.)

Now at this point many reductionist philosophers will naturally appeal to Lewisian "naturalness" (Sider 2011): maybe, for instance, P1 is way more "natural" than P2, and hence a "reference magnet". If P1 has great reference magnetism, reductive externalists like Tye and Dretske can explain how Mary manages to *easily* and *determinately* refer to it rather than P2.

But there are decisive problems with this idea. First, even if we were to grant that a naturalness constraint helps pin down the contents of our basic experiences and acts of attention, this would anyway not help for a simple reason: *we can suppose that P1 and P2 are equally natural* (assuming a way of measuring "degree of naturalness").

It is also worth mentioning that this idea relies on a misunderstanding of Lewis (see Pautz 2013). Lewis thought that the naturalness constraint only applies at the level of mental content to *belief.* In particular, he explicitly derived his naturalness constraint on belief from a more general *rationality constraint* on belief: roughly the beliefs an individual has are the ones that maximize his rationality. So his naturalness constraint only applies to *beliefs* and other intentional states that are *assessable for rationality*. But the problem I'm pressing doesn't

only concern Mary's *belief* that the quality is here, but arises at more foundational level. It concerns the content of her *experience* and her act of *attention*. What makes it the case Mary is *conscious of* P1 rather than P2, and *attends to* P1 rather than P2? Mary bears perfectly good "tracking" relations to *both* of these; and they all appear to be equally good candidates for being the referent of "x is conscious of y" and "x attends to y". The rationality-naturalness constraint, as Lewis understands it, simply doesn't apply to our pre-rational experiences and acts of attention. In fact, Lewis never really gave a theory of what fixes the determinate contents of our experiences and acts of attention—the "source intentionality" that lies at the foundation of his theory. My point is that the indeterminacy worries start here.

Let me make a final point before moving on to my other examples. We must guard against a simple solution to the problem of how consciousness enables determinate thought about qualities. Imagine a reductive externalist like Tye or Dretske giving the following response:

> "Look, my view doesn't imply the claim of radical indeterminacy (6). The correct answer to your problem in the case of Simple Mary is absolutely simple. It has two parts. *First*, there is a specific quality, Q, and Mary *experiences*, *attends to*, and thereby easily thinks about. Second, as a reductive externalist, I happen to think that there are reasons to think that Q is *identical with* a chemical property P1. One reason is the *generalization argument* discussed in §1. From these two claims, it follows that Mary is referring to P1—not P2. So my view doesn't predict the claim (6) that it is indeterminate whether Mary is thinking about P1or P2. I agree that this prediction would conflict with what Mary knows from the first person—but it is not a prediction of my view. My view gets you both Determinacy and Easiness."

But this speech cannot be a solution to the problem of the reference-grounding significance of consciousness. To quickly see this, consider an analogy. Suppose you are a reductive materialist and you are asked for a reductive materialist account of how you refer to Mark Twain. Suppose you reply by saying "well, the answer is simple and has two parts: first, you refer to Samuel Clemens, and, second, Samuel Clemens is *identical with* Mark Twain". This is not yet a materialist account of how you refer to Mark Twain! In the same way to say "first, Simple Mary experiences, attends to and refers to Q, and, second, Q = P1" is not yet to provide a materialist account of how Simple Mary experience, attends to and refers to P1 rather than P2.

Moreover, this speech is also guilty of a basic confusion between two things: (i) providing a *reason* to believe Mary refers to P1 rather than P2, and (ii) a materialist *explanation* of how she refers to P1 rather than P2. What is required to answer my problem is (ii) but at best the speech only provides (i). Of course, externalists can give a reason (e. g. the generalization argument from §1) for believing that the smell quality Q is determinately identical with the chemical property P1, and hence, that this is what Mary is referring to in this case. And

Block and Papineau can give reasons for thinking that Q is neural property P2, so that Mary instead refers to neural property P2. But even if a theorist can provide sophisticated reasons to think that simple Mary is in fact determinately referring to P1 (or P2), this is not yet to provide a reductive materialist account of how simple Mary might easily determinately refer to the one rather than the other (indeed, presumably, those sophisticated reasons won't figure in this account).

Finally, one way of reading the above speech is as follows. The suggestion might be that the answer to the question "What makes it the case that Mary is thinking about the external chemical property P1 rather than internal neural property P2?" is "In accordance with transparency, Mary is *conscious of* the external chemical property; she is not at all conscious of the internal neural property P2; and she thinks about the property she is conscious of." Now, in a way, I agree with the thought behind this statement. Intuitively, Mary thinks about the fine-grained smell quality that she is *conscious of*. *So, in the first instance, the problem I am raising most fundamentally concerns how reductive materialists can accommodate the fact that there is no radical indeterminacy in what Mary is conscious of when she smells the perfume (e. g. it is not indeterminate whether she is conscious of P1 or P2).* But, when it comes to answering the challenge, this statement just *passes the buck*. Just as there is a multiplicity of *physical properties* that are candidates to be what Mary is conscious of and thinks about in this example (e. g. P1 and P2), there is a corresponding multiplicity of candidate *physical relations* to be what fixes what she is conscious of. For instance, when she undergoes an internal brain state that has internal neural property P2, she undergoes downstream neural states (constituting her cognitive response) that are causally sensitive to P2. So she bears an especially intimate causal relation, causes*, to the occurrence of P2. She also bears another, more distant kind of causal relation, causes**, to the occurrence of P1 in in the external world (something like the tracking-17 relation discussed earlier). These are both perfectly good relations. It is not very plausible that one of these relations (say causes**) constitutes what we are conscious of (or what we "encounter" in Papineau's 2016 terminology) while the other does not. For what in our history of use of the expression "x is conscious of smell quality y" might determine that it picks out causes** rather than causes*? (Likewise, if some Martian observers wonder what Mary is "conscious of" in this case, what could determine that *their* use of "conscious of" refers to one of these physical relations rather than the other?)[13] We will return to this issue later in connection with "the exquisite identities response".

So much for the example of Simple Mary and the strong smell. My discussion so far has only been meant as an initial illustration of the basic challenge of determinacy-grounding significance of consciousness for reductionists. The full force of the problem, as well my argument that it has no adequate solution, will only be apparent once we've looked at my other two examples as well as what would be required for a response.

Second Example: Photons-vs-Newtons. Now I turn to a second illustration of the problem of the determinacy-grounding significance of consciousness: I call it the *Photons vs. Newtons Case*.

Suppose Mary views a tomato in the actual world. Mary is conscious of a distinctive quality that seems to pervade a round region. She attends to it. So there is a certain quality, and she thinks about it.

As I explained in §1, reductive materialists are under strong pressure to think that this quality is a reflectance-type co-instantiated with the shape. The pressure is provided by the "generalization argument". This is the only sensible way to be a reductionist about color qualia. Let us just assume that this is right for the sake of argument. That is, let us assume that it is determinate that Mary is conscious of, and thereby thinks about, *some reflectance-type or other*. Still, there is a problem about whether it is determinate *which one* she is conscious of. For there is a complexity here that is typically ignored. There is the *photonic-reflectance* of the tomato. There is also the *functional-reflectance* of the tomato. Just as in the case of Mary and the smell, there is a multiplicity of candidates. (There is also the microstructure of the tomato's surface, which is the basis of its reflectance; however, I will set this candidate aside.)

To appreciate what I mean, we can consider the actual world and then a possible world *W*. In the actual world, what we call "light" is made up of photons. And photons have the following features. They have zero rest mass. (If they had non-zero rest mass, they would violate relativity theory, which implies that as a massive object approaches the speed of light it acquires infinite mass.) They behave non-classically. In particular, they exhibit a wave-particle duality: wave-like properties in certain experiments and particle-like properties in others. Their energy and momentum depends inversely on their wavelength (λ). Now let us turn to the hypothetical situation, *W*. In *W*, let's us suppose that the true physics turned out to be more like Newtonian physics than in the actual world. What is called "light" in this world is made up of something like Newtonian corpuscles just as Newton thought, rather than our photons. They do have rest mass. They behave "classically" in the two-slit experiment. But otherwise they are like our photons. They travel at the speed of light. Their energy and momentum depends inversely on their wavelength (λ). In *W*, newtons interact with our photoreceptors just as photons interact with our photoreceptors in the actual world. Now, there is a property that the tomato of world has but that a corresponding tomato of *W* lacks. It is the property of having a disposition to reflect certain proportions of photons. This is what I mean by a *photonic reflectance*. It is a reflectance-as-realized-by-photons. There is also a property that the tomato of *W* has but that is not possessed by the tomato Mary is viewing here in the actual world. It is the property of having a disposition to reflect certain proportions of *newtons*. It is a reflectance-as-realized-by-newtons. Let's call this the *newtonic reflectance*. But there is also a property that the tomato of this world and the tomato of *W* have in common. In fact, if scientists of both worlds exhibited the reflectance curve of the tomato and the twin tomato, they would exhibit the same curve. This is

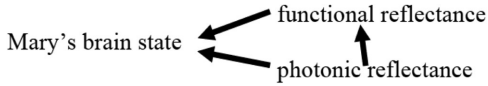

Figure 4. The more abstract functional reflectance of the tomato is realized by its photonic reflectance. Both cause Mary's brain state

the property of having a disposition to reflect certain proportions of particles playing a certain functional role (going the speed of light, having certain energies and momentum inversely related to wavelength). This is what I mean by the *functional reflectance*.

Now, as I have said, I think it is part of the determinacy-grounding role of consciousness that it enables more or less *determinate* beliefs about qualities. Once we have distinguished between photonic and functional reflectances, we see that reductive externalists face a challenge in accommodating this fact when it comes to our beliefs and other intentional states about color qualities. When Mary has her experience of the tomato, what could make it that case that Mary is *conscious of* and *thinks about* one type of reflectance property rather than the other? (This is analogous to the issue of whether pain is a neural state or a functional state. Since reductive externalists locate color qualia in the external word, they face a similar issue transposed to the surfaces of objects.)

It might be thought that the answer is simple: Mary is conscious of a reddish quality pervading the surface of the tomato. As a matter of fact, this quality is identical with (say) the photonic reflectance, not the functional reflectance. So, Mary is *conscious of* the photonic reflectance, not the functional reflectance. And she thinks about the photonic reflectance, not the functional reflectance.

But, as we just saw in our discussion of the smell case, this type of answer just passes the buck. What then makes it the case that Mary is *conscious of* the photonic reflectance rather than functional reflectance, rather than there being no fact of the matter concerning which one she is conscious of?

To drive the point home, consider a completely *insentient* robot or simple evolved creature that tracks reflectances. What does it "really" represent: photonic reflectances or functional reflectances? Surely, here *there is no fact of the matter*.

Now, in the previous case of Mary and the strong smell, the relevant candidates P1 and P2 were very different. In the present case, the functional reflectance and the photonic reflectance differ in a subtler way. So you might think, "maybe in this case moderate indeterminacy is acceptable—it is indeterminate whether Mary is conscious of the photonic reflectance or the functional reflectance, and so indeterminate which one she thinks of."

