How Can Brains in Vats Experience a Spatial World?
A Puzzle for Internalists

Adam Pautz

A "tricked" brain in a vat with exactly the brain activation that I now have would have the same experiences that I am having now despite lacking a body. — Ned Block (2012)

Internalist theories of experience hold that the character of an individual's experiences is entirely fixed by his intrinsic properties, so that any intrinsic duplicate (even a "brain in a vat") must have exactly the same experiences. Proponents include Block, Chalmers, Horgan, Kriegel, McLaughlin, Mendelovici, Papineau, and many others. Neuroscientists also typically favor internalism, for instance Koch, Zeki, and Lamme. One version is *type-type identity theory:* every distinct experiential property is *necessarily* identical with a distinct intrinsic neuro-computational property of the brain. On the other side are various externalist theories of experience. Examples include tracking representationalism (Dretske, Tye), naïve realism (Campbell, Martin), and active externalism (Noë, O'Regan).

I think that there is a strong empirical case to be made for internalism about experience and against externalist rivals. However, here my primary aim is not to argue for internalism but to develop an overlooked puzzle for it, a puzzle about the experience of space. My focus throughout will be on the type-type identity theory ("identity theory" for short), which is defended by Block (2009), McLaughlin (2003), Papineau (2014), and others. I focus on the identity theory because it is simple and because there has been some renewed interest in it. I also will suggest an answer to the puzzle, an answer having important consequences for our understanding of mental representation and of the place of the mind in the physical world.

Let me give you a sense of the puzzle. Suppose you have an experience that is in fact caused by a round tomato on an all-white background. I think that this experience is *necessarily* an experience as of a round item of some sort. In the terminology of Chalmers (2004), it necessarily involves "phenomenally representing" roundness. Now the puzzle is that identity theorists must apparently say that this representational relation is irreducible. To see this, consider a lifelong, lone brain in a vat ("BIV") that formed by chance and that happens to undergo the same brain state as you. On the identity theory, BIV has the very same tomato-like experience as you, and so also "phenomenally represents" roundness, despite its degenerate situation. In fact, given the identity theory, BIV might have *all* your rich visual experiences, representing a range of spatial properties and relations, such as *moving to the left*, *having so-and-so orientation*, and *being above*. In that sense, BIV could ostensibly "experience a rich spatial world". It is "tricked" because there aren't things before it with these properties. Now, obviously, these clusters of

properties are also not instantiated inside the brain by the neural states responsible for experience. The puzzle now is roughly this. How can experience be internally-determined and yet necessarily externally-directed? How can BIV represent properties that are instantiated "outside the head" (if they are instantiated at all), just on the basis of states "inside the head"? Evidently, BIV bears no interesting *physical-functional relations* to such spatial properties, including the relations invoked in our most sophisticated naturalistic theories of representation (asymmetric dependence, indication, and so on). So if, like you, BIV phenomenally represents such properties, then it appears that the phenomenal representation relation must be an *irreducible* relation. But this appears mysterious.[1]

In response to the puzzle, I think internalists should simply concede the irreducibility of the phenomenal representation relation. I think internalists who also subscribe to the doctrine of physicalism should accept what I will call the "internal grounding view" of phenomenal representation. On this view, even though the phenomenal representation relation is irreducible, our bearing this relation to certain shapes and other properties is always "grounded in" (in the sense of Fine, Rosen, and others) our being in certain brain states. I think that this is a defensible view. In fact, it meshes nicely with a general view of representation relations that Paul Horwich and Stephen Schiffer have proposed on independent grounds.

My plan is as follows. In §1, I will explain why we should take internalism seriously. In §§2-5, I argue that internalism leads to anti-reductionism about phenomenal representation. Finally, in §6, I describe the "internal grounding view" of phenomenal representation.

1. Why Take Internalism about Experience Seriously?

You will not be very interested in my puzzle for internalism if you are convinced that internalism about experience has no chance of being true. So I will first discuss the internalism-externalism debate and explain why I think that internalism deserves to be taken seriously. I will first explain what I consider to be the strongest argument for the rival position of externalism about experience. Then I will show that there is a stronger *empirical* case to be made for internalism, and explain why the argument for externalism is not decisive.

The most discussed argument for externalism about experience and against internalism depends on the controversial "transparency thesis": that whenever you know what your experience is like (even in hallucination), it is by attending to the objects and properties presented in the experience. I think that non-veridical experiences create problems for this thesis (Pautz 2007; see Tye 2014 for a response). I would like to sketch what I consider to be a stronger argument for externalism about experience, one with a more modest starting point. The argument is that externalists can provide the best explanation of the experience of space. I call it **the spatial argument.** It will help set the stage for

the rest of the essay. In effect, my puzzle for internalists will be about how they might answer this argument.

First, some terminology. Suppose again you see a tomato on a white background. Now pretend that later neuroscientists artificially reproduce exactly the same neural state you now have, so that you have a hallucination that perfectly matches your original experience. Intuitively, you have the same salient mental property in each case. Let '*R*' rigidly designate this property. Philosophers would call it *the property of having an experience with specific phenomenal character K*. This is an example of what I will call an *experience property*. Now internalism about experience implies that an individual's having *R* is entirely fixed by the intrinsic properties of his brain, so that even a life-long brain-in-a-vat could have *R*. Could this be right?

The spatial argument against this internalist view starts with a simple observation. Roughly, you couldn't have *R*, and fail to have an experience *as of* a *round* item of some sort. *R* is *necessarily* directed at a *round* item, a kind of item that needn't exist in your brain when you have *R*. Call this *External Directedness*. Many philosophers have argued for the point. I will have much more to say about it later because it figures in my puzzle for internalists about the brain-in-the-vat. For now, let's just assume it. I will also assume that having an experience as of a round item is a *representational* property in some good sense: you can have an experience as of a round item, even if there is no existing round item there. So I will put "External Directedness" by saying that having *R necessarily* involves "phenomenally representing" roundness.[2]

The next step in the spatial argument against internalism concerns the following question: assuming the physicalist view that the physical facts fix all the facts, what is the physical basis of your phenomenally representing roundness? For instance, how can you have an experience of a *round* thing even in hallucination, when there is no physical (or mental) round thing around?

Internalists about *R* are committed to the following answer: that your *intrinsic* physical state, considered in isolation from your environment, necessitates your phenomenally representing roundness, a property that is not instantiated by that physical state. For they claim that some intrinsic physical state of you necessitates your having tomato-like experience *R*. And, given "External Directedness", this in turn necessitates your phenomenally representing roundness.

However, as I indicated in the introduction, this internalist view of the built-in spatial intentionality of visual experience faces an apparent puzzle, which will be the main subject of this paper. In short, the puzzle is that internalism apparently requires "irreducible representational relations". In addition, it apparently requires somewhat arbitrary modal connections. For why should simply undergoing a mere neural pattern necessarily result in phenomenally representing roundness and not some other shape, even in possible cases (like the BIV case) in which the brain state is not causally connected to round objects? Considering the intrinsic character of the neural pattern alone, this looks arbitrary.

The best argument for externalism about the experience property *R*, I think, is that it follows from an apparently better externalist account of the spatial intentionality inherent in *R*, one avoiding these puzzles. Since Galileo there has been a question about whether *colors-as-we-see-them* are really are "out there". But, when it comes to the *spatial properties and relations* we phenomenally represent in experience, nearly everyone accepts *realism*: these really are occasionally instantiated out there in the world. (We will soon see that David Chalmers is an exception.) So an externalist account of how we represent *these* properties is natural: we phenomenally represent these spatial properties by having a visual system that interacts with instances of these very properties in the external world.

One simple version is the "tracking" account of Dretske (1995) and Tye (1995). Very roughly, on this account, the physical ground of your representing *roundness* (and not some other shape) isn't just your undergoing a certain neural pattern alone (as on internalism), but your undergoing a neural pattern that *in normal conditions* is caused by the presence of a round object (and that in turn causes behavior appropriate to such an object). For short, the ground is your *tracking* roundness. Call this a *tracking property*.

This externalist view of spatial experience apparently avoids the kind of puzzle facing the internalist view. It runs no risk of requiring irreducible representational relations: on this view the phenomenal representation relation that links the mind to external spatial properties is just an non-mysterious "tracking" relation. The externalist view also minimizes arbitrariness. On this view, the physical basis of phenomenally representing *roundness* is not merely undergoing a neural pattern, but rather undergoing a neural pattern that tracks the instantiation of that very property, *roundness*. By extending the physical substrate beyond the brain, we obtain an explanation of why you have an experience of one shape rather than another.

Now we can complete the spatial argument for externalism about *R*. The starting point was the modest claim that having tomato-like experience *R* necessarily involves phenomenally representing roundness. The next step was that phenomenally representing roundness depends on more than your intrinsic state: it depends on tracking round things. It follows that having experience property *R* depends on more than your intrinsic state, just as externalists maintain. For instance, contrary to internalism, this approach entails that an isolated, life-long brain-in-vat simply could not have *R*, even if it is an intrinsic duplicate of your brain, because it doesn't suitably track round things (more on this in §5).

The spatial argument is general. The character of visual experience is inseparable from representing spatial properties, like *moving to the left*, *having so-and-so orientation*, and *being above*. And in general (so the argument goes) we must explain the representation of such spatial properties in terms of links with the instantiation of these very properties in the external world, rather than in terms of mere internal neural patterns considered in isolation from that world. So visual experience properties are not intrinsic.

So far I have focused on the experience of spatial properties, examples of traditional "primary qualities". But what about the experience of traditional "secondary qualities"?

For instance, intuitively, having tomato-like experience *R necessarily* involves having an experience as of a *certain distinctive quality* along with roundness. I will call it *sensible redness*. (I will call it "sensible redness", rather than just "redness", to remain neutral between the view is that it is identical with "the color red" and the rival view that we should rather think of it as a "color-appearance property" that corresponds with but is not identical with the color red.) As Berkeley (1713, 157-8) noted, in general, sensible colors and shapes "appear as being in the same place" (though he himself located them both "in the mind"). Other examples of *sensible properties* are qualities of sound, bodily pain and pleasure, heat and cold, taste, and so on. We undeniably seem to experience sensible properties as co-instantiated with spatial properties, such as location and shape. Indeed, this is even so in *hallucination*. How is that possible?

I think that those who accept an externalist account of the experience of "primary qualities" like shapes might use a *generalization argument* to support a similar externalist account of the experience of "secondary qualities". So the spatial argument indirectly supports externalism generally. This generalization move is supported by considerations of uniformity. On the resulting view, sensible redness, like roundness, is an objective, mind-independent feature of tomatoes and other objects that is tracked by the visual system. In one natural version, it is a *reflectance property* of objects. Further, we *phenomenally represent* sensible redness (have experiences of it) *in the same way* we represent roundness, namely, by having a brain state that tracks it under biologically normal conditions.[3] This theory explains how, even in hallucination, we can ostensibly experience sensible colors in various locations and as conjoined with other spatial properties. Unless we are willing to accept sense data in a private mental space, or the mysterious "visual field regions" of Peacocke (2008), how else might we explain this? True, there are traditional arguments against the view that sensible colors are objective properties of external objects, concerning perceptual variation, spectrum inversion, and so on. But externalists (Dretske, Tye, others) have tried to answer those arguments.