It would be most natural for the reductive externalist to develop this indeterminacy option within a *supervaluationist* framework which locates the source of indeterminacy in language (e. g. Lewis 1994).[14] On this elaboration, expressions like "this color quality" or "red" are indeterminate in reference between

a photonic reflectance and a functional reflectance. Likewise, it is indeterminate whether "is a color" refers to the property *being a photonic reflectance* or the property *being a functional reflectance*. So it's indeterminate whether tomatoes in *W are red* or whether they *have an alien, non-color property*. And it is indeterminate whether "x is conscious of y" refers to a tracking relation F-187 that Mary bears uniquely to the functional reflectance as she views the tomato, or a tracking relation F-188 that she bears uniquely to the photonic reflectance of the tomato.

One initial problem with the indeterminacy option is that it disagrees with what is obvious to Mary from the first person. Mary attends to a distinctive quality that seems to fill the round region. There is *one* quality, and it is determinate that she thinks about *it*. Contrary to the indeterminacy option, it is not the case that there are two properties (the photonic reflectance and the functional reflectance), with very different possible-world extensions, and it is indeterminate which one is conscious of and is thinking about.

There is another, more decisive problem with the indeterminacy option. To see this, let us consider a twin of Mary in the Newton world *W* viewing a tomato. There are two possibilities concerning what properties Mary and Twin Mary are conscious of (and thereby have beliefs about):

I. Mary is conscious of the quality red, which is identical with the functional reflectance. Twin Mary is conscious of the very same quality *red*, which is identical with the functional reflectance. It is just that the quality red—that is, the functional reflectance - is differently realized across the two worlds at the micro-level. Given the character-consciousness link, if (and only if) this is so, Mary and Twin Mary have the very same phenomenal experience, namely, a reddish one.

II. Mary is conscious of the quality red, which is identical with the *photonic reflectance*. Similarly, Twin Mary is conscious of a *different* property, namely, the newtonic reflectance. This is an alien, non-color property that humans cannot be conscious of (because they cannot track it) and hence cannot imagine. In that case, Mary and Twin Mary have *different* phenomenal experiences. While Mary has a reddish color experience, Twin Mary has an alien experience that is not a *color* experience at all.

Now back to the indeterminacy option. On the indeterminacy option, it is indeterminate whether Mary is conscious of the photonic reflectance or the functional reflectance. What goes for Mary goes for Twin Mary: it is indeterminate whether Twin Mary is conscious of the *newtonic* reflectance or the functional reflectance. So it entails that it is *indeterminate whether (I) or (II) is correct*. That is, it entails:

7. It is indeterminate whether Twin Mary has a *reddish experience* of the tomato (phenomenally identical to Mary's) *or* an *alien experience*, one that is phenomenologically different from Mary's experience and indeed different from any possible color experience of ours. Indeed, if you could somehow have Mary's experience and then Twin Mary's experience, you

> could not know whether *they are the same*, or whether *they are very different*, for there is no fact of the matter to know.

But this is just evidently impossible. Maybe there can indeterminacy concerning when an individual's experiences begin after the individual is born. But, given that a creature has experiences, there cannot be radical indeterminacy concerning what those experiences are like. *Of course*, there is a fact of the matter about the phenomenal character of Mary's experience when she views the tomato; she knows this by introspection. (Recall that she is just an ordinary, actual person viewing a tomato.) Equally, there is a fact of the matter about the phenomenal character of Twin Mary's experience when she views the tomato reflecting newtonic light; she knows this by introspection. So, (7) just doesn't agree with the facts; contrary to (7), there is a fact of the matter about whether Mary and Twin Mary have the same experience, or radically different experiences.

In sum: many people (Quine, Field and Lewis) hold that there is indeterminacy of reference in many cases ("mass" and even "rabbit"). But it is absurd to suppose that it holds at the level of experience in the case of Mary and the photons and newtons, for the reasons I have given. It is absurd that it can be indeterminate whether another individual is conscious of the same quale you are, or an alien quale.

So reductive externalists need to avoid the indeterminacy option because they need to avoid (7). But it is hard to see how they might do so. Think of Mary and Twin Mary as purely physical systems. If the austere physical facts are all the facts there are, then they are not enough to determine whether Mary is conscious of the photonic reflectance or the functional reflectance, and they are not enough to determine whether Twin Mary is conscious of the newtonic reflectance or the functional reflectance. How could they? The difference is so subtle (even more subtle than the difference between rabbits and undetached rabbit parts).

To drive the point home, return to the completely *insentient* robot or simple evolved creature that tracks reflectances. What does it "really" represent: photonic reflectances or functional reflectances? Surely, *there is no fact of the matter*. Then, by parity of reasoning, the same applies to Mary and Twin Mary, if reductive materialism is true and (as with a robot) the austere physical description of them is the complete description. But, given the reductive externalist model, this indeterminacy option entails (7), as we have seen.

A final comment. This essay assumes the externally directed character of experience. Given this, there is strong pressure on reductive materialists to hold that sensible colors are reflectance properties outside the head. We saw that the only real alternative—that sensible colors are Shoemaker-style response-dependent properties of external objects—is not very plausible for reductionists. However, it is worth mentioning that the present problem is quite general. It arises even on a Shoemaker-style (1994) view. Even on this view, there are two options: sensible colors are dispositions to produce *functional* effects or they are dispositions to produce *neural* effects. So the very same radical indeterminacy

challenge arises: the view runs the risk of implying the possibility of cases where it is radically indeterminate what the phenomenology of someone's experience is (e. g., that it is indeterminate whether "Commander Data" of Star Trek fame has the same color experiences as you, or radically different experiences of a kind that we cannot imagine).

Third Example: Middle Earth. Now for my third and final example. For the sake of argument, let's once against be concessive to reductive externalists. Let us grant sensible colors are determinately identical with reflectance properties. For the purposes of the present example, it doesn't matter whether they are functional reflectances or photonic reflectances. My next threat of radical indeterminacy is totally independent of this. It is also quite elaborate. But you will see its point by the end.

The example starts with a variant of Harman and Block's "Inverted Earth" case (Block 1990). There are a man and a woman, Harry and Sally, who live on different planets. Harry is just an ordinary guy here on Earth. Sally grew up on Twin Earth. She belongs to a different species that, by a remarkable coincidence, is almost exactly like *homo sapiens*. Let us suppose on Twin Earth there is only one yellow thing, and this is the sky. That is, on Twin Earth, the color of the sky is "inverted". So we can call it *Inverted Earth*. However, there is a giant lens between the sky and the earth. So, even though the sky has a yellow reflectance, the light reaching twin Earthians eyes is "blue light". As a result, when Sally and other twin Earthians view the *yellow* sky, they get neural state B, the same neural state that that Harry and other humans get when the look at the *blue* sky. And proponents of the reductive externalist model, such Dretske and Tye, would say that among inverted Earthians B realizes the experience of yellow, whereas among Earthians it realizes the experience of blue. As is well known, they must say that a creature's *history* helps determine what features they are conscious of.

Now suppose that, because of global warming on earth and on twin earth, humans and twin humans flee those planets. Both wind up on *Middle Earth*, an earthlike planet that is exactly midway between Earth and Twin Earth. The only odd thing is that there are no blue things at all, and nothing gives off blue light. It's not just that there is a *single* "missing shade of blue"—nothing has *any* shade of blue. When they arrive on Middle Earth, each species is very surprised to learn of the existence of a nearly identical-looking species from a different planet. But the planet is very large, so they get along just fine. In fact, to their delight, they find that they can interbreed.

One day Harry meets Sally from Inverted Earth. Despite belonging to different species that evolved separately, they happen to be similar enough that they can interbred. And, when Harry met Sally, that is exactly what happened. So they have a child together. They name her "Mary". Mary grows up normally. But one day her brain goes wild. She spontaneously goes into visual brain state B. She is the first person in a long while to undergo brain state B: for as I just said,

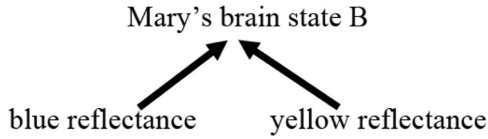

Figure 5. Mary's novel brain state B was Normally caused by the blue reflectance in her dad's population and was Normally caused by the yellow reflectance in her mom's population

on Middle Earth there is no blue light, so B doesn't occur naturally. Because she goes into brain state B, she has a vivid color hallucination.

Now here is the question for reductive externalists. What is Mary's hallucination like? Given the "character-presentation link", that depends on what color quale she is conscious of. And, given the externally directed character of experience, reductive materialists are under pressure to say that color qualia are reflectance properties (§1). So the question becomes: what reflectance property does Mary's brain state B represent, and which one does she thereby think about and know about?

Given our assumptions, there are two options. When Mary has B, she might be conscious of the *blue reflectance-type*: that is, she might be conscious of the reflectance-type that had been historically normally caused B in *her dad's species on Earth*. Alternatively, the occurrence of B in Mary might enable her to be conscious of the *yellow reflectance-type*: that is, the reflectance-type that had been historically normally caused B in his *mom's species on Inverted Earth*. (See Figure 5.) Which of these options is correct?

As before, one option is the *indeterminacy option.* We can define up a tracking relation, *tracking**, that Mary bears to the blue reflectance-type as she is having his hallucination: in particular, she *is in a state that, in his father's species, was Normally caused by the occurrence of* the blue reflectance-type. We can also define up a tracking relation, tracking**, that Mary bears to the yellow reflectance-type: she *is in a state that, in his mother's species, was Normally caused by the occurrence of* the yellow reflectance-type. These relations are equally eligible candidates to be the referent of the expression "x is conscious of quality y" in English. Given the reductive externalist theory of conscious experience, the indeterminacy option implies the following:

8. It is indeterminate whether Mary is conscious of blue or yellow. Therefore, it is indeterminate whether Mary has a bluish hallucination, or a yellowish hallucination. In fact, *Mary herself* cannot know from the first person which of these possibilities obtains—because, if it is indeterminate whether *p*, one cannot know that *p*.

But, as before, the indeterminacy option is incoherent. Sometimes indeterminacy is acceptable—for instance, indeterminacy in the reference of "mass" in Newton's mouth, or indeterminacy in whether the frog's brain state represents

frog food or *black dot*. However, radical indeterminacy in phenomenology is incoherent.[15]

You might think that reductionists can easily avoid the problem of Middle Earth: they should just give up the view that the color qualia we are conscious of are reflectance properties of surfaces in the external world. Instead, they are neural properties instantiated in the brain (Block 2010, Papineau 2016). But this is a mistake for a couple of reasons. First, I'm assuming external directedness. Given this, it is hard for reductionists to avoid the view that sensible colors are reflectance properties (§1). Second, the problem is in any case quite general. For instance, the very same type of problem applies to the consciousness of shape, which is certainly instantiated outside the brain. In a different version of the case, we could imagine that Mary is in a novel brain state B that normally tracked *round* in her mom's population and *oval* in her dad's population.[16]

The only possible reductive solution: exquisite identities. To sum up: consciousness makes possible determinate beliefs (and other propositional attitudes) about qualities. This is because the contents of our conscious states are themselves determinate. But for reductionists multiple candidate cases make a problem for this idea. They are under some pressure to accept the forms of radical indeterminacy embodied in 6–8. The Quinean indeterminacy problem applies at the very source of intentionality. The challenge of radical indeterminacy be most serious exactly where it is least plausible.

So far, I have just put the problem on the table. Now I want to get to the crux of the matter. In our discussion of Black-and-White Earth, we saw that reductionists must identify the conscious-of relation with a very fine-grained physical relation, "tracking-17". The lesson of our new multiple candidate cases is that, to accommodate the determinacy of consciousness and consciousness-based thought, the relevant physical relation must be even more fine-grained. Proponents of the reductive program must advocate a system of arbitrary, lucky-looking identities. And this leads to a host of problems. Let me take these points in turn.

To avoid 6–8, reductionists need to maintain that there are determinate facts about what physical properties Mary is conscious of and thereby thinks about. For the sake of discussion, let us just suppose that they are as follows:

9. In the strong smell case, it is determinate that Mary is conscious of and thinking about the chemical-type P1 rather than the neural property P2 (so, the smell quale that grabs her attention is in fact P1 rather than P2). In the Photons-vs-Newtons Case, it is determinate that both Mary and Twin Mary are conscious of and think about the same functional reflectance rather the photonic reflectance or the newtonic reflectance; thus, Twin Mary is determinately conscious of the same reddish quality as Mary and it is determinately identical with a functional reflectance. Hence it is determinate that they have phenomenally identical experiences.

> Finally, in the Middle Earth case, it is determinate that Mary is conscious of the yellow functional reflectance rather than the blue one (even though her relation to the two reflectance-types is totally symmetrical). So it is determinate that she has a yellowish experience.

Now, given reductive materialism, (9) entails that the identities obtain, which are unknowable even given total physical information:

10. The conscious-of relation is identical with a physical relation, call it ***F-187***, that Mary bears to P1 rather than P2, to functional reflectances rather than photonic reflectances, and (in the Middle Earth case) to the yellow reflectance rather than the blue one. (We cannot even begin to gesture at this relation. But, since relations are abundant, we know that there is such a relation. Perhaps it is some super specific version of the relation: x is in a state that Normally tracks y in the population of the *mother* of x.) The thinking about relation is identical with a coordinate physical relation, call it ***G-187***, that Mary bears to the same properties (rather than the other candidates).

Previously we saw that reductive externalists must identify the conscious-of relation with the tracking-17 relation that Mary bears the outer colors and not inner colors in the black-and-white earth case. Now we see that their reductive theory must be even more specific to avoid radical indeterminacy in my other multiple candidate cases. The conscious-of relation must be identical with the **F-187** relation. Of course, it is very hard to specify what **F-187** (and **G-187**) might be. But, given the abundance of physical relations, there is bound to a huge abundance of relations F-186, **F-187**, F-188 and so on with different extensions. The present view is that it is just a "surd metaphysical fact" that the conscious-of relation is identical with **F-187** rather than the alternatives (compare Putnam 1981, 46–48). The exquisite identities view not only accommodates Determinacy (the denial of 6–8); it also accommodates Easiness, for even unsophisticated creatures can easily bear **F-187** and **G-187** to various properties.

Now, the proponent of such exquisite identities faces a question: how does the expression "x is conscious of y" in the language of thought or in public language *refer to* **F-187** rather than F-186, F-188, and so on? And how does "x thinks about y" in the language of thought or public language *refer to* the coordinate physical relation **G-187** rather than G-186, G-188, and so on? If you looked at our use of these expressions, you would never be able to tell that. That is why I said above that the relevant identities are unknowable.

To answer this question, the present response requires *yet another exquisite identity*, namely, that the *reference relation* is identical with a physical relation **R-187** that "x is conscious of y" bears to **F-187** and that "x thinks about y" bears to G-187 (and that "x refers to y" bears to **R-187**). What the present response requires, then, is that there is a *whole system of coordinate identities*.

Here is a way of putting what has been shown. On Timothy Williamson's epistemicism about vagueness (1994), there are inscrutable semantic facts. For instance, "bald" refers to a perfectly precise hair condition **C-187**—but we will never know what it is. So, on his view, the reference relation is identical with a relation **R-187** that "bald" bears to this precise condition **C-187** rather than other. My discussion has shown that, if they wish to avoid an absurd form of indeterminacy in consciousness and thought at the most basic level, *all reductive materialists* require similar totally arbitrary-looking facts about consciousness and intentionality, no matter what theory of indeterminacy they accept.[17]

One potential problem with the exquisite identities response is that it requires *a posteriori materialism*. This is indeed a strange view. Here is a way bringing this out. Typically, when you learn an identity, such as water = H_20, you learn interesting contingent (physical) facts, for instance, that people use "water" to refer to H_20. In short, we can articulate what you learn in other terms. This helps to explain why identities can be informative. But present response requires that learning identities can be informative even if it doesn't come with learning any interesting contingent physical facts. For instance, suppose that Frank Jackson's Mary (to revert from my Mary to Frank Jackson's Mary) knows all the fundamental physical facts. On the present view, there remains a whole system of coordinate identities for her to learn. For instance, suppose God whispers to her that the conscious-of relation = **F-187**, that the thinking-of relation is **G-187**, that the reference relation is F is **R-187**, that sensible colors are functional reflectances (not photonic reflectances), and so on. On the present view, she thereby learns very significant information about the world. But this information cannot be articulated in any other terms. In particular, she doesn't learn any knew physical facts about word usage and so on: she already knows all these facts. Therefore, the alleged new information is supposed to be at the same time highly significant but also in a way impenetrable or opaque. For this and other reasons, *a posteriori* materialism is hard to believe. It is even hard to understand.

But I do not want to press this problem. My view is that, even if we allow inscrutable identities, the exquisite identities response is deeply problematic. Therefore, my determinacy problem goes beyond the standard "epistemic gap" problem.

What the present section has shown is that, to avoid radical indeterminacy in consciousness and consciousness-based thought (viz. 6–8), reductive externalists also require a bizarre form of *metaphysical arbitrariness*. For instance, as Mary looks at a tomato, she bears a physical relation **F-187** to the functional reflectance and a physical relation F-188 to the photonic reflectance. The present response requires that it is just an arbitrary, surd fact that the conscious-of relation is identical with **F-187** rather than F-188, and that the sensible color is identical with the functional reflectance rather than the photonic reflectance. And it is just an arbitrary fact that bearing **F-187** to the functional reflectance (the quality red, on this view) enables Mary to easily have *de re* thoughts about it, while her bearing the intrinsically near identical relation F-188 to the photonic reflectance

(the *realizer* of the quality red, on this view) doesn't enable her to easily have *de re* thoughts about *it*. (On this view, experience directly enables *de re* thought about the quality—the functional reflectance—but not about its physical realizer—the photonic reflectance.) Likewise, in the Middle Earth case, it is just an arbitrary fact that that when Mary undergoes B, she is conscious of the quality yellow (the yellow reflectance that B tracked in her *mom's* population) rather than the quality blue (the blue reflectance that B tracked in her *dad's* population). The same applies to Mary and the strong smell: it is just a brute, inscrutable fact that **F-187** (a relation that Mary bears uniquely to P1) rather than F-188 (a relation that Mary bears uniquely to P2) that constitutes the conscious-of relation (see footnote 13).

This whole system of extremely arbitrary, coordinate identities is self-consistent but just *intrinsically unbelievable* because it is exceedingly arbitrary. Why does one exquisite system of brute identities hold, and not an ever-so-slightly different system?[18] Here is another way to put it. Nearly everyone rejects epistemicism (Williamson 1994) because it requires bizarre arbitrary semantic facts. But by the same token we should all reject reductive externalism because it requires a similar kind of arbitrariness. And if I am right that, given external directedness, reductive materialism leads to reductive externalism, this means that we should reject reductive materialism.

If reductive materialists accept the exquisite identities view for Mary, shouldn't they accept it for the Simple System too? After all, the Simple System is just another purely physical system. On such a view, the Simple System "really" determinately represents only functional reflectances, not photonic reflectances (because the representation relation is exquisitely identical with a relation, **R-187**, that the Simple System bears to functional reflectances and not photonic reflectances); likewise, it "really" represents chemical properties of odorants and not their neural signatures; and so on and so forth. There are all these exquisite intentional facts about the Simple System we could never know about. This is just madness! But if it is a crazy view about the Simple System, it is a crazy view about Mary too.

At the very least: if we *can* avoid extreme arbitrariness in our explanation of the role of consciousness is fixing determinate intentionality, we should. And we can—with a non-reductive theory. In fact, with a non-reductive theory, we can provide a satisfying explanation of *all* the ways in which consciousness is significant. At the same time, we can retain allegiance to materialism. I now turn to these points.[19]

5. Sketch of a Nonreductive Explanation of the Significance of Consciousness

I have argued that the reductive externalist model of the conscious-of relation cannot plausibly accommodate the various ways in which consciousness is significant: its dissimilarity-grounding significance, reason-grounding significance, and

determinacy-grounding significance. I will now sketch a nonreductive-internalist model of the conscious-of relation, and then I show that it can provide an attractive and unified explanation of these central facts about consciousness. The resulting view has some similarities to Russell's (1912) view on which acquaintance lies at the foundation of the mind. It provides a *consciousness-first picture* of the mind, somewhat as Williamson (2002) has advocated for a knowledge-first picture in epistemology. As long as we remain in the shackles of reductionism, we cannot explain these facts. By contrast, if we turn to a non-reductive view, we can give a non-trivial explanation of them.

I will begin by just laying out, without argument, the form of nonreductionism I think we should favor. Then I will show, in a series of steps, that it can provide a satisfying, unifying explanation of the ways in which consciousness is significant.

The nonreductive internalist model of the conscious-of relation. To begin with, the alternative model I would like to propose is internalist, holding that the character of consciousness is fully determined by the state of the brain. This is in contrast to the reductive externalist model, which explains consciousness in terms of informational-teleological relations to the environment. The Significance Argument suggests that externalism about experience is just a mistake: it is bound to result in radical phenomenal indeterminacy and is bound to violate the significance of consciousness. Elsewhere I have developed an Empirical Argument against the reductive externalist model that also suggests that externalism about experience fundamentally misguided. The whole history of psychophysics and neuroscience suggests that the explanation of consciousness is to found in the brain.[20]

The nonreductive internalist model I favor can still accommodate what I have called the "externally directed" character of experience—the starting assumption of this essay. The picture is that consciousness is both internally dependent *and* externally directed. There is nothing problematic about this: thought about numbers is externally directed but internally dependent. In the same way, the idea is that perceptual intentionally is externally directed but internally dependent.