The spatial argument and the generalization argument together suggest that experience properties like *R* are necessarily connected with tracking properties. If this is right, it is natural to go further and claim that they are just *identical with* tracking properties. So, for instance, having the tomato-like visual experience *R just is* having some or other "suitable" internal state that would, under biologically normal conditions, track the co-instantiation of roundness together with a certain sensible color (identified with a reflectance property of surfaces) in a certain viewer-relative place *p*. In general, different experiences (auditory experience, bodily sensations, taste experiences) involve tracking, and thereby representing, different clusters of external properties.

Thus, we have arrived at a general externalist view of experience. In fact, we have arrived at the "tracking representationalism" of Dretske (1995) and Tye (1995). It is "representationalist" in that it holds that phenomenal differences among individuals' experiences reside in differences in what perceptible properties those individuals phenomenally represent. True, there are difficult cases for tracking representationalism (blur, affective differences, attentional differences, itches and tickles), but proponents hope that all, or nearly all, aspects of experience can be handled in this way. For reasons, I won't go into here, I consider it to be the best form of externalism.

Tracking representationalism is radically externalist. To see this, it may be helpful to compare it with an internalist theory of experience. Consider, for instance, the brain-based *identity theory*. To a first approximation, on the identity theory, each experience property is *necessarily* identical with a unique internal neuro-computational property (e. g. a unique *spatio-temporal pattern* of neuronal firing), rather than with a tracking property involving the external world. (I will provide a more complete formulation in §2.) Differences among experiences are constituted by differences in these patterns (see Prinz 2012, 126-133 for a very important discussion). If you want to know the complete essence of having a reddish experience, you would have to look at the corresponding spatio-temporal neural pattern. True, a reflectance property of external objects normally *causes* this neural pattern; but the experience has no *essential* connection to that reflectance property. Likewise, the essence of the smell of peppermint is another (and presumably radically different) internal spatio-temporal neural pattern. By contrast, on the tracking representationalism of Dretske and Tye, experience properties are identical with tracking properties of the form: having some appropriate internal state or other that tracks cluster of spatial and other external properties P, Q, R, . . . Differences between experiences (within a species or across species) reside wholly in differences in the *external physical properties* normally tracked, and thereby represented, by those experiences. So if you want to know the essence of a reddish experience of a tomato, look at the reflectance property that it tracks (constituting the "sensible redness" you perceive, on this view) – not the neural "content vehicle". And if you want to know the essence of a smell experience, look at the chemical property that it tracks. As Tye says:

> Peer as long as you like at the detailed functioning of the brain . . . that is not where phenomenal character is to be found (Tye 1995, 162-3) . . . phenomenal character is in the world (Tye 2009, 119).

So far, I have sketched a seductive "spatial argument" for externalism about experience, which led us to the "tracking representationalism" of Dretske and Tye. Let us now turn to the argument for the other side, internalism about experience.

There are many arguments against externalist views like tracking representationalism and for internalism about experience. Many of them are

armchair arguments. In my view, such arguments are unconvincing (Pautz 2013b). For instance, the *inverted spectrum argument* says that it is conceivable, and therefore possible, that two people should have "inverted color experiences" but normally track the same reflectance properties in the external world (Block and Fodor 1972, Shoemaker 1994). But the move from conceivability to possibility is questionable (Tye 2000, 109-110). Block's (1994) well-known *inverted earth argument* – which is the flipside of the inverted spectrum argument – can be shown to depend crucially on the assumption of internalism about experience (Levine 2001, 113). In his formulation of the argument, Block asserts without argument that "we can assume the supervenience of qualia on the brain" (1994, 518). But this is exactly what is at issue. We need an *argument* for this assumption.

In my view, the best argument against externalism and for internalism is an empirical argument, which I call **internal-dependence argument**. I recommend that internalists add this argument to their arsenal. To illustrate the "internal-dependence" argument, I will focus on the identity theory, but I think a similar argument could be developed for other internalist theories. Elsewhere I have explained the argument in detail, and have distinguished it from more standard arguments.[4] Here I can only briefly describe some of the recent empirical work it is based on. Much of this work uses the emerging technique of "multivariate pattern analysis". While the spatial argument for externalism starts with our experience of spatial properties (traditional "primary qualities"), the internal-dependence for internalism starts with the experience of sensible properties ("secondary qualities").

For instance, suppose you successively experience blue, purple and green. Then your first color experience is more like your second than your third. Now, on tracking representationalism, all facts about the character of our experiences derive from fact about the physical properties tracked and thereby represented by our experiences. But it is simply not the case that the "blue" reflectance-type is more like the "purple" reflectance-type than the "green" reflectance-type, where these are the reflectance-types (colors) tracked and thereby represented by your consecutive experiences. In fact, if anything, the blue reflectance-type is more like the *green* one than the purple one (Byrne and Hilbert 2003, Pautz 2006b).

At the same time, important recent research in neuroscience shows that phenomenal similarities and differences map nicely onto *neural* similarities and differences. Brouwer and Heeger sum up this research as follows:

> The visual system encodes color by means of a distributed [neural] representation [in area V4]. . . similar colors evoke similar patterns of [neural] activity, and neural representations of color [in V4] can be characterized by low-dimensional "neural color spaces" in which the positions of [experienced] colors capture similarities between corresponding patterns of activity (2013, 15454)

Indeed, when ordered according to similarity, these neural representations formed the circle, akin to the familiar hue circle. So your distributed *internal V4 neural representation* of the blue object resembles your *V4 neural representation* of the purple object more than your *V4 neural representation* of the green object. Other recent important work on the neural basis of color experience includes Danilova and Mollon (2012) and Schmidt, Neitz and Neitz (2014). In light of this recent research, a broadly internalist account of color experience is evidently more reasonable than an externalist theory such as tracking representationalism. Elsewhere I have developing this point more precisely by appealing to hypothetical "coincidental variation" cases, which differ from both inverted spectrum and inverted earth cases.[5]

The point extends to non-visual modalities. For instance, phenomenal resemblances among smells radically fail to line up with objective resemblances among the corresponding chemical properties tracked, and thereby represented, by our olfactory system (Pautz 2013a). By contrast, recent research conducted by Youngentob *et al*. (2006), Howard *et al*. (2009) and others has shown that neural patterns in the olfactory system fall into a neural similarity space nicely matching phenomenal similarity space (similar to Brouwer and Heeger (2009) for color vision).[6] Similar results have been found for the experience of taste (Crouzet *et al*. 2015). Finally, psychophysics has shown that there is huge mismatch between the qualitative structure auditory experiences (ratio relations among perceived loudness levels, categorical changes in phoneme perception, etc.) and the structure of the physical properties our auditory experiences track and thereby represent. To find the explanation, we must look inside the head (Chang 2010, Pautz 2013a). This is exactly the opposite of what tracking representationalists like Dretske and Tye suggest.

In short, in many cases, externalists like Dretske and Tye have it backwards. Peer at the external physical properties tracked and represented by our experiences however much you like. That is not where you will find an explanation of the phenomenal structure of our experiences. So the explanation *must* reside in the "detailed functioning of the brain". And the more we learn about the brain, the more this hypothesis is corroborated. My point here is not just that in the actual world experiential differences are always accompanied by neural differences. Contrary to what some have suggested (e. g. Prinz 2012, 19), that fact alone provides no evidence against externalism about experience (Pautz 2013b, 168). My point relies on two additional empirical facts. First, in many cases, psychophysics shows that *structural relations* among experiences (similarity and difference, equal intervals, proportion) are not matched by the structural relations among the (highly unnatural) external physical properties that those experiences track. Second, at the same time, recent neuroscience has shown (often using the new technique of multivariate pattern analysis) that they are much better matched by structural relations among their neural correlates. These twin facts suggest that our experiences of sensible properties (sensible colors, smells, audible qualities) depend on neural processing, in a

way that can be demonstrated to be in conflict with tracking representationalism (Pautz 2013a).

Once we accept an internalist account of the experience of "sensible properties", the simplest and most uniform view is that the experience of *all* perceptible properties (including spatio-temporal ones) generally is internally determined. The brain-based identity theory is a view of this kind. Like the argument for externalism we considered earlier, this is a "generalization argument" – only it proceeds in the opposite direction.

So there is a strong empirical case for internalism about experience. But then what about the more *a priori* "spatial argument" against internalism and for externalism that we started with? That argument suggests that *spatial* phenomenology, at least, *does* depend constitutively on links to the environment, contrary to a general internalist theory like the identity theory.

In my view, the spatial argument is far from decisive, for a couple of reasons. To begin with, the spatial argument for externalism about experience depends on realism about the spatial properties that we phenomenally represent. The thought is that, since these properties are instantiated out in the external world, it's natural to explain how we phenomenally represent them in terms of our having a visual system that interacts with instances of these very properties in the external world.

But some have recently suggested a kind of *irrealism* about experienced spatial properties. This view may seem far-fetched but several philosophers have advocated it on the basis of contemporary physics (for discussion see Ney 2013, pp. 177-181). One example is David Chalmers. In fact, Chalmers advocates a kind of *generalized* irrealist view. To illustrate, suppose you view a tomato. Then it seems to you that a certain quality, sensible redness, is coinstantiated with roundness. Since the 17th century scientific revolution, many have suggested that such sensible colors are not really instantiated by external objects. These "irrealists" about sensible color would admit the tomato has a reflectance. But, in their view, this reflectance is *nothing like* the sensible color presented in experience. Chalmers accepts this irrealist view of sensible colors on *a priori* grounds (2006, 82). I have also argued for the same view on entirely *empirical* grounds (Pautz 2006b, 2013a). What Chalmers does is to take irrealism one step further. The "real tomato" does not even instantiate the property *being round* that you phenomenally represent as coinstantiated with sensible redness, that is, the property *having edges roughly equidistant from a common point*. Chalmers (2006, 107) calls this "perfect roundness"; if you like, it is "roundness-as-we-see-it". On his view, the tomato only has "imperfect roundness", which is some arcane quantum mechanical property that is nothing like the familiar roundness you are directly acquainted with (2012, 296-297). In general, the basic spatial and temporal relations given in experience are just not out there. Chalmers thinks that this irrealist view is supported by relativity and certain interpretations of quantum mechanics ("wavefunction fundamentalism"). The result is a kind of *uniform* Kantian picture: the real world is "noumenal".[7] As Chalmers puts it:

> In spatial experience, I think we are presented with certain primitive spatial properties. . . But I think that there is little reason to think that they are instantiated in our world. Certainly, it is not easy to see how [they could be instantiated] in a relativistic world, or in a string-theoretic world . . . (2012, 333)

Now, if Chalmers thinks that "perfect" roundness and redness not instantiated in the external world, where *does* he think they are instantiated? Does he perhaps think that they are mental *qualia* instantiated in the mind or the brain when you experience the tomato? No – of course, when you experience a tomato, perfect roundness (that is, the property *having edges roughly equidistant from a common point*) need not be instantiated inside your brain! Instead, Chalmers adopts a representationalist view. In having the tomato-like experience, you *phenomenally represent* perfect roundness and redness. So, they *appear* to be instantiated out there. Chalmers, then, accepts "External Directedness" where 'roundness' is understood to mean perfect roundness. But he thinks that, as it happens, *nothing* instantiates these properties - not even mental items in the head. It is only some other possible worlds ("Edenic worlds") that they are instantiated by external objects, according to Chalmers.

Of course, if such an irrealist view of experienced spatial properties is correct, then the spatial argument for externalism fails at the first step. We cannot explain how we phenomenally represent roundness-as-we-see-it (much less redness-as-we-see-it) in terms of being in an internal state that typically tracks the instantiation of that very property under biologically normal conditions, for the simple reason that it could never be instantiated in this world! The irrealist view goes with internalism about the spatial intentionality built into normal visual experiences.[8]

My point here is not that this irrealist view is definitely right. My point is just that it *may* be correct. So, the spatial argument for externalism about visual experience is not so cut and dried. (However, in what follows, for the sake of discussion, I will often write as if realism about experienced spatial properties correct.) In fact, although I cannot go into this here, there are other empirical reasons to doubt externalist theories of the experience of space. So it is worthwhile to consider the question of whether we can develop an alternative internalist theory.[9]

There is another, more basic reason why the spatial argument isn't a decisive argument against internalism about experience. That argument is just an inference to the best explanation. The argument is that an externalist theory, such as the "tracking theory", provides the best explanation of the spatial intentionality built into many experiences. So, one way to block the argument would be to show that there is a viable rival *internalist* explanation of the experience of spatial features. Such a theory would explain how BIV, just on the basis of its internal neural states, can "phenomenally represent" roundness, even though it fails to track round things.

This rest of this essay is devoted to the question of what such an internalist theory might look like. I begin by developing the puzzle about spatial representation that is the main focus of this essay. In particular, I will argue (§§2-5) that internalists must concede that the "phenomenal representation relation" is an *irreducible* relation. This may look like a *reductio* of internalism. However, I will briefly suggest (§6) that this is not the right way to look at it. Internalists can live with this result by accepting a non-reductive "grounding" account of the phenomenal representation of space.

2. From Envatted Brains to Irreducible Intentionality: The BIV Argument Sketched

The **BIV argument** is designed to establish a conditional claim: if internalism about experience is right, then a non-reductive view of phenomenal representation follows.

First I will only describe the setup and list the steps of the argument. Afterwards (§3-5) I will explain and defend those steps in turn. I will continue to focus on the tomato-like experience property *R*. However, just about any visual experience could illustrate the argument. I will also continue to focus on the brain-based identity theory defended by Block, McLaughlin, Papineau, and others.

I begin by saying more about the identity theory. I have said that the identity theory holds that *R* is necessarily identical with an intrinsic neuro-computational property of the brain ("intrinsic" in the sense that any duplicate brain must share the property). But what kind of property? Suppose that whether or not a subject has tomato-like experience *R* co-varies with their having some fairly local neural property *V* in the visual cortex. No identity theorist holds that *V* all by itself – say, isolated in a bottle - would constitute *R*. Rather, identity theorists hold that *R* is necessarily identical with some more *global* neural state, incorporating certain further conditions. Let *N* be the more global neural state that, by contrast to *V* alone, *is* necessary and sufficient for *R*, according to identity theorists.

What might *N* involve? Ned Block (2005, Box 1), one of the most prominent defenders of a brain-based approach to experience, suggests that we can approach the issue by asking: what could be removed from your brain, and what must be kept, while you still continue to have *R*? Exactly which brain areas are required is unknown at present. But Block tentatively suggests that *N* probably has to incorporate a "recurrent feedback loop". Block (2007a, 482) also notes that "there is some evidence that there is a single neural background of all experience involving connections between the cortex and the upper brain stem including the thalamus".

But there are other things we could remove. As Block (2012, sect. 10) writes, citing recent research, "the basic phenomenology of vision can survive vast destruction in motor areas and early sensory areas on both sides of the brain". You could have *R* because of direct stimulation of your visual cortex, without involving the eyes. In addition, we could remove connections to the

body: for, contrary to behaviorists, a quadriplegic, or someone with complete locked-in syndrome, could have *R*. Therefore *N*, the minimal sufficient condition for *R*, doesn't involve the eyes or to the motor output systems (e. g. motor neurons in the spinal cord and brainstem).[10]

Now for the setup of the argument. Suppose that, in another possible world, a brain-in-a-vat ("BIV") just pops into existence "out of the blue" (like a "Botlzman brain"), and then starts to undergo neural state *N* for five minutes purely by chance, without any external cause (Block 2012, sect. 12). Suppose, further, that BIV *only* has the neural machinery required to have *N* and hence *R*. So BIV doesn't have, for instance, receptor systems (eyes, ears, etc.) or a motor output system. So, although I called it a "brain-in-a-vat", BIV is not exactly like a complete brain. Indeed, it has no evolutionary history and belongs to no species (in fact, we could assume that BIV occupies an otherwise empty universe, so that it is a "brain in the void"). So it doesn't really count as a *human brain*. It is merely an intrinsic *duplicate* of a significant part of your actual brain as you view a tomato.

Still, the identity theory implies that BIV has the tomato-like experience property *R*. Likewise for any internalist theory. *This cannot be disputed*. For I have stipulated that BIV has *N*, where "*N*" is just shorthand for whatever neural property is, on that theory, the minimal neural basis of *R*.

This is not a trivial implication. As we shall see, Dretske and Tye deny it, on the basis of their "tracking representationalism".[11] In what follows, I will assume it for the sake of argument.

My BIV argument will now proceed as follows:

> **Assumption:** An internalist theory of experience is right, so that, by having neural property *N*, BIV has experience property *R*. (Assume for conditional proof.)
> **External Directedness about *R*:** necessarily, if any individual has *R*, then that individual has an experience of a *round* item of some sort; that individual has an experience that "matches the world" only if some item is present that is *round*. So, in having *R*, BIV has such an experience.
> **Phenomenal Representation:** If BIV has an experience that "matches the world" only if some is present that is round, then it stands in the following dyadic relation to the property *being round*: it has an experience that "matches the world" only if some item has this property. Call this the "phenomenal representation relation".
> **Irreducibility:** If BIV bears the dyadic phenomenal representation to the property of being round, then this relation is *irreducible*. For BIV bears no suitable dyadic *physical-functional* relation to the property of being round, with which the phenomenal representation relation might be identified (e. g. the kind "tracking relation" invoked by Dretske and Tye).

> **Conclusion:** *If* an internalist theory of experience is right, so that BIV has *R*, *then* the dyadic phenomenal representation relation is *irreducible*.

The conclusion of the BIV argument may be represented as follows:

[Insert Pautz Figure "23-01" about here]

> Figure 1: Internalism about experience implies that the phenomenal representation is irreducible (represented here by the arrow sticking out from the brain).

If the internalist accepts the BIV argument, then he faces the question of how to understand the relationship between the irreducible phenomenal representation relation and the physical world. This constitutes the **puzzle of the phenomenal representation relation** for internalists.

This puzzle is bound up with another puzzle for internalists, which we might call the **puzzle of sensible properties**. As we noted before, in having tomato-like experience *R*, BIV has an experience *as of* a certain distinctive *reddish quality* filling a round area in space; that is, BIV phenomenally represents *sensible redness* as well as roundness. We saw that externalists like Tye and Dretske have an attractively straightforward (though arguably false) view of sensible redness: sensible redness is an external reflectance property that *really is* co-instantiated with roundness. But *internalists* about experience cannot take this objectivist view. For they hold that internal neural state *N* suffices for the experience of sensible redness, even in BIV cases where it doesn't track any reflectance property in the external world. The empirical research on color vision I cited earlier supports this brain-based view. But then, according to internalists, what in the world might sensible redness be, if not an objective property of external things in physical space?

Ned Block would apparently say that, when BIV (or, for that matter, an actual person) has the tomato-like experience, sensible redness (or what he calls the "red *quale*") is somehow instantiated "in the mind" (2007b, 74). On one version of this view, sensible redness is instantiated by a literally round "visual field region". But there exists no reddish and literally round "visual field region" anywhere within BIV (Peacocke 2008, 14). On Block's own view, sensible redness (the "red *quale*") is instantiated by the BIV's *experience itself*, which, on his identity theory, is just a neural state *N*. Now, it is an evident fact that, if the BIV (or an actual person) has the tomato-like-experience, then it at least *seems* to the subject that sensible redness (the "red *quale*") fills a round region. (If Block's term "red quale" does not refer to the distinctive quality that seems to fill a round region when one has the tomato-like experience, then I do not know what it could refer to.) So Block's view requires that a quality that does not in fact fill a round region (because it is in fact a quality of a non-round neural state) somehow appears to fill a round region (see Pautz 2003a, 286-290). Sydney Shoemaker (1994) is an internalist

who suggests a different view of sensible redness. He suggests that it is a response-dependent "appearance property" (roughly, a disposition to produce a certain neural state in a population) that is really instantiated by external objects (like round tomatoes) in our environment but not in BIV's environment. Other internalists have argued that sensible redness is a primitive property that is not instantiated *anywhere*, on both *a priori* grounds (Chalmers 2006, Horgan 2014) and empirical grounds (Pautz 2006b, 2013a).

If internalism is true, then which of these views on the nature and whereabouts of sensible redness is best? And what makes it the case that BIV ostensibly experiences sensible redness as *bound with roundness*? These questions make up the puzzle of the sensible properties for internalists.

In developing the BIV argument for the irreducibility of phenomenal representation, I will mostly ignore the issue of the nature of sensible properties like sensible redness. I will mostly focus on how the BIV might phenomenally represent *spatial properties* like roundness. We will see that the BIV argument goes through no matter what view of sensible properties the internalist adopts.

Next I turn to elaboration and defense of the steps of the BIV argument for the irreducibility of phenomenal representation.

3. First Step: External Directedness

The first step of the BIV argument is this:

> *External Directedness: necessarily,* if any individual has tomato-like experience *R*, then that individual has an experience as of a *round* item of some kind. Further, that experience *fully matches the world* only if there is some item present that *is round*, that is, that *has edges roughly equidistant from a common point*.

I call this 'External Directedness', because when our experiences are "directed at" items of various shapes and standing in variation spatial relations, there need not be such items in our head. I already briefly alluded to External Directedness in the seductive "spatial argument" for externalism about experience (§1). Now I will clarify and defend it in more detail.

The notion of "matching the world" employed in the formulation of External Directedness can be explained by examples. Suppose you have the tomato-like experience *R* while hallucinating. Your experience does not match the world if there is really only a *rectangular table* before you. But if there happens to be before you a tomato on a white background, then we can all recognize a sense in which your hallucination *does* match the world (many philosophers have discussed such "veridical hallucinations").

My BIV argument only requires that the single experience *R* is necessarily directed at a *round* item. But I think similar claims apply to other visual experiences: other experience-types might be necessarily directed at a *square*

object, or an *object moving from left to right*, or one *object being above another*.

Why accept External Directedness? The initial argument is based on reflection. Consider the sentence "a round thing is present". Since language is conventional, we can easily imagine hypothetical cases in which those very marks mean that a triangular thing is present, or mean nothing at all. But could someone have *R* (an experience *exactly like* your experience of a stationary tomato on a white background), while *not* having an experience as of a *stationary round thing*, a thing whose edges are roughly equidistant from a common point? For instance, could someone have *that very experience*, and yet have an experience that is correctly characterized as an experience *as of* a *triangular thing moving to the right*? That just seems impossible. So even when BIV has *R*, BIV has an experience as of a *round* thing. This, together with the fact that there is no round thing there, is the only explanation of the evident fact that BIV has a *non-veridical* experience. As Block (Block 2012, sect. 12) puts it, BIV is "tricked".