For instance, consider the "brain in the void". On the nonreductive internalist model, thanks to its internal states, the brain in the void is conscious of external properties, for instance, shapes, positions, distances. On this view, the brain just has an innate capacity to enable one to be conscious of a certain range of basic perceptible properties, properties that are typically not instantiated in the brain itself. (Compare the sense datum view of Russell 1912.) Of course, the brain in the void doesn't bear the tracking relation to any of these properties. So, on this view, the conscious-of relation cannot be identical with the tracking relation.

This brings me to a second defining feature of the reductive internalist model: there is *no* interesting identification of the form "for x to conscious of property y is for x to $\ldots$ y". It is not even identical with some massive disjunction of

physical relations. There is just nothing interesting to say about *what this relation is*. In this sense, the relation is primitive, just as Russell held that acquaintance is primitive. In fact, the nonreductive internalist model I favor is obviously very Russellian. But rather than holding that consciousness relates to us things ("sense data") that have properties, the reductive internalist model I favor holds that, in illusion and hallucination at least, consciousness relates us to properties without relating us to things having the properties.

The nonreductive internalist model I favor also endorses a nonreductive account of *sensible properties*. Colors are not identical with reflectance properties, smell qualities are not identical with chemical properties, and so on. In general, there is no completion of schemas like *for something to be red is for the thing to be F* (Pautz 2018).

The nonreductive internalist model is compatible with both *realism* and *irrealism* about the sensible properties. For instance, Colin McGinn (1996) combines this view with realism. He holds that, although sensible properties (traditional "secondary qualities") are primitive, they are response-dependent: a thing has a primitive sensible property iff it normally appears to have that sensible property. This is not true of "primary qualities" (shapes, locations, and so on): following tradition, he treats secondary qualities and primary qualities differently. So, on his view, before the evolution of sentient creatures, things only had primary qualities and no secondary qualities. Then brains evolved that have the intrinsic capacity to enable creatures to be conscious of things as having secondary qualities as well as primary qualities. On his view, those things thereby *acquired* those sensible properties. Other proponents of the nonreductive internalist model are *irrealists,* for instance David Chalmers (2006), Terry Horgan (2014) and myself (Pautz 2006, 2018). We reject McGinn's "rule" that guarantees that things have the sensible properties they normally appear to have. In their view, things don't have any of the "secondary qualities" that we perceptually represent, even if they may have the "primary qualities" that we represent. (Nor do sensible properties qualify our experiences or parts of our own brains: the idea is that they qualify absolutely nothing at all.) When it comes to sensible properties, the brain is a projective apparatus; they only live in the contents of our experiences. Of course, this is a traditional Lockean view.

The nonreductive internalist model may seem to require a kind of dualism on which there are contingent psychophysical laws linking brain states with bearing the primitive conscious-of relation to primitive sensible properties. This is not so. It is also quite compatible with the materialist doctrine that everything is grounded in the physical. Such an approach would be very similar to dualism, but it would replace psychophysical laws with "grounding laws". To make this clear, I offer an analogy.

Many philosophers endorse *reasons fundamentalism*, for instance Parfit (2011) and Scanlon (2014). They claim that there is no interesting identification of the form *for p to be a reason for x to a is for it to be the case that* $\varphi(p, x, a)$, where the right-hand side is filled with non-normative vocabulary. In

this sense, the reasons-relation is primitive, just as nonreductive internalists hold that the conscious-of relation is primitive. Why accept this view of the reasons-relation? For one thing, no one has provided a halfway plausible example of such an identification. For another, facts about what you have reason to do, or what you *pro tanto* ought to do, just "seem different" from non-normative facts about what is the case. Now, even if reasons fundamentalism is true, this doesn't mean that normative facts about what you ought to do can completely "float free" from non-normative facts about what is the case. On the contrary, it is quite intuitive that they don't float-free in this way: whenever you have a reason to do something, this is grounded in some non-normative facts (together perhaps with a normative principle). There can be *grounding without reduction.*

In the same way, nonreductive internalists could say that, whenever you stand in the irreducible conscious-of relation to a property, this is grounded in (say) your being in a certain brain state by way of necessary grounding laws governing how this relation is dependent on brain states. Likewise, following McGinn, they might say that sensible properties are primitive, but grounded in complex dispositional relations between objects and perceivers. Again, the model here is *grounding without reduction*. On this view, even though states of consciousness are not identical with arrangements of the fundamental physical properties, they cannot float free from such arrangements (contrary to dualism, "Zombies" are impossible), any more than the normative facts can float free from the non-normative facts. There is, then, on this view a sense in which the conscious-of relation is primitive (it has no bi-conditional real definition in physical terms) and a sense in which it is not primitive (it is nevertheless *grounded in* the physical). I am myself neutral between such a robust nonreductive materialism and out-and-out dualism (Pautz 2010).

Of course, in any form, the nonreductive internalist model is complex. It provides a "layered" picture of reality rather than a "flat" picture of reality. But there are strong reasons to prefer it to the reductive externalist model. First, the reductive-externalist approach has all the marks of a "degenerating research program" beset by problems of detail (the depth problem, the disjunction problem, and so on) that have never been adequately solved (just as Williamson (2002) has noted there is a history of failed attempts to reduce knowledge). Second, there is an Empirical Argument: because it is internalist, the nonreductive internalist model fits better than the externalist reductive model with empirical evidence of the role of the brain in determining conscious experiences (Pautz 2010, 2016, 2018). Third, there is the Significance Argument of this essay: while, as we have seen, the externalist reductive model cannot accommodate the various ways in which consciousness is significant, the nonreductive internalist model can provide a satisfying, unified explanation, as we shall see. Rather than taking consciousness to be something that must be explained in other terms, we can make better progress in understanding the mind if we take it to be a starting point from which to explain other things. In what follows, I will develop such an explanation in three steps, with each new step building on the previous one.

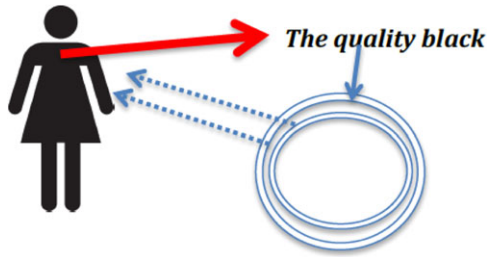

Figure 6. The nonreductive internalist account of the Black-and-White Earth case

Step 1: the dissimilarity-grounding significance of consciousness. Recall that the reductive externalist model violates the dissimilarity-grounding significance of consciousness (§2). The conscious-of relation is the tracking-17 relation, which is intrinsically just like other tracking relations, such as tracking-16, tracking-18, and so on, that are *not modes of consciousness at all*. So it implies "ludicrous similarities" in the black-and-white case involving Mary. It entails that a state S can be *intrinsically exactly like* a state of consciousness even though S is not a state of consciousness at all! We know that this cannot happen, just as we know that something cannot be intrinsically just like a color without also being a color, or intrinsically just like a number without also being a number.

By contrast, the nonreductive internalist model is quite compatible with the dissimilarity-grounding significance of consciousness. For, on this view, the conscious-of relation is *not* to be identified with a mere tracking relation, which is intrinsically like other tracking relations that are not modes of consciousness at all. Instead, proponents of this model can say that the conscious-of relation is *unique*: there is no relation R* that is not a form of consciousness but that is intrinsically similar to the conscious-of relation.[21]

Here, then, is how this model accounts for the Black-and-White Earth case. (Figure 6.)

On this view, the conscious-of relation (represented by the bold, red arrow) that Mary bears to the quality black is *nothing like* any of the mere tracking relations (represented by dotted arrows) that Mary bears to reflectance properties. So it avoids the ludicrous result that a state S can be *intrinsically exactly like* a state of consciousness even though S is not a state of consciousness at all.

Step 2: the reason-grounding significance of consciousness. When you are conscious of qualities, this enables you to know what they are like—to know their patterns of similarities and difference. Moreover, when you are conscious of an ostensible state of affairs (when, to use Pryor's 2000 terminology, you are presented with it with "phenomenal force"), then you have a *prima facie* reason to believe that this state of affairs obtains. Consciousness has a reason-grounding significance.

Recall that Black-and-White Earth shows that reductive externalists cannot plausibly accommodate these obvious facts (§3). On their view, the conscious-of relation is the tracking-17 relation, and the achromatic colors Mary is conscious of are the reflectance properties of the outer objects that she bears this relation to. But what should proponents of this view say about the tracking-18 relation that Mary bears to the colors (reflectance properties) of inner objects on Black-and-White Earth? If they say that this relation has the same reason-grounding significance as tracking-17, then they get an absurd over-generation of reasons: Mary has a reason to believe all kinds of things about the chromatic color properties of the inner objects *even though she is not at all conscious of them and is only ever conscious achromatic colors*. If, on the other hand, they say that the tracking-18 relation *doesn't* have the same epistemic significance as the tracking-17 relation, then they violate the small difference-principle. They get "extreme normative arbitrariness". It would be like saying that friendship-17 has great normative significance but friendship-18 has none, or that pain-17 has great normative significance but pain-18 has none.

The problem disappears once we accept the nonreductive model. The solution flows directly from my first point: unlike reductive externalists, nonreductive internalists can accept the dissimilarity-grounding significance of consciousness. They can say that there is a profound difference between this relation and all other relations. Because of this, they can say that this relation has a reason-grounding significance that other relations don't also have, *without violating the small-difference principle.* To illustrate, return to Mary on Black-and-White Earth. On the nonreductive model, thanks to her brain state, Mary bears the irreducible conscious-of relation to various irreducible achromatic colors that are distinct from reflectance properties (as illustrated in Figure 6). She thereby has a reason to believe various things. For instance, following Russell, she has a reason to believe things about the resemblances among those achromatic qualities. True, she bears the tracking-17 relation to the reflectances of outer objects, and the tracking-18 to the reflectances of inner objects. But this doesn't mean that she *also* has immediate reason to believe *anything* about these reflectance properties (an "over-generation of reasons"), for on this view these mere tracking relations don't have the same reason-grounding significance as the conscious-of relation (indeed by themselves they have *none*). And this doesn't violate the small difference principle, because there is a *big* difference between the conscious-of relation and these mere tracking relations. Analogy: to say that the property *being in severe pain* but not the property *being a rock has* a normative significance doesn't violate the small difference principle (it doesn't require "normative arbitrariness"), since there is a *big* difference between these properties.

Of course, even if nonreductionists can *accept* the reason-grounding significance of consciousness without violating the small-difference principle, this doesn't mean that they can *explain* the reason-grounding significance of consciousness. So it by itself doesn't answer a question raised by Jim Pryor (2000, footnote 37): *why* does the conscious-of relation have a reason-grounding

significance while other relations (e. g. mere tracking relations you can bear to things, or mere spatial relations you can bear to things while you're asleep) do not? My answer is that there is no answer. This is a fact about the *constitutive essence* of the conscious-of relation, and in general facts about constitutive essence don't have any explanation (Fine 1994). Compare: *why* does being in severe pain have a normative significance but being a rock doesn't have that normative significance? It is not implausible to think that here we have reached normative bedrock. Above I used Parfit and Scanlon's "reasons fundamentalism" about the reason-grounding relation as a model for my own nonreductive theory of the conscious-of relation. I also happen to think that their theory is the correct model for reasons. The resulting combined picture is that facts about our brain states (together with nomic laws or grounding laws) explain facts about what ostensible states of affairs we are conscious of; these facts in turn explain what reasons we have.