There are other arguments for External Directedness. For instance, if a BIV with the general capacity for thought had *R*, it would thereby be in a position to have a false (but justified) *thought* with the content *something is that way*, a thought that is true only if something is *round*. Despite its sorry state, it could acquire a demonstrative concept of *roundness*. In general, having visual experiences is what explains our ability to form concepts of spatial properties and relations. It has a unique explanatory significance. How could this be so, if visual experience did not itself have built-in spatial content? (For another important argument for External Directedness, see Chalmers 2006, 74.)

The assertion that some visual experiences are necessarily externally directed is neutral on many questions. For instance, one question is: when you view a *tilted penny*, is the type of experience you have necessarily as of a thing that is *elliptical*, or of a thing that is *tilted and round*, or as of a thing that is "*elliptical-from-here"* (the view-point relative but objective property of having a shape that would be occluded by an ellipse placed in a plane perpendicular to the line of sight)?[12] Or are multiple answers correct? And do the spatial predicates needed to characterize the full accuracy conditions of experience express spatial properties that physical objects sometimes really have (Horgan 2014), or does contemporary physics show that this is not so (Chalmers 2012). As we shall see, my BIV argument is neutral on these issues.

External Directedness is also neutral between the main theories of experience. For instance, while it is implied by all versions of representationalism (including those of Tye, Dretske, Chalmers, and Horgan), it is also strictly speaking compatible with a purely non-representational view of visual experience, which denies that visual experience is essentially representational. In particular, the internalist might deny that BIV's tomato-like experience *R* has any "representational content" at all (perhaps on externalist grounds). Still, the internalist might accommodate External Directedness by invoking the traditional sense datum theory or Peacocke's sensationalism (2008). On this approach, BIV counts as having an experience

"as of a round item", in accordance with the letter of External Directedness, because BIV has an experience of a reddish and literally round *visual field region*. On this view, the experience "matches" a scene, only if the scene contains an object with the same shape as the visual field region.

But such a "visual field region" would be a peculiar non-physical item, for there is certainly no such round physical object inside BIV's brain. So internalists should reject this view. They should not recognize in this case the real existence of any such object (Peacocke 2008, 14).

So I think internalists should accept a broadly "representational" interpretation of External Directedness. BIV's experience simply has a false "representational content" to the effect of *there is a round item right there*, or *that is round*.

This still doesn't amount to "representationalism" about experience. Even on a "representational" interpretation, External Directedness only implies that in *some* cases visual experiences are *necessarily connected with* representational properties, specifically, representational properties involving space. This falls short of the representationalist thesis that, necessarily, *all* experiential facts *consist in* facts about the contents of our experiences. So even anti-representationalists could accept External Directedness.

In fact, Ned Block, a well-known anti-representationalist, favors External Directedness. He writes:

> [Representationalism] is the view that the phenomenology of an experience is the experience's representational (intentional) content. I am an opponent. But I am willing to allow that every phenomenological state has representational content, even that the phenomenology consists – in part – in its having that representational content. (2007a, 538).

Elsewhere Block (1995, 278) explicitly says that some types of visual phenomenology are "intrinsically" tied to certain *spatial* representational contents (*there is a circle there*, *there is a square there*, etc.), in agreement with External Directedness.

External Directedness is a modal claim. It says that *R* is *necessarily* directed at a round item. So the case for it is incomplete until consider whether it holds up in hypothetical cases.

Imagine that, on a Twin Earth, tomatoes are (what we would call) *ellipsoid*. Nevertheless, suppose that our Twin Earthians are wired up so that they normally have experience *R* (which *we* have on viewing *round* tomatoes) on viewing these ellipsoid tomatoes. Is *R* in any sense an experience as of a *round* thing, even on this Twin Earth where it is normally caused by ellipsoid objects? External Directedness implies a "yes" verdict. There are two reasonable accounts of the case compatible with this verdict. (i) On one account, while we get it right, our counterparts are regularly subject to a mild shape illusion. Horgan's realism about experienced spatial features (2014) implies this verdict. (ii) There is also Chalmers's irrealist view, which I

described before. On his view, in accordance with External Directedness, *R* is necessarily directed at a "round" thing (or in his terminology, a "perfectly" round thing) on Twin Earth as well as Earth. But neither tomatoes on Twin Earth nor tomatoes here on Earth possess this property, "perfect roundness", or "roundness-as-we-see-it". Rather, on the different planets, tomatoes merely have different versions of "imperfect" roundness; they only have (different) arcane quantum-mechanical properties, neither of which is anything like the "perfect" or roundness given in experience (roundness-as-we-see-it). The result is that *neither* our experience of our tomatoes *nor* our twins' experience of their tomatoes is perfectly veridical![13]

Let me address a final issue. You might think identity theorists, and internalists in general, should reject External Directedness, blocking my BIV argument at the first step.

In fact, the BIV case itself can be used to illustrate the thought. On internalism about experience, in having neural state *N*, BIV has the tomato-like experience *R*, despite being isolated from the external world. But, assuming a standard externalist approach to perceptual representation (e. g. Burge 2010), even though BIV has *R* (the very same experience you have on viewing a tomato), BIV cannot count as having an experience as of a *round* thing, contrary to External Directedness. For, in BIV, the neural state *N* does not have the required biological function of *tracking* round things, or indeed things of any kind. In fact, assuming standard externalism, BIV's highly detailed tomato-like visual experience *has no content at all* (somewhat like a state of undirected depression). In the BIV scenario, there is simply no sense in which the tomato-like experience *R* is essentially an experience as of a round item (e. g. no non-physical round "sense datum" or "visual field region" is present either), even though it is *exactly* like your experience of a tomato. In this regard, the tomato-like experience *R* is like the marks 'a round thing is present': this very experience could have had no spatial content at all, or any spatial content you please.

However, I think that internalists about experience (including identity theorists) should accept External Directedness and reject this argument. For there are very strong arguments in favor of External Directedness, as we have seen. (Again, if BIV's tomato-like experience has no content at all, why do consider it "non-veridical"?) Chalmers (2006) and Horgan (2014) are examples of philosophers who combine internalism about experience with External Directedness. As we have seen, Block, too, says that some visual experiences are "intrinsically" linked to certain spatial contents.

To block the above argument against External Directedness, internalists about experience should simply reject the assumption of externalism about phenomenal representation on which it depends. As Chalmers (2006, 83) and Horgan (2014) have emphasized, even if externalism is right in some cases (representational states about individuals or natural kinds like *tomato*), it simply does not follow it is true for *all* representational properties, including the representational properties of experience. *Some* basic forms of phenomenal representation must be internally-determined, if they are (as Block puts it)

"intrinsically" linked with phenomenology and phenomenology is internally-determined. So, for instance, by having the tomato-like experience *R*, BIV counts as having an experience as of a *round* object. This, together the absence of a round thing, explains why the experience is *non-veridical*. I don't see how the internalist could plausibly deny this.[14]

4. Second Step: Phenomenal Representation

The ultimate aim of my BIV argument is to show that internalism about experience requires a non-reductive theory of phenomenal representation. The first step was External Directedness: if BIV has the tomato-like experience *R*, then BIV thereby has experience as of a *round* thing, an experience that matches the world only if a *round* thing is present. The next step is to argue that, if this is right, then BIV bears the "phenomenal representation relation" to the property of being round (as depicted in Figure 1).

I assume a minimal *realism about properties*: there exists a rich set of spatial properties in the BIV scenario, instantiated by various objects (objects with which BIV cannot causally interact).

If there are properties, then the claim that BIV has an experience that is matches the world only if something is round is equivalent to the claim that BIV has an experience that matches the world only if something has *the property of being round*. This claim in turn immediately implies that BIV bears a representational relation to the property or attribute of being round. A rough gloss on this relation is as follows: $x\lambda x\lambda y$(*x* has an experience that matches the world only if something has property *y*). In the introduction, I called this the *phenomenal representation relation*, following Chalmers (2004). I will continue to use this terminology. As Burge (2010, 380) puts it, experience involves the *perceptual attribution* "of certain types or attributes—such as roundness, being to the left of". (My focus here on the general element of perceptual content is of course consistent with acknowledging a singular element.) In general, if you have the general capacity for thought, and if you phenomenally represent a certain property, then you thereby have the capacity to predicate that property of things in thought.

So if identity theorists accept External Directedness, as I have argued they should, then they must claim that, when BIV undergoes neural state *N* (on this view, visual experience property *R*), BIV thereby *phenomenally represents* roundness. As I have defined it, this is a dyadic relation between *subjects* and *properties*. Many internalists already accept this result, for instance Chalmers (2006, 107) and Horgan (2014). It is a very minimal claim that doesn't go beyond the pretheoretical claim of External Directedness. For instance, it doesn't require the additional claim that BIV's experience "aims at the truth" in the way beliefs do, or "lays claim" to the presence of a round thing (Papineau 2015, sect. 15). In fact, on some interpretations, I myself would reject this additional claim (Pautz 2010c, fn. 11).

The argument generalizes to other spatial and temporal properties. Assuming internalism, BIV might have all the same experiences as you, by

having the same underlying neural states. Then, like you, BIV stands in the phenomenal representation relation to *a large variety* of spatial properties and relations, such as *having orientation l*, *being above*, *having an edge at viewer relative place p*, etc. That is to say, it has experiences that match the world on the condition that there are things before BIV having these properties.

True, BIV does not see any *instances* of these properties. But the general properties still exist, and (by my argument) BIV bears a representational relation to them, in the above sense. In the actual world, whenever an individual has a hallucination, they likewise bear the phenomenal representation relation to perceptible properties that aren't instantiated before them. Indeed, something similar happens whenever you have a false belief. If you mistakenly believe something is the next room is round, then you bear the following relation to *being round,* even though it is not instantiated in your vicinity: you have a belief that is true on the condition that something in the next room have this property.

Recall that I am assuming that in the BIV scenario all of the relevant properties are instantiated by some objects or other (objects BIV cannot causally interact with). So even internalists who are leery of *Platonic, uninstantiated* properties – for instance, Mendelovici (2010) and Kriegel (2011) - must admit that these properties exist in the BIV case, and that in this case BIV bears the phenomenal representation relation to them. So they too face the question I am leading up to of whether this relation is reducible (§5).[15]

Of course, BIV phenomenally represents what I previously (in §1 and §2) called *sensible colors* in addition to spatial properties. For BIV has exactly the same experience *R* you had on viewing a particular tomato. And the following expanded version of External Directedness is hard to deny:

> *External Directedness II:* Necessarily, if an individual has the tomato-like experience *R*, that individual has an experience as of an item that is round *and* red_s.

Here the predicate 'red_s' could be defined ostensively: it expresses that familiar, salient quality which seems to you and your BIV counterpart to fill a round region when you have the tomato-like experience *R*. (I take it that such properties exist, for we can say true things about them: for instance, that red_s is more like $orange_s$ than $green_s$.) It appears to BIV that this property is co-instantiated with roundness. One view is that it is a property of a literally round "visual field region" (Peacocke 2008), but we should reject such items. We should rather say that the BIV *phenomenally represents* this sensible property. (Chalmers 2006, Horgan 2014).