It only remains to explain the determinacy-grounding significance of consciousness (§4). How does consciousness make determinate intentionality possible? I will propose as the best explanation that this is deeply related to the reason-grounding significance of consciousness that we just discussed, and that the connection is mediated by a general Lewisian theory of intentionality. So first I provide a quick explanation and defense of this approach.

Interlude: a Lewisian theory of intentionality. Lewis (1994) defended a holistic *best-systems theory* of content-determination for belief and desire. The main rival is the atomistic, language-of-thought approach of Jerry Fodor (1987). Let me explain these two views by way of an example; this example will also enable us to construct a novel argument for the Lewisian approach over the Fodorian approach.

For the sake of argument, let us suppose that there is something like a language of thought (Fodor 1987, 2008). The example to be discussed shows that, *even if* there is in some sense a subpersonal language of thought, it does not determine the contents of your person-level beliefs and desires, contrary to Fodor. To keep the example simple, let us also suppose that you have lost have the capacity to speak English or any other public language. In fact, you do not even have the capacity to think in English *sotto voce*.

Here then is the example. Suppose you take part in a simple psychophysical experiment. In the experiment, you are shown color patches side by side, and you are trained to press a button if their *apparent colors are distinct*. Maybe you get some kind of reward for doing this. Suppose that, after being trained to do this, you are shown the patches in Figure 7 below in ideal perceptual conditions. Then it clearly seems to you that the apparent colors are distinct and you notice that they are. This guides you in tapping on the button. Evidently, in this trial of the experiment, these facts about your experiences and behavioral dispositions are enough to make it the case that you *believe* that the apparent colors are distinct, and that you *want* to touch the button when they are.

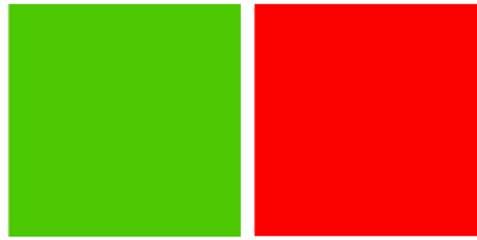

Figure 7. you are shown these color patches in trial 1 and trial 2

Now imagine that you are shown the very same, differently-colored patches over again. Suppose that, in this second trial, all your experiences, behavioral dispositions, and functional organization are exactly the same as in the first trial. The only difference is that all this is *realized differently* at the neurocomputational level. In particular, because of some short-lived neural aberration, your subpersonal neural states are re-organized in such a way that the Fodorian would say that the sentences "the apparent colors are *the same*" and "if the colors are *the same* and I press on the button, I will be shocked" enter your "belief-box", while the sentence "I will be shocked" enters your "desire-box" (in the sense of Fodor 1987).

What is the correct verdict on what you believe and desire in the second trial? Let us consider the Fodorian theory of content-determination first. We have stipulated that in this second trial everything is the same from the inside. So, for all the world, it seems that you once again believe what is obvious, that the apparent colors are *different*. You have a vivid *experience* of these two distinct colors, and they *phenomenally seem* distinct to you. You *notice* their difference. And this guides you in tapping on the button—which your history of experiences tells you will result in a reward. Nevertheless, on Fodor's view, you do *not* in fact believe that the apparent colors are different! For, on his view, *you count as believing the contents of the sentences that go into your "belief-box".* So, on his view, you "really" secretly believe that the apparent colors are *the same*, even though you have a vivid experience of their difference. Thus, you "really" have an extremely irrational belief, which is totally out of whack with your experience. On Fodor's view, then, you have undergone a radical doxastic shift in your introspective color belief between the first trial and the second trial, but it is a radical doxastic shift that you *do not and cannot notice*. Likewise, for all the world, it appears that you do *not* want to receive an electric shock by pressing on the button. For, if you *did* receive a shock, you would do everything in your power to make it stop. Nevertheless, on Fodor's view, in the second trial, you "really" secretly desire to receive an electric shock, because the sentence "I will receive an electric shot" enters your desire-box. This belief and this desire—the obviously false belief that the apparent colors are the same and the insane desire to be shocked—*are totally out of whack with your experiences and would not show*

up in any of your possible behavior (including your inner "mental actions" as well as your publicly observable behavior). For, by stipulation, all that has remained the same. In that sense, they are entirely "secret". Thus, Fodor's view entails the possibility of *secret scrambling* for belief and desire. Intuitive, this prediction about the case is absolutely crazy.

In my view, the rival Lewisian "best systems" approach provides a much more reasonable account of this case. To begin with, let us think of the neural reorganization that takes place in a *neutral* way. Fundamentally, all that is going on is that there is a neural reorganization that *preserves total functional organization*. In my view, we should think of this case in the same way we think of other cases where there are neural differences but functional organization stays the same: as a case of *different neural realization of the same mental states*. That is, contrary to the Fodorian approach, in the second trial as well as in the first, you believe that the apparent colors are *distinct*, and you *don't* want to receive an electric shock; it's just that your mental states are realized differently at the neuro-computational level. And this is exactly what the Lewisian "best systems theory" says about the case. Roughly, Lewis's idea is that *you have such-and-such beliefs and so-and-so desires iff all the "best systems" assign to you those beliefs and desires, where the best systems are the ones that have you departing least from the principles of theoretical and practical rationality, given your total history of conscious experiences and conscious behavioral dispositions*. In both trials, the "best systems" are ones according to which you believe that the apparent colors are distinct and you want a reward and you believe that you can receive a reward by touching the button when the apparent colors are distinct. So, in both trials, that is what you believe and want. Unlike Fodor's atomistic approach, Lewis's holistic approach doesn't imply "secret scrambling" in this case. Notice that this is a *rationality-based* theory of belief and desire. It holds that there is a constitutive connection between an individual's *reasons* and her beliefs and desires.[22]

A clarification. As Lewis notes, his best systems theory is compatible with the *existence* of something like a "language of thought". For it is not a theory about how beliefs and desires are in fact realized in the brain; rather, it is a theory of *content-determination,* which is officially neutral on how beliefs and desires are in fact realized in the brain (Lewis 1994, 422).[23] It says that, *even if* there is something like "language of thought", the *contents* of beliefs and desires are *not* fixed in an atomistic, building-block way by the contents of elements of the language of thought; rather, they are fixed in a more holistic way. This is supported by the above thought-experiment. Here is an analogy that may help bring out the force of the thought-experiment. Everyone agrees that the public language sentences we utter don't always determine what we *believe.* That is much too simple. For instance, we *misspeak.* For instance, the other day I pointed out a tree to my daughter and said "that's an *assiduous* tree", but I did not believe that it is a hardworking tree. But if you accept this point, you should accept that the above thought-experiment undermines the simple Fodorian approach to content-determination. For the moral of that thought-experiment is just that *the same*

point applies to the subpersonal language of thought, if such there be. Even if there is a language of thought, the contents of our beliefs and desires aren't necessarily the same as the "contents" of subpersonal, language-like states in the brain. That is much too simple. Rather, they are determined in a more holistic fashion.[24]

Step 3: The determinacy-grounding significance of consciousness. If we now combine the unique reason-grounding significance of consciousness (step 2) together with the best systems theory that connects an individual's reasons with his beliefs and desires, we can finally explain the determinacy-grounding significance of consciousness. Moreover, we can do this in a way that avoids the extreme arbitrariness worries that plague the only available reductive externalist account of determinacy ("exquisite identities"). To illustrate, let us apply these ideas to the "many-candidate cases" from §4.

Consider first the case of Mary and the strong smell. There is a unique smell quality and Mary thinks of *it*. On the reductive externalist view, given the multiple candidates (P1, P2, and all the other elements of the causal chain), there is a puzzle about how the austere physical facts could make it the case that Mary thinks determinately about *one* property rather than the others, without positing a bizarre form of arbitrariness. On the present account, the problem goes away. On this view, the mistake is to think, with reductionists, that the austere, narrowly physical facts are all the facts. There are more facts than this. In addition, there is the fact that Mary bears the irreducible conscious-of relation to a certain irreducible smell quality Q; this fact depends on, but is additional to, her brain state. The conscious-of relation is distinct from any mere tracking relation. And this smell quality Q is distinct from P1 and P2. Of course, once we accept this view, we still face the question: what makes it the case that Mary thinks of Q rather one of the many physical properties P1, P2, ... that she bears various tracking relations to? But now we have an answer—*and one that avoids extreme arbitrariness*. Owing to her brain state, Mary is only *conscious of* Q. The conscious-of relation, unlike all those mere tracking relations, has a unique reason-grounding significance: Mary therefore has a reason to believe that Q is instantiated before her, *not* P1, P2, or ... Moreover, as we saw, it is my view that it is in the nature of belief to be sensitive to reasons. Therefore, Mary's belief determinately concerns Q, and *not* P1, P2, ... and so on: for it is the attribution of *this* belief to her that has her responding to her reasons. It is not arbitrary to suppose that the conscious-of relation—rather than all these tracking relations—plays a thought-grounding role in this case: the reason is that it—rather than all these tracking relations—plays a reason-grounding role, and thought is constitutively sensitive to reasons.[25]

This view also explains how Determinacy and Easiness can be true together—indeed, why they are at their maximum when it comes to thought about qualities. This is puzzling if the austere physical facts are the only facts that they are (for they are too impoverished), but becomes expected on my approach. When you are conscious of a content involving a quality, then, because

this relation has a unique reason-grounding significance, you have a strong reason to believe that content; the content thus becomes a "belief-magnet". Given the present consciousness-based best systems theory, unless conditions are not normal, you easily and automatically count as determinately believing that content, because the assignment to you of the belief maximizes your rationality. Nothing more (behavioral dispositions, descriptive fit, etc.) need be in place.

Consider next the even tougher cases for reductionists: Photons-vs-Newtons and Middle Earth. In each case, the reductive externalist faces a problem in accounting for the determinacy of consciousness and consciousness-based thought. In the tomato case, Mary's brain state tracks the photonic reflectance and the functional reflectance. Which one *really is* the color quale that she is conscious of? A determinate answer would require that the conscious-of relation is **D-187** rather than D-188. In the Middle Earth case, she is in brain state B, which was normally caused by the blue reflectance in her dad's population and the yellow reflectance in her mom's population. What color quality is she conscious of? Again, a determinate answer would require a bizarre kind of metaphysical arbitrariness.