BIV might phenomenally represent various other sensible properties. For instance, if it has an auditory hallucination, it has an experience *as of* an event having a certain location and certain audible qualities (sensible pitch and loudness). Its auditory experiences would fully match the world only if these properties are co-instantiated in a space around it. In that sense, BIV

phenomenally represents the conjunction of a certain location and certain audible qualities.

To sum up so far: internalists must hold that, simply in virtue of its neural states, BIV somehow phenomenally represents a large variety of spatial properties and relations as well as certain sensible properties.

5. Final Step: Irreducibility

The final step of the BIV argument is that, if BIV bears the phenomenal representation relation to spatial and other properties, then this relation is irreducible.

Before developing the argument, let me briefly explain what I mean by saying that a property or relation is reducible. Roughly, I say that a property *P* is *reducible to properties and relations Q, R, S . . .* just in case *P* is identical with a complex property built up from *Q*, *R*, *S . . .* (so that *Q*, *R*, *S . . .* are "ontologically prior" to *P*).[16] The *identity theory* that is the focus of my discussion is reductive theory of *monadic experience properties*. On this theory, the experience property *R just is* the complex neural-computational property *N*.

Many advocate reductionism about *all* "manifest image" properties: they hold that they are one and all nothing but complex properties built up from some limited set of properties and relations from the "scientific image" (e. g. the properties of the physical sciences plus certain "topic-neutral" properties and relations *causes* and *is a part of*). In this sense, they defend "reductive physicalism". They include Armstrong, Field, Jackson, Lewis, Papineau, Sider, and Smart, among many others. I myself think that this approach is right for *nearly* all manifest image properties instantiated in the world (with the phenomenal representation relation being a major exception).

I generally favor reductionism because it provides the simplest explanation of the relation between the manifest image and the scientific image. It is ontologically simple, since it holds that manifest image properties are *just identical with* enormously complex properties built up from scientific image properties. It is also simple in its stock of brute principles. It only requires general principles of property construction that everyone accepts. True, it requires brute identities between manifest image properties and complex physical-functional properties. But, as many have emphasized (e. g. Block 2003), identities have a very attractive feature: they do not cry out for further explanation. They are "explanation-stoppers". And, intuitively, identities do not add to the complexity of a theory.

Further, as Sider (2011, chap. 7) has recently emphasized, standard arguments against reductionism about the manifest image fail. For instance, along with Sider, I use "reduction" broadly enough that *functionalist theories are versions of reductionism*. So reductionism in my sense accommodates *multiple realizability* (e. g *being a chair* might be identical with a functional property). In addition, the *history of failed attempts* to provide complete reductions does not show that such reductions do not exist. The complete

reductions might be deeply *a posteriori*, or just too complex for us to specify (perhaps even infinitary).

Now for an important point emphasized by Hartry Field in his seminal "Mental Representation" (1978) and in a recent (2001) postscript to that essay. Philosophers often only focus on the reduction of monadic properties (with one "argument-place"). For instance, identity theorists focus on monadic experience properties like *having a headache*. But, as Field notes, there also exist relations. And, just as we can ask whether a monadic property is reducible, we can ask the same of a relation. To suppose the question is less pressing for relations would be an unjustified double standard.

Field focuses on relations concerning cognitive and linguistic representation: *individual x believes proposition y*, *individual x is thinking of existing concrete object y, name x refers to object y*, and *predicate x is satisfied by object y*. Other relations of philosophical interest include *x causes y* and *fact x provides a reason to perform action y*. I have added a relation to list: the phenomenal representation relation.

My focus is on identity theory. Identity theorists (Block, McLaughlin, Papineau) must heed Field's point. Even if they are right that *monadic* experience properties reduce to *monadic* neuro-computational properties of subjects, I have shown that they must also recognize the dyadic *phenomenal representation relation* between subjects and perceptible properties. And we can ask: according to identity theorists, is this relation reducible to some dyadic physical-functional relation between subjects and those perceptible properties, or not? This would require an interesting identity-claim of the following form:

> [#] $\lambda x \lambda y$(subject x phenomenally represents property y) = $\lambda x \lambda y$(x . . . y)

I will now argue that, even if identity theorists are right that *monadic* experience properties are reducible to monadic neural properties, they must say that the *dyadic* phenomenal representation *relation* is irreducible.[17]

The strategy of the argument is simple. As we have seen, on internalism, BIV bears the *phenomenal representation relation* to roundness and other perceptible properties. But BIV is isolated from the world. So it bears *no interesting dyadic physical-functional relation* to such properties, such as the tracking relation invoked by Dretske and Tye. In brief, given internalism about experience, our standard externalist models for reducing representational relations fail in the special case of the phenomenal representation relation. I will also provide a principled reason (the *disjunction problem*) for thinking that internalists cannot accept any alternative internalist model for reducing this relation. Let me explain these points in turn.

Broadly speaking, standard externalist theories of representational relations fall into two categories: input-based theories that emphasize what a state is apt to be caused by, and behavior-oriented theories that emphasize what actions a

state is apt to cause. But, given internalism, both are ruled out for the phenomenal representation relation.

The tracking representationalism of Tye and Dretske discussed in §1 provides an example of an externalist, input-based theory of phenomenal representation. Recall that tracking representationalists identify all perceptible properties, including sensible colors, with objective physical properties of external objects. Further, according to tracking representationalists, when individuals have experiences, the dyadic phenomenal representation relation that they bear to such properties is nothing but a complex tracking relation:

> $\lambda x \lambda y$(x phenomenally represents property y) = $\lambda x \lambda y$(x is in a inner state that realizes an experience and that *would* be caused by the instantiation of y were conditions biologically normal)

Call the relation named on the right-hand side the *tracking relation*. Then the idea is that *the phenomenal representation relation just is the tracking relation*.

Let me unpack this. What makes an inner state "realize an experience"? One idea is that its content must be cognitively accessible (Tye 2000, 62; Dretske 1995, 19; Prinz 2012). Another idea is that it must only satisfy some general neural background condition (for discussion see Block 2007a). This issue will not play a role in what follows. *Biologically normal conditions* are conditions in which the sensory systems are "operating as they were designed to do in the sort of external environment in which they were designed to" (Tye 2000, 138). (This is Tye's version; Dretske's is very similar.) This view provides a neat account of hallucination. In hallucination, you can bear the phenomenal representation relation to properties that are not currently instantiated in your vicinity (e. g. *being round and reddish*), because you are in a state that *would* be caused by the instantiation of those properties under biologically normal conditions.

But *internalists* about experience cannot identify the phenomenal representation relation with the tracking relation.

To see this, consider BIV. BIV is not attached to a body. It lacks eyes and the other receptor components of the visual system. So its inner neural state, *N*, is causally cut off from the environment. It doesn't track *anything* under "biologically normal conditions". Indeed, since BIV did not naturally evolve, there is no such thing what is "biologically normal" for it.

Granted, BIV's neural state *N would* track roundness, if *N were* "plugged into" a certain situation: for instance, if it were linked with a receptor system and a body in the same way it is in *normal humans* (Papineau 2014, 30). But recall the basic physical facts. *N* is just a distributed neural pattern occurring in BIV. It *could* have been causally connected to any shape you please (in general, anything can cause anything). (Compare: the expression "round" could have been use refer to any shape you please.) For instance, if *N* were plugged into a different neural environment, it could equally have tracked *being a triangle*. It also could have tracked any external color you please. In

fact, if it were hooked up to a computer, it might even track the *patterns of bits* in that computer providing the "sensory inputs"! Since BIV lacks an evolutionary history and belongs to no species, there is nothing to select one of these counterfactual situations as the "right" or "biologically appropriate" situation for BIV.[18]

The conclusion that tracking theorists like Dretske and Tye would draw is that, since BIV does not track any unique set of properties (even counterfactually), it does not phenomenally represent any. So, on their view, it entirely lacks experience. But, as I have argued, internalists must say that BIV does indeed have tomato-like experience *R*, and does phenomenally represent roundness and sensible redness and so on. So the conclusion they must draw is rather that the phenomenal representation is distinct from any tracking relation.

Consider next behavioral theories of representation. For instance, Evans (1985, 385) famously said that an experience "acquires a spatial content for an organism by being linked with behavioral output". A toy behavior-based reduction of the phenomenal representation relation might go as follows:

> $\lambda x \lambda y$(x phenomenally represents property y) = $\lambda x \lambda y$(x is in an inner state that realizes an experience and that, in typical members of the appropriate population, grounds the disposition to behave in ways "appropriate to" an object with property y)

Call the relation named on the right-hand side the *behavioral relation*. Then the idea is that *the phenomenal representation relation just is the behavioral relation*.

So, for instance, when you (a normal human) have *N*, then you have behavioral dispositions appropriate to a red and round object at place *p*: for instance, to reach out to place *p* and grasp exactly as if a round thing is at *p*, and to draw a round image if asked to draw a picture of what you see, to say "that's red", and so on. And when you have an experience of a green triangle you have another suite of sensorimotor dispositions. Such behavioral dispositions determine what properties you phenomenally represent, according to the behavioral theory.

Now for general reasons the behavioral theory of phenomenal representation is hopeless. There is simply no backward road from behavior back to the content of experience. What does it even mean to say that some set of behavioral dispositions is "appropriate" to *an object with property y*? For instance, what are the behavioral dispositions that are uniquely "appropriate to" or "fit" an object of specific shade of red, or a specific shade of white? The possibility of behaviorally undetectable spectrum inversion suggests that there is no such thing.

Even setting aside these problems, *internalists* about experience cannot identify the phenomenal representation relation with the behavioral relation.

To see this, consider BIV. BIV is not a complete human body. It is merely a duplicate of *part* of a brain. So it has no interesting behavioral dispositions. If you throw a tomato at it, it will just sit there.

Of course, *in normal actual humans*, *N* grounds certain behavioral dispositions. But this point cannot save the behavioral theory. For, in some other possible species, *N* might be hooked up to quite differently to a body, resulting in quite different behavioral responses. Indeed, it could be hooked up to a computer, so that its "behavioral responses" are digits on a computer screen. Since BIV belongs to no species and has no evolutionary history, there is nothing that could select one possible species as "the appropriate population" or one embodiment as the "normal embodiment".

It follows that BIV has no unique set of behavioral dispositions. Therefore, it does not bear the behavioral relation to any properties whatever. Nevertheless, as I have argued, internalists must say that BIV phenomenally represents roundness and sensible redness and so on. So they must conclude that the phenomenal representation relation is distinct from the behavioral relation.

In sum, BIV doesn't bear to its environment any of the causal, informational, theological, or sensorimotor relations invoked in any of our standard theories of representational relations. So, when it comes to providing a reductive theory of the phenomenal representation relation, it is as if internalists about experience have both hands tied behind their backs. If they can provide a reductive theory of phenomenal representation, it would have to be radically different from all current theories. How might such a theory go?

I will consider such a reductive theory. I call it "disjunctivism". But it fails. Indeed, it will lead us to a *principled* reason why any internalist reductive theory of the phenomenal representation relation must fail.

Let us focus on the brain-based identity theory of monadic experience properties. Recall that on the identity theory the tomato-like experience property *R* is identical with the neural property *N*. Let *N1, N2, N3*, . . . be the indefinitely-many other neural properties which, on identity theory, are identical with some experience property or other that individuals actually undergo (for instance, the experience of a *square* object, the experience of an object *moving to the left*, an experience of one object *above* another, and so on).

We can reach "disjunctivism" about the phenomenal representation relation in two steps. First of all, if what I have argued for so far is correct, then identity theorists are committed to indefinitely-many entailments of the following form:

Having *N* entails phenomenally representing *being round* (if it exists)
Having *N* also entails phenomenally representing *being red*$_S$
Having *N1* entails phenomenally representing *being square*
Having *N2* entails phenomenally representing *moving to the left*
Having *N4* entails phenomenally representing *being green*$_S$
Having *N5* entails phenomenally representing *having pitch p and lo cation l . . . etc*.

The identity theorist is committed to these entailments, because each of the listed neural properties is identical with a unique experience property, which (by my argument) necessitates phenomenally representing a unique cluster of properties.

Now this list does not specify a dyadic physical-functional relation (having two argument-places) with which the dyadic phenomenal representation relation might be identified. It does not specify an identity of form [#] above.

But you might think that the list leads naturally to such an identity. For, if all the above entailments obtain, perhaps identity theorists can just identify the phenomenal representation relation with some *disjunctive relation* of the following kind:

> $\lambda x \lambda y$(*x* phenomenally represents property *y*) = $\lambda x \lambda y$(*x* has neural property *N* and *y* = *being round*, or *x* has neural property *N1* and *y* = *being square*, or x has neural property *N2* and *y* = *moving to the left*, or *x* has neural property . . . and *y* = . . . *and so on*).

In short, the idea is that the *phenomenal representation relation just is the disjunctive relation*. This relation is basically just a "big" list of ordered pairs. Notice this is not disjunctivism about *monadic experience properties such as R* (we're assuming those are identical with non-disjunctive neural properties); rather, it is disjunctivism about *the phenomenal representation relation*.

Take an example. If you plug 'BIV' into '*x*' and '*being round'* into '*y'*, you get a truth (because then the first disjunct becomes true). Hence BIV bears the disjunctive relation to *being round*. So, on disjunctivism about phenomenal representation, BIV bears the *phenomenal representation relation* to *being round*, as desired.

On disjunctivism about phenomenal representation, nothing "unifies" the disjuncts. For instance, we saw that the BIV case shows that internalists cannot say that what is common between the disjuncts is that the relevant neural states *track* the corresponding properties.

Disjunctivism about phenomenal representation is exactly analogous to a reduction of the *name-object reference relation* famously considered and rejected by Field in "Tarski's Theory of Truth" (1972). On this theory, word *x* refers to object *y* iff *x* is the word 'France' and *y* = France or *x* is the name 'Eiffel Tower' and *y* = the Eiffel Tower . . . and so on for every name of English. Field notes we would never accept such disjunctive reductions in other cases (he discusses *being in pain* and *having valence n*). So we should not accept such a theory of the reference relation.

Block (2002, 412) is also skeptical of such disjunctive identities because they are not "explanatory". McLaughlin is skeptical for a different reason (2003, 181). So these identity theorists presumably would be skeptical about disjunctivism about the phenomenal representation relation (though neither discusses this case).

Indeed, I think that identity theorists (Block, McLaughlin, Papineau) and other internalists (Chalmers, Horgan) certainly cannot accept disjunctivism about phenomenal representation, for two reasons.

First, there is the *modal problem*. The disjunctive relation is defined in terms of a list of all the experience-constituting neural states actual creatures undergo, and the properties they actually phenomenally represent while undergoing those neural states. But surely there is a possible world where creatures have quite different neural structures and so phenomenally represent "alien" perceptible properties that cannot be on this list because they do not exist in the actual world.[19] Hence the across-worlds extension of the phenomenal representation relation exceeds the across-worlds extension of any such disjunctive or "big list" relation. It follows that the phenomenal representation relation is distinct from any such disjunctive relation.

There is another problem, the *indeterminacy problem*. I have developed the problem elsewhere (Pautz 2010a, 47-48), so here I will be brief. There is actually a huge abundance of variant disjunctive relations that are candidates to be the phenomenal representation relation. They might agree in extension when it comes to actual humans, but differ slightly or radically when it comes to remote actual or possible non-humans that we never interact with. (Compare the plus and quus functions in Kripke's (1982) discussion of Wittgenstein.) For instance, suppose that in the future we come across an alien creature (perhaps an alien brain in a vat). It has a complex sensory system and undergoes a radically different kind of neural state from us. Now one disjunctive relation, *D*, might pair its alien brain state with perceptible property *P*. Another, *D**, might pair its alien brain state with another, radically different perceptible property, *P**. Yet another might pair its alien brain state with another property, *P***. And so on. All these arbitrary disjunctive relations exist.

Now here is a problem for the disjunctivist. Intuitively, I can pretty easily refer to the phenomenal representation in my own case. And then I can go on to formulate various hypotheses about what the alien creature phenomenally represents. For instance, I might guess "the alien is aware of a shape and a color". Now the disjunctivist faces the following question: are there any physical facts that could determine that when I make such a guess about what properties the creature is "aware of" (that is, phenomenally represents), I am determinately glomming onto *one* of the variant disjunctive relations of the sort described above, rather than any of the others? I think that disjunctivists must answer "No". After all, all the variants fit my history of use of the predicate "*x* is aware of *y*" equally well. And it is not as if one of them stands out as being very natural and hence a "reference magnet" (Dorr and Hawthorne 2013). Rather, they are all equally unnatural and disjunctive. They are on a par. (In this respect, the puzzle here is unlike the puzzle about *plus* and *quus*, which might be solvable because *plus* is more natural than the other quus-like variants and hence a "reference magnet".) So the disjunctivist must say that it is *radically indeterminate* what disjunctive relation it is that I'm talking about.[20] That is to say, if disjunctivism is right, then there is no determinate fact of the matter about what disjunctive relation the phenomenal

representation relation *is*. But this has unacceptable results. For instance, it entails that I can truly say "it is indeterminate whether the alien is aware of (represents) spatial properties, or whether it is aware of (represents) properties that are nothing like spatial properties" and "it is indeterminate whether the alien is aware of (represents) sensible colors, or whether it instead represents properties belonging to a wholly alien quality-space". Intuitively, this is absurd. Given that there is some necessary connection between phenomenology and representation, it would mean that it is radically indeterminate *what it is like* for the alien.

These problems undermine any possible internalist reduction of the phenomenal representation relation. For there is no general, non-disjunctive algorithm, applicable to all actual and possible individuals, going from the intrinsic characters of those individuals' neuro-computational states to the properties they phenomenally represent. So any internalist reduction of the phenomenal representation relation will inevitably identify it with a massively disjunctive relation (where there are many variant disjunctive relations in the vicinity, having different extensions). Consequently it will be open to the problems I have identified.[21] Call this cluster of issues the *disjunction problem*.

The conclusion I draw is that, if an internalist theory of experience such as the identity theory is correct, then the phenomenal representation relation is *irreducible*. This view avoids all the problems I have developed for the reductive position. It may seem too radical. However, as I will explain in the conclusion, it may not be as radical as it seems.

My BIV argument for the irreducibility of the phenomenal representation relation has been entirely neutral between different solutions to the "puzzle of sensible properties" (§2) for internalists. Indeed, some views on this strengthen my case for irreducibility. For instance, given Chalmers's general irrealist view (discussed in §1) that in the actual world the sensible properties and even the spatial properties that we phenomenally represent are entirely uninstantiated, we can immediately rule out the claim that the phenomenal representation relation is reducible to a mind-world physical-functional relation like the tracking relation, even before we consider BIVs. Another view is Shoemaker's (1994) appearance property view, according to which (roughly) sensible properties are identical with properties of the form *normally causing internal neuro-functional state F*. Shoemaker only applies this view to the sensible properties we phenomenally represent; obviously, it wouldn't be plausible to generalize it to the *spatial properties* we phenomenally represent. Obviously, even if this view is right, it does not absolve the internalist of the need to answer the further question: what is the dyadic phenomenal representation relation that BIV bears to such properties? My BIV argument for irreducibility of this relation applies even if Shoemaker's view is right. For instance, even if Shoemaker's view is right, the internalist cannot identify this relation with the tracking relation, for exactly the reasons I have given. Of course, the internalist might combine Chalmers's or Shoemaker's view about sensible properties with "disjunctivism" about the phenomenal representation relation. That is, he might identify the phenomenal representation relation with a disjunctive or

"big list" relation (for this theory is neutral on the metaphysical status of the perceptible properties that feature on that list). But we have already dismissed disjunctivism.

A final point about the BIV argument. The internalist about experience might accept the premise of External Directedness but still try to somehow block my BIV argument for the irreducibility of the phenomenal representation relation. But let me remind the internalist what this would require. The internalist would have to at least gesture at a general *dyadic* physical-functional relation, *R,* between individuals and perceptible properties (with two arguments places, *x* and *y*), which is a good candidate to be the dyadic phenomenal representation relation. In other words, he would have to gesture at a completion of the general schema [#]. He would also have to make it plausible that this relation *R* has the same extension as the phenomenal representation relation (e. g. that your BIV-duplicate bears this relation *R* to all the relevant perceptible properties, the same ones you phenomenally represent). Until the internalist does this, he has not provided a response. As Sider (117) says, if we cannot provide even a toy "metaphysical analysis" of a relation, we have excellent evidence that it is irreducible (or in a sense "fundamental" as he puts it). And the disjunction problem provides an in-principle reason for thinking that this cannot be done in the case of the phenomenal representation relation.

We have arrived at our puzzle. If the BIV argument is sound, internalism implies that the phenomenal representation relation is irreducible. But isn't this a spooky view? Doesn't it require that internalists give up a physicalist view of the mind?

6. Sketch of A Possible Solution: The Internal Grounding View of Phenomenal Representation

I think that for internalists the most reasonable response to the BIV is to accept its conclusion of the irreducibility of the phenomenal representation relation. For instance, I think that accepting this conclusion is more reasonable than rejecting the premise of External Directedness, since the case for that premise is so strong. But others may not be so sure. They will regard this conclusion as extremely puzzling. So in closing I would like to briefly sketch a view that may help to reduce our sense of puzzlement. I call it the "internal grounding view". I will continue to focus on the brain-based identity theory, but other internalists could accept the same view. As it happens, the view meshes nicely with a general view of representational relations proposed by Paul Horwich and Stephen Schiffer on independent grounds.

First let me introduce the notion of grounding. Recently there has been a lot of enthusiasm about the explanatory potential of this notion (e. g. Fine 2001, Rosen 2010). The notion can be introduced by examples. The fact that John's action was done with the sole intention of harming *grounds* the fact that it is wrong. Or again, the fact that the apple is red grounds the fact that it is colored. Grounding is stronger than mere necessitation or entailment: in

addition, grounding involves an explanatory or determinative connection. It also differs from reduction. For instance, as Rosen (2010) notes, followers of G. E. Moore might say that natural properties ground normative ones, but deny that natural properties are reducible to normative ones. Or again, *being red* grounds *being colored*, but there is no obvious reduction in the vicinity (unless *being colored* identical with a disjunction with *being red* as a disjunct). Many think that, at a minimum, physicalism about the mind requires mental properties and relations to be *grounded in* physical (and topic-neutral) ones, even if they may not be reducible to them.