The source of the radical indeterminacy problem is the same in each case: the "externalist" hypothesis that sensible qualities are objective physical properties and that we are conscious of them by tracking them. This is just the wrong approach. *It is just a mistake to think that the source of phenomenal intentionality is tracking-teleological relations to the environment*. The right approach is rather this. In each case, Mary's *fine-grained internal brain state* is the basis of her being conscious of a determinate color quality C. The conscious-of relation is not a mere tracking relation and the color quality C is not any reflectance property. Once again, we still face the question: In these two cases, what makes it the case that Mary thinks determinately about C on this occasion and not any other the other candidates, namely the physical properties that she tracks: the photonic reflectance, the functional reflectance, the red reflectance, the green reflectance, and so on? But the nonreductive internalist can provide an answer—*and one that avoids extreme arbitrariness*. The conscious-of relation, unlike all those mere tracking relations, has a unique reason-grounding significance: Mary therefore has a reason to believe that C is instantiated before her, *not* any one of the many reflectance properties. Moreover, as we saw, it is my view that it is in the nature of belief to be sensitive to reasons. Therefore, Mary's belief determinate concerns C, and *not* P1, P2, . .: for it is the attribution of *this* belief to her that has her responding to her reasons.

A final comment on the determinacy-grounding significance of consciousness. Many, like Bertrand Russell (1912) and contemporary naïve realists, take it to be a brute fact about the relation of conscious acquaintance (unlike, say, the relation of *being spatially close to*) that, when we bear this relation to an item, it is easy to determinately think about the item. I have proposed that this is *not* a brute fact: the explanation derives from more basic facts, namely the intermingling essences of conscious-of relation and the belief-relation. In particular,

it is in the essence of consciousness to provide reasons (Pryor, Russell) and it is in the nature of beliefs to listen to reasons (Lewis, Davidson). Thus, conscious experience and thought fit together like hand and glove.[26]

Summary. It may be helpful to summarize this theory of the significance of consciousness

Step 1: Because the conscious-of relation is not reduced to some variant of a tracking relation (where there are many intrinsically similar tracking relations R*, R**, R***, in the vicinity), this relation can have a unique dissimilarity-grounding significance and be unlike any other relation in nature.

Step 2: Because the conscious-of relation has a unique dissimilarity-grounding significance and is unlike any other relation in nature, it can also have a unique reason-grounding significance that no other relation has, *without violating the small-differences principle.*

Step 3: Because the conscious-of relation has a unique reason-grounding significance that no other relation has, and since thought is constitutively sensitive to reasons, bearing *this* relation to property *non-arbitrarily* grounds determinate thoughts about *that property*, even in "multiple candidate cases".

Let me make a final point. On the externalist reductive model, thought and reasons are *explanatory prior to* (and independent of) conscious experience: a conscious experience is a representation of an external property in the brain that is poised to lead to thoughts or provide reasons (Dretske 1995, Tye 2000). My view also recognizes a deep connection between consciousness on the one hand and reasons and thought on the other, but it reverses the order of explanation. Instead of taking consciousness as something to be reductively explained, it takes it to be a starting point from which to explain otherwise puzzling mental capacities. In this way, once we are free from the shackles of the failed program of reductive materialism, we can make progress on the traditional puzzles about the mind. My account, then, takes a *consciousness-first approach.*

6. Conclusion

I have developed a novel argument for irreducibility of consciousness, the *Significance Argument*. It takes the form of an inference to the best explanation. Given the externally directed character of experience, the only way of reducing the conscious-of is some version of the externalist-reductive program: for our only reductive models for explaining directedness are externalist. But the externalist reductive model cannot plausibly accommodate the various ways in which consciousness is significant. If we move to a nonreductive model of the conscious-of relation, then we can adequately explain these obvious facts.

The nonreductive internalist model of consciousness I have sketched provides a complex, layered picture of reality, rather than a simple, "flat" reductive picture. But it is quite compatible with a robust materialism and the causal closure of the physical world. And it is a complexity required by the facts. Maybe this is just the way the world is.[27]

Notes

1. I discuss these points in greater detail in my forthcoming book (Pautz forthcoming). To clarify, the proto-theory I have just laid out only says that there is a necessary co-variance between experiences and the consciousness of properties. This does not require that to have an experience *just is* to bear the conscious-of relation to a cluster of properties (an abstract object of a certain kind). Maybe experiences are more basic, concrete states that *ground* bearing the conscious-of relation to a cluster of properties (Pautz 2016, p. 36 and fn. 7).
2. The reductive externalist model is defended by Armstrong (1968), Byrne and Hilbert (2003), Dretske (1995), E. J. Green (in discussion), and Tye (2000), among others. Neander (2017) and Williams (MS) defend a reductive externalist account of the intentionality of conscious experience. If they think that the intentionality of conscious experience is bound up with how things phenomenally appear (otherwise I do not know what they mean by "the intentionality of experience"), then they defend a version of the reductive externalist model of phenomenal consciousness. If *naïve realists* (e. g. Fish 2009) were to develop their view in a reductive way (with the sensible properties identified with ordinary physical properties and the relation of conscious-acquaintance identified with a complex world-brain causal relation), then they would also defend a version of the reductive externalist model. All these views are vulnerable to the Significance Argument of this essay (as well as the Empirical Argument of Pautz 2010, 2016, 2018).
3. One might hope that a theory of the conscious-of relation that invokes an "interventionist condition" (Neander 2017, 270-271) would imply that Mary is conscious of the color of the outer object and not the color of the inner object. However, this idea faces several general problems (see E. J. Green's contribution to this volume for discussion). In any case, as I say in the text, I will be assuming for the sake of argument that the "depth problem" has *some* solution; it does not really matter what it might be.
4. Another wrongheaded reaction to [#] would be this: "The similarity claim [#] is false because tracking-17 a property (that is, really being conscious of it) involves the capacity to *cognitively access* the property, whereas this is not true of tracking-18 a property." Against this, there is no big difference here. For notice that *both* tracking-17 and tracking-18 were defined above in terms of cognitive accessibility. In particular, tracking-17 can be defined in terms of *cognitive accessibility-17*: when Mary undergoes an internal sensory neural state that is caused-17 by the grey of the outer object, she is easily able to *think-17* about it, that is, to form a thought-representation S in her brain that is *caused-17* by it. Tracking-18 is similarly defined in terms of cognitive accessibility-18: when Mary bears cause-18 relation to the red of the inner object, she is equally easily able to *think-18*

about it, that is, to form a thought-representation S in her brain that is *caused-18* by it. So in all respects tracking-18 and tracking-17 - and hence Mary's physical relation to outer black and inner read - are nearly identical. They have the same real definition, except that "cause-18" and "cause-17" and interchanged throughout. A Martian observer who knew all the physical facts about Mary could truly say that her tracking-18 relation to the inner red is barely different from her tracking-17 relation to the outer black. (Here I am indebted to discussion with Jeff Speaks.)

5. A bit more precisely: if (necessarily, whenever one has P, this directly grounds one's having a reason to take attitude A - belief or desire, say - to state of affairs S) and (P* is intrinsically very similar to P), then (necessarily, whenever one has P*, then this grounds having a reason to take that *same* attitude A to a *similar* state of affairs S*). Notice that this principle allows that small non-normative differences (in the Ps) can add up to big normative differences, so it doesn't lead to sorites-paradoxical reasoning. Notice also that it is not being assumed that a property P has normative significance only if it is an intrinsic property. Indeed, one of my examples above was about friendship-17 with someone, which is extrinsic. The principle implies that, since this extrinsic property is very similar to friendship-18 (another extrinsic property), friendship-18 must have a similar normative significance.
6. It might be replied that the reductive externalist who provides a reductive account of acquaintance in terms of tracking-17 and not tracking-18 might provide a *coordinate reductive account* of epistemic properties like justification based on tracking-17 and not tracking-18. This would entail the restrictive option (4). But this doesn't at all avoid the problem I'm now raising for (4). For it still violates the small difference principle. There is just no way of getting around this.
7. This bears on a view recently defended by Geoff Lee in a fascinating essay (2018). Lee argues that reductive materialism leads to a view that he calls *deflationary pluralism*: non-experiences as well as experiences can be equal in epistemic and normative status. (For related discussions, see Hawthorne 2006 and 2007.) However, Lee only considers between-subject cases, such as a human and an unconsciousness robot. I think that his deflationary pluralism comes to grief when it comes to within-subject cases like Mary on black-and-white earth. In this case, deflationary pluralism would recommend accepting (5), which we found to be absurd. (In fact, given certain assumptions, deflationary pluralism leads to the result that a single subject can have equal justification for believing *incompatible* things.) And this would be on top of the ludicrous similarities 1-3.
8. Mark Johnston has discussed the issue of whether materialist theories in general are subject to "arguments from below" (2010, 306-316). The argument of this section against reductive materialism has been totally different from arguments from below. In the present case, a simple argument of this kind might go as follows: since the tracking-17 relation obviously has no epistemic significance whatever and since, on the reductive externalist model, the conscious-of relation *just is* the tracking-17 relation, this model implies that the conscious-of relation has no epistemic significance whatever. Unlike this argument, my argument did not start with an unsupported assertion that the tracking-17 relation has no epistemic significance whatever. Rather, I have shown that, given the reductive externalist model, the claim that the tracking-17 relation (that is, on this

model, the conscious-of relation) has epistemic significance would require either a *normative discontinuity* or else an *overgeneration of reasons*, neither of which is plausible. So my present problem for reductive is quite different from a simple "argument from below". I would also not endorse "argument from below" reasoning in connection with *nonreductive* materialism. Indeed, in §5, we will see that the discontinuity/overgeneration dilemma can be avoided by nonreductive materialism and I am open to such a view. I would not deploy an "argument from below" argument against this view, because I see nothing wrong with the idea that, even if our underlying physical states (e. g. brains states) do not have intrinsic epistemic and normative significance, they ground experiences distinct from them that do have such intrinsic significance.

9. I have supposed that Mary is a conceptually unsophisticated child or animal. So, in the version of the case I am imagining, she only thinks of the determinate strong smell quality (or the trope that is the instantiation of this quality) that grabs her attention; she does not think about what might have caused it. Of course, if Mary were an adult, she might first think about the strong smell quality that grabs her attention, and then *wonder what caused it*. On this version of the case, Mary would be determinately thinking of more than one item: she would be thinking of the smell quality and what caused it. However, even in this version of the case, the *indeterminacy* claim (6) would be false. And this would be enough for my present challenge to reductive materialists; for the challenge to them is to explain how they might avoid the kind of indeterminacy asserted by (6). (Thanks here to Matt Duncan.)
10. Note well: even though I use the name "Determinacy", I do not mean that we must accept *absolute* determinacy. Because of the "problem of the many", *no materialist would accept this.* As we shall see, the claim I will be relying on should be put like this: there is not *radical indeterminacy* in consciousness or consciousness-based thought, in particular, the claims 6-8 in this section are false.
11. One might think that Determinacy and Easiness are at another maximum when it comes to our consciousness-based thought about *ordinary particulars*. But, in fact, I think that determinate reference to ordinary particulars (e. g. reference to *the statue* rather than to *the clay* that constitutes it) is much harder to achieve than determinate reference to basic qualities, often requiring sophisticated referential intentions (Pautz 2017, section 3).
12. Moreover, the descriptions-plus-causation view has a bizarre prediction: that if you changed the descriptive information you associate with mentalese demonstratives of smell qualities (if you changed your proto-theory of smell), this would switch the reference of such thoughts—say from external properties to internal properties. But this is just evidently wrong: instead, you'd still be thinking of the same qualities as before, and just have a different theory of them. When it comes to our consciousness-based thought about qualities, a strong form of *Semantic Stability* is correct.
13. Or, to put the problem in another way, let us say that Mary attends* to a property when she bears the intimate causal relation, causes*, to that property (so that Mary only attends* to neural properties of her own internal states such as P2). Let us say that Mary attends** to a property when she bears the more remote causal relation, causes**, to the occurrence of that property (so that she

only attends** to externally-instantiated properties like the chemical-type P1). Given these stipulations, it is undeniable that Mary both attends* to internal properties and attends** to external properties. Now, how could it be that one of these relations is thought-grounding while the other is not? For instance, it would be arbitrary to suggest that Mary only thinks about the chemical-type P1 that she attends** to, and that she is entirely unable to think about the neural property P2 that she attends* to (as externalists like Dretske and Tye who favor "transparency" would suggest). It would be equally arbitrary to say the opposite: that, in this example, Mary only thinks about the neural property P2 that she attends* to, and that she does not think about the chemical type P1 that she attends** to (as an internalist like Block would say).