Now return to the tomato-like experience property *R*. On the identity theory, *R* is *identical with* neuro-computational property *N*. I have argued that identity theorists must hold that having *N* (on this view, *R*) *entails* phenomenally representing roundness. More generally, they are committed to the raft of the neural-representational entailments gestured at in the previous section. I have also argued that they must hold that the phenomenal representation relation is *irreducible*.

Now for the internal grounding view. It only adds one claim: these neural states do not merely entail, but also *ground*, standing in the irreducible phenomenal representation relation to certain clusters of properties. This is in line with the general physicalist creed that all mental facts are *grounded in* physical facts.

In short, I suggest that the identity theorist must accept quite different theories for experience properties and the associated representational properties. In the case of experience properties, he can retain the identity theory: they are *identical with* neural properties. Not so for the representational properties involved in experience. Unlike neural properties, they essentially have the form: *standing in the irreducible phenomenal representation relation to spatial and other properties P, Q, R, . . .* For such representational properties, the right model is grounding, not reduction. They are *grounded in* neural properties, but *not reducible to* them.

Maybe this is a workable view. In fact, in a discussion of an earlier version of the present essay, Jeff Speaks has endorsed its central argument, saying that "the truth of [the identity theory] would have as a surprising consequence of the irreducibility of [the phenomenal representation relation]"; but he adds "this is no immediate objection to [the identity theory]" (2015, 272).

For instance, the idea is that the monadic experience property *R* is identical with the neural property *N*. This property grounds, but is not identical with, the representational property of bearing the irreducible phenomenal representation relation to roundness. Likewise, the fact that two individuals (e. g. you and your BIV-duplicate) undergo the same monadic neural states grounds the fact that they bear the irreducible phenomenal representation relation to the same perceptible properties.

I conclude with some comments about this view.

(I) As I have mentioned, the internal grounding view of phenomenal representation accords nicely with a general theory of representation suggested by Horwich (1998) and Schiffer (2003, 162) and taken seriously by Field

(2001). On this view, representational relations (*believing*, *meaning*, etc.) are *generally* irreducible. For instance, Schiffer writes:

> What on earth could be the non-intentionally specifiable reducing relation in which "immaterial" stands to the property of being immaterial and by virtue of which the word means that property? (2003, 162)

Nevertheless, Horwich and Schiffer hold that, whenever a thing (a word, an individual) stands in an irreducible representational relation to some item, this is grounded in (or "constituted by") its having certain monadic physical property (a functional property, a use property, or whatever). So the view is still physicalist. As Field (2001, 71) puts it, according to this view, "these distinct monadic properties need have nothing to do with each other, and they certainly don't need to involve a common physical relation".

The internal grounding view of phenomenal representation is perfectly analogous. It concedes the irreducibility of the dyadic relation of phenomenal representation. For what on earth could be the non-intentionally specifiable relation in which BIV stands to the property of being round (etc.), with which the phenomenal representation relation might be identified? Nevertheless, the internal grounding view holds that, whenever an individual bears the irreducible dyadic phenomenal representation relation to certain perceptible properties, this is *grounded in* his being in a distinct monadic neural state (which is also an experiential state).

(II) Kit Fine has suggested a *congruence constraint* on grounding (2001, 20-21). Fine's constraint implies that if item *P* is real, then the ground of standing in a relation to *P* must itself involve standing in a relation to *P*. As Fine explains, this seems *generally* true. For instance, typically, you *refer to* a thing by virtue of standing in certain underlying (informational, etc.) *relations* to *that thing*. But if the internal grounding view is true, then the congruence constraint fails when it comes to phenomenal representation, because it holds that standing in the phenomenal representation relation to a shape property (for instance) is grounded merely in having a neural state, which is not itself a relation to that shape property. The Horwich-Schiffer view also violates Fine's principle. They conclude that the constraint is not generally valid. Horwich speaks in this connection of the "constitution fallacy" (1998, 25).

In fact, there may be other counterexamples to Fine's congruence constraint. For instance, the concrete, non-relational state of an object's having a certain mass *grounds* the relational state of its bearing the *mass-in-grams relation* to a certain number.[22] Also, the concrete, non-relational state of an object's being red grounds the relational state of its *instantiating* the abstract universal, *redness* (Horwich 1998, 25).

(III) The internal grounding view holds that in some cases there is a necessary connection between experience and representation. Yet it differs in several interesting ways from standard representationalist theories of experience (e. g. those defended by Dretske and Tye).

On the internal grounding view, *experience grounds phenomenal representation*. For instance, on this view, the state of having the tomato-like experience *R* is identical with the non-relational, concrete state of being in neural property *N*. This non-relational, concrete state then *grounds* the relational state of standing in the phenomenal representation relation to the abstract property *being round*. This fits with the key idea of the recently popular "phenomenal intentionality program" that representation is grounded in experience (e. g. Horgan 2014, Kriegel 2011, Mendelovici 2010). And it avoids the somewhat counterintuitive claim made by some representationalists that having an experience *consists in* standing in a representation relation to an *abstract object* (on the oddness of this claim see Pautz 2010c, 292ff and Papineau 2015, sect. 13).

The internal grounding view also provides an *internalist* account of phenomenal representation. Since experience is internally determined and grounds the representation of perceptible properties, the representation of perceptible properties is also internally determined. This makes phenomenal representation unique. Our standard externalist accounts (tracking accounts, teleological accounts) don't apply to it. The brain simply has an *intrinsic capacity* to phenomenally represent a certain clusters of basic perceptible properties (sensible colors, shapes, etc.) that need not be instantiated *in* the brain.

The internal grounding view is also quite compatible with the anti-representationalist idea that *some experiential differences do not correspond to differences in the phenomenal representation of properties* (Block 2007a, 538). For instance, maybe the difference between a blurry experience of a tomato and a clear one is a mere difference in the neural "content-vehicle", one that doesn't ground the phenomenal representation of any different perceptible properties.

(IV) You might think that in the end the internal grounding view doesn't provide totally satisfying solution to the puzzle of phenomenal representation. In particular, you might think it has two disadvantages.

First, the internal grounding view appears to be more complex than the kind of thoroughgoing reductive physicalism defended by philosophers like Armstrong, Field, Jackson, Lewis, Papineau, Sider, and Smart. On reductive physicalism, *all* properties and relations of the manifest image are just identical with complex properties and relations built from some basic stock of properties and relations from the scientific image. By contrast, the internal grounding view holds that individuals have properties of the form *standing in the irreducible phenomenal representation relation to so-and-so perceptible properties* that are *distinct from* all such complex properties, even if they are *grounded in* some of them. So it requires that individuals have "extra" properties. It also appears to require extra brute principles that a thoroughgoing reductive physicalist position would avoid. In particular, unlike reductive physicalism, this view requires "grounding connections" of the following kind: *if an individual has complex neural property N, then this grounds the distinct fact that this individual phenomenally represents the property of being round*.

And, on the internal grounding view, these appear *brute*. For, on this view, what could possibly explain such grounding connections? On the internal grounding view, they are not derivable from a general reductive theory of the phenomenal representation relation (e. g. the disjunctivist or "big list" view that this relation is identical with a "disjunctive relation" of the kind I described previously), for the simple reason that this view *rejects* any such reductive theory. Despite the recent enthusiasm for grounding, brute grounding connections can be objectionable in much the same way that brute psychophysical laws, or brute supervenience connections, are objectionable. My point here is not just that the internal grounding view requires an "explanatory gap" (something all standard physicalists have learned to live with); rather, the point is that this view requires extra metaphysically brute principles, which increases the complexity of the view.[23]

Second, as a non-reductive view, the internal grounding view implies that there is a certain kind of non-uniformity in nature that is avoided by reductive physicalism. (David Lewis has raised a similar complaint against non-reductive views of *normative properties*, as discussed by Jackson 1998, 27.) On reductive physicalism, everywhere in nature the only properties that are instantiated are the basic stock of properties and relations from the scientific image, together with complex physical-functional properties *C1*, *C2*, *C3* . . . built up from them. But on the internal grounding view, in some cases, there is more to say. On this view, *some* of the complex physical-functional properties *C1*, *C2*, *C3* . . . are "special" in that they ground *distinct* properties of the form *standing in the irreducible phenomenal representation relation to so-and-so perceptible properties*. So, for instance, the neural property *N* of the human brain that we have discussed in "special" in this way: it grounds the distinct property of *phenomenally representing roundness*. By contrast, the neural properties of the early visual system (which can be possessed in the absence of experience) do not ground any such distinct irreducible property. Likewise, the complex physical-functional properties of (say) an automobile engine do not ground any "extra" irreducible properties. This looks a bit arbitrary. Why do *some* complex physical-functional properties in nature (viz. certain ones of the brain) ground distinct irreducible properties, while others do not?

The internal grounding view, then, appears complicated and arbitrary. However, I think it could be replied that this appearance is largely due to the fact that we lack detailed knowledge of the brain, the most complex and amazing thing in the world. Maybe there are general, *systematic* grounding connections between our intrinsic neural patterns and what perceptible properties (shapes, sensible colors, etc.) we phenomenally represent, even if we have not yet discovered them (Pautz 2010b; but see Adams 1987 for interesting grounds for skepticism). And maybe, if we only knew them ("cracked the neural code"), we could look into a human brain, or BIV, and systematically "decode" what shapes and other perceptible properties the subject phenomenally represents. Then the internal grounding view of phenomenal representation would appear much less complicated and arbitrary.

Some neuroscientists have recently worked on such "brain-reading" (e. g. Haynes 2009). But it's still early days.

In sum, internalists face a choice. They can either try to find a fault with the BIV argument, or they can accept a non-reductive view of the phenomenal representation relation, such as the internal grounding view. I think that their most reasonable option is to accept such a non-reductive view of the phenomenal representation relation.[24] But, again, here my primary aim has been to raise a puzzle for internalists about experience – not to solve it.[25]

References

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

Brouwer, G. and Heeger, D. 2009. Decoding and Reconstructing Color from Responses in Human Visual Cortex. *Journal of Neuroscience* 29: 13992–14003.

Brouwer, G. and D. Heeger. 2013. Categorical Clustering of the Neural Representation of Color. *Journal of Neuroscience* 33: 15454-15465.

Berkeley, G. 1713. First Dialogue Between Hylas and Philonous. In D. M. Armstrong (ed.) *Berkeley's Philosophical Writings*. New York: MacMillian.

Block, N. 1994. Qualia. In S. Guttenplan (ed.) *A Companion to the Philosophy of Mind*. Oxford: Blackwell, 514-520.

—1995. On a Confusion about the Function of Consciousness. *Behavioral and Brain Sciences* 18: 227-287.

— 2002. The Harder Problem of Consciousness. *Journal of Philosophy* 99: 391-425.

— 2005. Two Neural Correlates of Consciousness. *Trends in Cognitive Science* 9: 46-52.

— 2007a. Consciousness, Accessibility and the Mesh between Psychology and Neuroscience. *Behavioral and Brain Sciences* 30: 481–548.

—2007b. Wittgenstein and Qualia. *Philosophical Perspectives* 21: 73-115.

Block, N. and J. Fodor. 1972. What Psychological States are Not. *Philosophical Review* 81: 159-181.

Block, N., and K. O'Regan. 2012. Discussion of J. Kevin O'Regan's *Why Red Doesn't Sound like a Bell: Understanding the Feel of Consciousness*. *Review of Philosophy and Psychology* 3: 89–108.