14. Here is why it would be most natural for the reductive externalist to develop the indeterminacy option within a supervaluationist framework. The main alternative to supervaluationism is epistemicism (Williamson 1994). But if the reductive externalist is an epistemicist, then she is more likely to accept the "exquisite identities" response to be discussed at the end of this section (see also footnote 17).
15. See Chalmers 2004 uses a case where an Earthian moves from Earth to Twin Earth in order to argue that proponents of the reductive externalist model are committed to the possibility of radical phenomenal indeterminacy. However, proponents of the reductive externalist model have effectively answered Chalmers's case by appealing to teleological considerations (Lycan 2001). My elaborate Middle Earth case is designed to block this response; even if they bring in teleological considerations, their view implies radical phenomenal indeterminacy in this case.
16. Cian Dorr (a contemporary proponent of the reductive picture of the world that is the target of this essay) has suggested to me in discussion that accepting radical indeterminacy at the level of phenomenology may not be not absurd after all. He noted that, once we describe the physical truths corresponding to 6-8—once we describe their "precisifications"—we find nothing strange: just the unproblematic physical facts. But I don't see how this suggestion takes away the problem. After all, on Dorr's suggestion, 6-8 are still true, and they are still absurd. And it is still true that there is massive semantic indeterminacy in what our phenomenal terms refer to. Here is an analogy: as Kripke (1979) noticed in connection with his puzzle about belief, if we describe the facts in belief-neutral terms, there is no puzzle. But that doesn't take away the fact that there is a puzzle about belief.
17. I have compared the exquisite identities response to Williamson's (1994) view of being bald, being a heap, and so on. However, I should clarify that there is also an important difference between Williamson's view of these examples and the exquisite identities theory of consciousness as I am imagining it. In the case of being bald, being a heap, and so on, Williamson accepts (i) exquisite identities *and* (ii) ignorance due to semantic instability. The exquisite identities view as I am imagining it likewise accepts (i) exquisite identities in the case of consciousness but—and this is the important difference—it *denies* (ii) extreme ignorance due to semantic instability in this case. Thus it denies radical indeterminacy in conscious experience in the sense of Williamson's epistemic view. For instance, on the view I am imagining, when Mary has her hallucination on Middle Earth, she is conscious of *yellow* (not blue) because she bears the **F-187** relation to the yellow reflectance (not the blue reflectance) and the conscious-of relation is identical

with the **F-187** relation. But there is no semantic instability. Rather, her term "conscious of" *stably* refers to **F-187** (because it stably bears **R-187** to it) and her term "yellow" *stably* refers the yellow reflectance (because it stably bears **R-187** to it). So, on the exquisite identities view that I am imaging, it is *not* the case that she is bizarrely ignorant of the fact that she is having yellowish experience due to the semantic instability of her phenomenal vocabulary. Such a claim of ignorance due to semantic instability would be deeply implausible in this case: surely, here there is *no* radical semantic instability in her phenomenal vocabulary, and she is *not* ignorant of what her experience is like.

18. The exquisite identities response also requires a weird form of *luck*. To see this, suppose that Mary looks at the tomato, focusing on its color. Suppose that, for whatever reason, she thinks "I am thinking about the very same property I am conscious of"—maybe this sentence is in her "belief-box". This sentence is of course true. The proponent of the exquisite identities theory accommodates this by accepting the theory described in the text. On this theory—call it *Theory-1* - the reference relation is identical with a physical relation **R-187** such that "thinking about" and "conscious of" bear **R-187** to relations **F-187** and **G-187** (respectively), where **F-187** and **G-187** are *coordinate relations* in that Mary bears these relations to the *same functional reflectance*. But consider another, more twisted theory, *Theory-2*. On this theory, the reference relation is identical with a physical relation R-187* such that these expressions refer to *non-coordinate* relations **F-187** and G-189. On this theory, the conscious-of relation is identical a relation F-187 that Mary bears only to the *functional reflectance,* but the *thinking-of relation* is identical with a relation, G-189, that Mary bears only to the *photonic reflectance*. On this theory, Mary's thought "I am thinking of the same quality that I am conscious of" is false. (And the supervaluationist gambit of "penumbral connections" cannot avoid this, for, on the bizarre theory of (partial) reference that I am imagining, there is not even a single "precisification" of this sentence on which it comes out true: that is, the expressions "think of" and "conscious of" *partially refer* only to physical relations that Mary bears to different properties in this case.) The proponent of Theory-1 faces the question: why is Theory-1 true and not Theory-2? That is, why is one set of identities true and another false? On *a posteriori materialism*, the identities have no explanation. But then it is inexplicable—and also lucky-looking—that Theory-1 rather than Theory-2 is true.
19. It may be helpful to compare my argument in this section to an interesting suggestion of David Papineau's. Papineau (2002, chapter 7) suggests that reductive materialists must accept radical indeterminacy concerning some major issues about consciousness. He considers this to be an interesting consequence of reductive materialism rather than a *reductio ad absurdum* of it. However, aside from a couple of remarks (198), he does not give an argument for the indeterminacy view. Moreover, the interpretation of Papineau is not straightforward, for when he clarifies his view (199-200), he says things that are inconsistent with the indeterminacy view. My discussion has been very different. To begin with, whereas Papineau focuses on thought about *experiences*, I have focused on our consciousness, and conscious-based thought, about individual *qualities*, which I consider to be more basic than our thought about experiences. I think that a Papineau-style radical indeterminacy view, when applied to my cases, is *evidently false*. Whereas

Papineau favors reductive materialism, I have used Determinacy and Easiness in a series of novel cases to argue *against* it. And I have argued against responses appealing to descriptive fit, inferential dispositions, naturalness, and exquisite identities (responses Papineau does not consider).

20. In what follows, I will explain how the nonreductive *internalist* model of consciousness that I favor answers the Significance Argument. But it might be wondered whether a non-reductive ("primitivist") *externalist* view could also avoid the Significance Argument (for a recent example, see Allen 2017). In my view, such a view is vulnerable to a different form of the Significance Argument. In particular, such a view may accommodate the various ways in which the conscious-of relation is significant, but only at the cost of positing *extremely irregular and arbitrary grounding connections*. For instance, return to Mary on Black-and-White Earth viewing a black-looking object (§§2-3). On such a view, the outer object has a "primitive" black color while the inner object has a "primitive" red color. These primitive colors are grounded in, but distinct from, the reflectance properties of the outer object and the inner object. (And it is just a brute fact that they are grounded in, say, photonic reflectances rather than mere functional reflectances.) Non-reductive externalists must give the following account of this case: There is some narrowly-physical relation **R-187** that Mary bears *uniquely* to the primitive grey color of the outer object, and that she *doesn't* bear to the other chromatic states in this situation—for instance, the underlying *reflectance* of the outer object, the primitive red color of the inner object, and the underlying reflectance of the inner object. True, there are also various ever-so-slightly-different relations—R-186, R-188, R-189—that Mary bears to these other property-instantiations. But, it is just a brute and arbitrary grounding principle that Mary bears the primitive acquaintance relation to the states that she bears the ***R-187*** relation to, rather than the other states that she bears these ever-so-slightly different relations to. This would be required in order to explain why she is only acquainted with the *primitive red color* of the outer object, and *not* the *underlying reflectance* of the outer object, nor *any* states of the *inner* object. So a nonreductive *externalist* view requires an arbitrary "discontinuity" or "singularity" in nature, with an *exceedingly minute* (indeed hard-to-specify) physical difference grounding a *utterly monumental* mental difference. This view is intrinsically implausible. It is also subject to an Empirical Argument (Pautz 2016). In my view, the nonreductive *internalist* model is more plausible: for the history of psychophysics and neuroscience suggest that a nonreductive *internalist* model can provide much more regular, systematic and non-arbitrary grounding connections (see footnote 25 for more on this).
21. The nonreductive internalist model doesn't however *entail* the dissimilarity-grounding significance of consciousness. For even if the conscious-of relation is a relation R that has no physicalist definition, there may be *another* relation R* (e. g a relation that insentient robots bear to states) that has no physicalist definition, that is not a form of consciousness, but that is intrinsically very similar to R*. But the nonreductive internalist can say that it is not the case and cannot be the case (anymore than a non-color could be intrinsically like a color).
22. Let me make a couple of remarks in order to support my conditional claim that, if Fodor's atomistic language of thought approach to content-determination is correct, then the "secret scrambling case" I have described is possible.

(i) To explain "systematicity", Fodor needs a theory of the individuation of LOT expressions on which they can retain their identity despite being recombined (see Fodor 1987, appendix). But that is exactly what happens in my hypothetical case. (ii) Fodor advocates a functionalist way of spelling out the metaphors of "belief-box" and the "desire-box": one counts as believing* a sentence B and desiring* a sentence D iff B and D tend to lead to actions that satisfy the content of D given the truth of the content of B (Fodor 1987, 69; Fodor and Lepore 1992, 116). However, I intentionally constructed my hypothetical case so that, even though the sentences are "scrambled", *you still count as believing* and desiring* them under Fodor's functionalist account*.

23. This point is sometimes missed. For instance, Quilty-Dunn and Mandelbaum (2017, section 3.1) say that, if the holist and the atomist were to agree that our belief-related behavior is explained by a language of thought, then "the debate [between them] would start to look verbal". Against this, there is still a non-verbal difference between them: they provide different theories of *content-determination*; for instance, they differ in their predictions about what content-attributions we would make with respect to hypothetical cases like my "scrambling case" above (just as a superficialist theory of "water" and a natural-kind theory deliver different predictions about how we would apply "water" to Twin-Earth cases).
24. In order to state my "secret scrambling" argument against the atomistic, language of thought approach of Fodor, I have just granted for the sake of argument that there is some theory for determining the contents of sentences in the language of thought. But another deep problem with the approach is that no one has even sketched a plausible theory of this kind; therefore, the view doesn't even get off the ground. The approach still has adherents (e. g. Quilty-Dunn and Mandelbaum 2017) but for the most part they have just ignored this problem and moved on to other things. Fodor's himself (1987) suggested a *building-block model*: the contents of sentences in the language of thought are explained by (i) the contents of the sub-sentential expressions of the language of thought and (ii) its grammar. But the project of giving a single naturalistic theory of (i) totally dead-ended in the 90s, even though it mostly focused on simple terms like "cow" and largely ignored tougher cases, like names for *abstracta*, determiners, modifiers, connectives, and so on. Moreover, Fodor totally ignored the problem of giving a naturalistic theory of (ii). This is a real problem because, just as there can be deviant assignments of semantic values to sub-sentential expressions, there can be deviant grammars for the language of thought. (For instance, where ^ is a computational relation among terms in the language of thought, what makes it the case that 'a'^'F' in the language of thought means *that a is F* and not *that a is either F before 3000 AD or green afterwards*.) Given a building-block approach, it is not clear what naturalistic facts can be availed upon to rule out deviant grammars for the language of thought. By contrast, in the most basic cases, the Lewisian can rule out deviant interpretations on the ground that they gratuitously attribute *unnecessary irrationality* to the subject (what *other* feature do deviant interpretations have in common that could mark them out as incorrect?).
25. Still, you might think that my own non-reductive internalist model requires arbitrariness at a more basic level: in the connection between Mary's *brain state* and her *being conscious of irreducible smell quality Q*. But, understood one way, this is just an instance of explanatory gap that everyone faces in one form or

another. Moreover, in my view, there is a profound difference (as yet unknown) between the brain states that realize consciousness sand those that don't. So, in my view, the profound conscious/non-conscious divide is grounded in a big physical divide in nature. Moreover, I believe that research in psychophysics and neuroscience supports the view that there are general, *systematic* grounding connections between our intrinsic neural patterns and what perceptible properties (shapes, sensible colors, smell qualities, etc.) we are conscious of, even if we have not yet discovered them (Chalmers 2012, 279, 341; Kriegeskorte and Kievit 2013; Pautz 2018; but see Adams 1987 for interesting grounds for skepticism). If we only knew them ("cracked the neural code"), then we would find them to be very regular and non-arbitrary.

26. In fact, it falls out of this view that *in a creature without conscious experiences (e. g. the Simple System example used earlier) determinate intentionality is not possible* (see Pautz 2013). The view therefore supports a form of Russell's claim that *"all cognitive relations—belief and desire—presuppose acquaintance"* (1914, 1) and Chalmers's more recent claim that *"acquaintance is a condition on the possibility of thought and justification"* (Chalmers 2012, 467). The present model not only supports this claim but also explains why it should be true: conscious acquaintance is linked to reasons and (by the best systems theory) determinate thought depends on reasons.
27. This essay was presented at Leeds and in a seminar at MIT co-taught by Alex Byrne, Jack Spencer, and myself. My thanks to my discussants on those occasions. And thanks to Brian Cutter, Cian Dorr, Matt Duncan, Philip Goff, Uriah Kriegel, Geoff Lee, Heather Logue, Carla Merino-Rajme, and Robbie Williams. This publication was made possible through the support of a grant from the John Templeton Foundation. The opinions expressed in this essay are those of the author and do not necessarily reflect the views of the John Templeton Foundation.

References

Adams, R. 1987. Flavors, Colors, and God. In his *The Virtue of Faith and Other Essays in Philosophical Theology*. Oxford: Oxford University Press, 243–262.

Allen, K. 2017. *A Naïve Realist Theory of Colour*. Oxford: Oxford University Press.

Armstrong, D. 1968. *A Materialist Theory of Mind*. London: Routledge.

Berkeley, G. 1710. *Of the Principles of Human Knowledge*. http://www.earlymoderntexts.com/assets/pdfs/berkeley1710.pdf

Block, N. 1990. Inverted Earth. *Philosophical Perspectives* 4: 53–79.

———. 2010. Attention and Mental Paint. *Philosophical Issues* 20: 23–63.

Byrne, A. and D. Hilbert. 2003. Color Realism and Color Science. *Behavioral and Brain Sciences* 26: 3–21.

Chalmers, D. 2004. The Representational Character of Experience. In B. Leiter (ed.) *The Future for Philosophy*. Oxford: Oxford University Press.

———. 2006. Perception and the Fall from Eden. In T. Szabo Gendler and J. Hawthorne (eds.) *Perceptual Experience*. Oxford: Oxford University Press.

———. 2012. *Constructing the World*. Oxford: Oxford University Press.

Dorr, C. 2007. There Are No Abstract Objects. In J. Hawthorne, T. Sider, and D. Zimmerman (eds.) *Contemporary Debates in Metaphysics*. Oxford: Blackwell, 32–64.

———. 2016. To Be F Is To Be G. *Philosophical Perspectives* 30: 39–134.

Dretske, F. 1995. *Naturalizing the Mind*. Cambridge, MA: MIT Press.
Fine, K. 1994. Ontological Dependence. *Proceedings of the Aristotelian Society* 95: 269–290.
Fish, W. 2009. *Perception, Hallucination, and Illusion*. Oxford: Oxford University Press.
Fodor, J. 1987. *Psychosemantics*. Cambridge, MA: MIT Pres.
———. 1994. *The Elm and the Expert*. Cambridge, MA: MIT Press.
———. 2008. *LOT 2: The Language of Thought Revisited*.
Hawthorne, J. 2006. Postscript: Sider's Argument from Vagueness. In his *Metaphysical Essays*. Oxford: Oxford University Press, 104–109.
Hawthorne 2007. Craziness and Metasemantics. *Philosophical Review* 116: 427–440.
Horgan, T. 2014. Phenomenal Intentionality and Secondary Qualities: The Quixotic Case of Color. In B. Brogaard (ed.) *Does Perception Have Content*? Oxford: Oxford University Press.
Johnston, M. 2007. Objective Minds and the Objectivity of Mind. *Philosophy and Phenomenological Research* 75: 233–268
———. 2010. *Surviving Death*. Princeton, NJ: Princeton University Press.
———. 2011. On a Neglected Epistemic Virtue. *Philosophical Issues* 21: 165–218
Kriegel, U. 2013. The Phenomenal Intentionality Research Program. In U. Kriegel (ed.) *Phenomenal Intentionality*. Oxford: Oxford University Press, 1–26.
Kriegeskorte, N. and Kievit, R. A. 2013. Representational Geometry: Integrating Cognition, Computation, and the Brain. *Trends in Cognitive Sciences* 17: 401–412.
Kripke, S. 1979. A Puzzle about Belief. In A. Margalit (ed.) *Meaning and Use*. Reidel, 239–283.
Lee, G. 2018. Alien Subjectivity and the Importance of Consciousness. In A. Pautz and D. Stoljar (eds.) *Blockheads: Essays on Ned Block's Philosophy of Mind and Consciousness*. Cambridge, MA: MIT Press.
Levine, J. 2001. *Purple Haze*. Oxford: Oxford University Press.
———. 2006. Phenomenal Concepts and the Materialist Constraint. In T. Alter and S. Walter (eds.) *Phenomenal Concepts and Phenomenal Knowledge: New Essays on Consciousness and Physicalism*. Oxford: Oxford University Press, 145–166.
Lewis, D. 1994. Reduction of Mind. In S. Guttenplan (ed.) *A Companion to the Philosophy of Mind*. Oxford: Blackwell, 412–431.
Lycan, W. 2001. The Case for Phenomenal Externalism. *Philosophical Perspectives* 15: 17–35.
McGinn C. 1996. Another Look at Color. *Journal of Philosophy* 93: 537–553.
Neander, K. 2017. *A Mark of the Mental: In Defense of Informational Teleosemantics*. Cambridge, MA: MIT Press.
Papineau, D. 2003. *Thinking about Consciousness*. Oxford: Oxford University Press.
———. 2016. Against Representationalism (about Conscious Sensory Experience). *International Journal of Philosophical Studies* 24: 324–347
Parfit, D. 2011. *On What Matters, Volume One*. Oxford: Oxford University Press.
Pautz, A. 2006. Can the Physicalist Explain Colour Structure in terms of Colour Experience? *Australasian Journal of Philosophy* 84: 535–564.
———. 2010. A Simple View of Consciousness. In R. Koons and G. Bealer (eds.) *The Waning of Materialism: New Essays*. Oxford: Oxford University Press, 25–66.
———. 2013. Does Phenomenology Ground Mental Content? In U. Kriegel (ed.) *Phenomenal Intentionality*. Oxford: Oxford University Press, 194–234.
———. 2014. The Real Trouble with Armchair Arguments against Phenomenal Externalism. In M. Sprevak and J. Kallestrup (eds.) *New Waves in Philosophy of Mind*. Basingstoke: Palgrave Macmillan, 153–181.
———. 2016. Experiences are Representations: An Empirical Argument. In B. Nanay (ed.) *Current Controversies in Philosophy of Perception*. New York: Routledge, 23–42.
———. 2017. The Perceptual Representation of Objects and Natural Kinds. *Philosophy and Phenomenological Research* 95: 470–477.

———. 2018. How Does Color Experience Represent the World? In D. Brown and F. MacPherson (eds.) *Routledge Handbook of the Philosophy of Color*. New York: Routledge.
———. Forthcoming. *Perception: How Mind Connects to World*. New York: Routledge.
Peacocke, C. 2008. Sensational Properties: Theses to Accept and Theses to Reject. *Revue Internationale de Philosophie* 62: 7–24.
Prinz, J. 2008. Has Mentalese Earned its Keep? *Mind* 120: 485–501.
Pryor, J. 2000. The Skeptic and the Dogmatist. *Noûs* 34. 517–549
Putnam, H. 1981. *Reason, Truth and History*. Cambridge: Cambridge University Press.
Russell, B. 1912. *The Problems of Philosophy*. London: Williams and Norgate.
———. 1914. On the Nature of Acquaintance I: Preliminary Description of Experience. *The Monist* 21: 1–16.
Scanlon, T. 2014. *Being Realistic about Reasons*. Oxford: Oxford University Press.
Schaffer, J. 2017. The Ground Between the Gaps. *Philosopher's Imprint* 17: 1–26.
Shoemaker, S. 1994. The Phenomenal Character of Experience. *Philosophy and Phenomenological Research* 54: 291–314.
Sider, T. 2011. *Writing the Book of the World*. Oxford: Oxford University Press.
Tye, M. 2000. *Color, Consciousness and Content*. Cambridge, MA: MIT Press.
Williams, R. MS. *The Nature of Representation*.
Williamson, T. 1994. *Vagueness*. New York: Routledge.
Williamson, T. 2002. *Knowledge and Its Limits*. Oxford: Oxford University Press.
Yi, B. 2017. Nominalism and Comparative Similarity. *Erkenntnis*. https://doi.org/10.1007/s10670-017-9914-2.