Burge, T. 2010. *Origins of Objectivity*. Oxford: Oxford University Press.

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, 153-182.

Chalmers, D. 2006. Perception and the Fall from Eden. In T. Szabo Gendler and J. Hawthorne (eds.) *Perceptual Experience*. Oxford: Oxford University Press, 49-125.
Chalmers, D. 2012. *Constructing the World*. Oxford: Oxford University Press.
Chang E. F, Rieger J. W, Johnson K, Berger M. S, Barbaro N. M, et al. 2010. Categorical Speech Representation in Human Superior Temporal Gyrus. *Nature Neuroscience* 13: 1428–1432.
Coghill, R., McHaffie, J., Yen, Y. 2003. Neural Correlates of Interindividual Differences in the Subjective Experience of Pain. *Proceedings of the National Academy of Sciences* 100: 8538–8542.
Crane, T. *Aspects of Psychologism*. Cambridge: Harvard University Press.
Crouzet, S. M., Busch, N.A. and Ohla, K. 2015. Taste Quality Decoding Parallels Taste Sensations. *Current Biology* 25: 1-7.
Danilova, M. and Mollon, J. 2012. Cardinal Axes are not Independent in Color Discrimination. *Journal of the Optical Society of America* 29: 157–164.
Dorr, C and J. Hawthorne. 2013. Naturalness. *Oxford Studies in Metaphysics* 8: 3-77.
Dretske, F. 1995. *Naturalizing the Mind*. Cambridge: MIT Press.
Evans, G. 1985. Molyneux's Question. In G. Evans (ed.) *The Collected Papers of Gareth Evans*. London: Oxford University Press, 364-399.
Field, H. 1972. Tarski's Theory of Truth. *Journal of Philosophy* 13: 347-375.
Field, H. 1978. Mental Representation. *Erkenntnis* 13: 9–61.
Field, H. 2001. Postscript to 'Mental representation'. In *Truth and the Absence of Fact*. Oxford University Press, 68-82.
Fine, K. 2001. The Question of Realism. *Philosopher's Imprint* 1: 1-30.
Haynes, J. 2009. Decoding Visual Consciousness from Human Brain Signals. *Trends in Cognitive Science* 13: 194–202.
Hill, C. 2009. *Consciousness*. Cambridge: Cambridge University Press.
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, 329-350.
Horwich, P. 1998 *Meaning*. Oxford: Oxford University Press.
Howard, J. D., Plailly, J., Grueschow, M., Haynes, J. D., Gottfried J. A. 2009. Odor Quality Coding and Categorization in Human Posterior Piriform Cortex. *Nature Neuroscience* 12: 932–939.
Jackson, F. 1998. *From Metaphysics to Ethics*. Oxford: Oxford University Press.
King, J. 1998. What is a Philosophical Analysis? *Philosophical Studies* 90: 155-179.
Kriegel, U. 2011. *The Sources of Intentionality*. Oxford: Oxford University Press.
Kriegeskorte, N. and Kievit, R. A. 2013. Representational Geometry: Integrating Cognition, Computation, and the Brain. *Trends in Cognitive Sciences* 17: 401-412.
Kripke, S. 1982. *Wittgenstein on Rules and Private Language*. Cambridge: Harvard University Press.

Lee, G. 2013. Materialism and the Epistemic Significance of Consciousness. In U. Kriegel (ed.) *Current Controversies in the Philosophy of Mind*. London: Routledge, 222-245.
Levine, J. 2001. *Purple Haze*. Oxford: Oxford University Press.
Lycan, W. 1996. *Consciousness and Experience*. Cambridge, MA: MIT Press.
Masrour, F. 2015. The Geometry of Visual Space and the Nature of Visual Experience. *Philosophical Studies* 172: 1813-1832.
McLaughlin, B. 2003. A Naturalist-Phenomenal Realist Response to Block's Harder Problem. *Philosophical Issues* 13: 163-204.
Mendelovici, A. 2010. Mental Representation and Closely Conflated Topics. PhD diss., Princeton University.
Ney, A. 2013. Ontological Reduction and the Wave Function Ontology. In A. Ney and D. Albert (eds.) *The Wave Function: Essays in the Metaphysics of Quantum Mechanics*. Oxford: Oxford University Press. 168-183.
Papineau, D. 2014. I—The Presidential address: Sensory Experience and Representational Properties. *Proceedings of the Aristotelian Society Hardback* 114: 1–33.
Papineau, D. 2016. Against Representationalism (about Conscious Sensory Experience). *International Journal of Philosophical Studies* 24: 324-347.
Pautz, A. 2006a. Sensory Awareness is not a Wide Physical Relation: An Empirical Argument Against Externalist Intentionalism. *Noûs* 40: 205-240.
— 2006b. Can the Physicalist Explain Colour Structure in terms of Colour Experience? *Australasian Journal of Philosophy* 84: 535–565.
— 2007. Intentionalism and Perceptual Presence. *Philosophical Perspectives* 21: 495-541.
— 2010a. A Simple View of Consciousness. In G. Bealer and R. Koons (eds.) *The Waning of Materialism*. Oxford: Oxford University Press, 25-66.
— 2010b. Do Theories of Consciousness Rest on a Mistake? *Philosophical Issues* 20: 333–367.
— 2010c. Why Explain Experience in terms of Content? In B. Nanay (ed.) *Perceiving the World*. Oxford: Oxford University Press, 254-309.
— 2011a. Can Disjunctivists Explain our Access to the Sensible World? *Philosophical Issues* 21: 384–433.
— 2011b. Review of Hill's *Consciousness*. *Analysis Reviews* 71: 393-397.
— 2013a. The Real Trouble with Phenomenal Externalism: New Empirical Evidence for a Brain-Based Theory of Consciousness. In R. Brown Ed., *Consciousness Inside and Out: Phenomenology, Neuroscience, and the Nature of Experience*. New York: Springer, 237-298.
— 2013b The Real Trouble with Armchair Arguments Against Phenomenal Externalism. In M. Sprevak and J. Kallestrup (eds.) *New Waves in Philosophy of Mind*. London: Palgrave, 153-181.
Peacocke, C. 2008. Sensational Properties: Theses to Accept and Theses to Reject. *Revue Internationale de Philosophie* 62: 7–24.

Prinz, J. 2012. *The Conscious Brain*. Oxford: Oxford University Press.
Rosen, G. 2010. Metaphysical Dependence: Grounding and Reduction. In B. Hale and A. Hoffman (eds.) *Modality: Metaphysics, Logic and Epistemology*. Oxford: Oxford University Press, 109-136.
Schiffer, S. 2003 *The Things We Mean*. Oxford: Oxford University Press.
Schmidt, B., M. Neitz, and J. Neitz. 2014. Neurobiological Hypothesis of Color Appearance and Hue Perception. *Journal of the Optical Society of America* 31: 195-207.
Shoemaker, S. 1994. Phenomenal Character *Noûs* 28: 21-38.
Sider, T. 2011. *Writing the Book of the World*. Oxford: Oxford University Press.
Soames, S. 1998. The Modal Argument: Wide Scope and Rigidified Descriptions. *Noûs* 32: 1-28.
Speaks, J. 2015. *The Phenomenal and the Representational*. Oxford: Oxford University Press.
Thompson, E. 1995. *Color Vision*. London: Routledge.
Tye, M. 1995. *Ten Problems of Consciousness*. Cambridge, MA: MIT Press.
Tye, M. 2000. *Consciousness, Color, and Content*. Cambridge, MA: MIT Press.
Tye, M. 2009. *Consciousness Revisited*. Cambridge: MIT Press.
Tye, M. 2014. Transparency, Qualia Realism and Representationalism. *Philosophical Studies* 170: 39-57.
Weatherson, B. 2003. What Good are Counterexamples? *Philosophical Studies* 115: 1–31.
Youngentob, S.L., Johnson, B.A., Leon, M., Sheehe, P.R. and Kent, P.F. 2006. Predicting Odorant Quality Perceptions from Multidimensional Scaling of Olfactory Bulb Glomerular Activity Patterns. *Behavioral Neuroscience* 120: 1337–1345.

---

[1] I first developed this form of argument in Pautz (2010a, sect. 7) and Pautz (2013b). In those earlier discussions, I focused on internalism about the experience of traditional "secondary qualities" (sensible colors, audible qualities, and so on). In the present essay I focus instead on the prospects for internalism about the experience of spatial properties ("primary qualities"), a doctrine that I think raises unique puzzles. For further discussion of the puzzle I will be raising, and other puzzles for internalism about experience, see Speaks (2015).

[2] As we shall see (§3), External Directedness is much weaker than "representationalism", the thesis that, necessarily, all experiential facts *consist in* representational facts.

[3] Various non-uniform theories are possible, but elsewhere (2010b, 353-354; 2013a, 286-288) I argue that they are problematic.

[4] In fact, I develop three distinct empirical arguments: the internal-dependence argument (Pautz 2010a, 2013a), the generalized structure argument (2006b, 2013a), and the judgment-explanation argument (2010b, 363, n.23, 358-359). Geoff Lee has suggested in discussion that there are yet other arguments against tracking representationalism in particular based on the fact that the stimulus/signal mapping is merely probabilistic and highly context-dependent.

[5] For the hypothetical "coincidental variation cases", see Pautz 2006a, 2010a, 2013a. These cases not only differ from inverted spectrum and inverted earth cases. They also differ from Block's more recent cases of "shift spectrum" (2007b, section 7) and cases of "pseudonormal vision" (2007b, section 10). In my view, there are problems with these cases and they fail to

undermine tracking representationalism (Pautz 2013a, 252-254; Byrne and Hilbert 2003, 19). My internal dependence, illustrated with "coincidental variation cases", avoids these problems.

[6] The recent empirical work I have cited suggests that the physical ground of phenomenal similarity is neuro-computational similarity. This neural-based view of phenomenal similarity also has the advantage of allowing that across-individual comparisons of experience make sense. For instance, on the neural-based view, my color experience of a tomato might resemble a monkey's more than a dog's, *if* our corresponding neural patterns stand in this resemblance-order. (See also Coghill *et al*. 2003 and Schmidt *et al*. 2014 on the neural ground of across-individual comparisons.) The only alternative to this neural-based view of phenomenal similarity is a behavioral-functional view according to which facts about phenomenal similarity, within an individual, can somehow ultimately be cashed out in terms of that individual's discriminatory and other behavioral dispositions. This view has the disadvantage of implying that across-individual comparisons of experiences never make sense (the "Frege-Schlick view"). True, the neural-based view requires that (i) there are well-defined measures of neural similarity, and (ii) that within an individual neural similarity in the relevant sense is the causal basis of that individual's similarity judgments and ordering behavior. But I think these claims are plausible: see Pautz (2013a, n.8) and Kriegeskorte and Kievit (2013).

[7] Nevertheless, Chalmers (2012, 439) holds that when we say "the tomatoes is round" we speak truly, so there is also a sense in which his view is realist. For more on this, see note 13 of the present essay.

[8] I have said that if Chalmers's irrealism is true, then internalism about spatial experience follows. It may be that we can also say that internalism about spatial experience is plausible *only if* irrealism is also true. For those who instead combine internalism with *realism* (e. g. Horgan 2014) face an explanatory challenge: if, contrary to externalist accounts, the brain entirely determines what properties we phenomenally represent, in a manner that is constitutively independent of links to the actual properties of external objects, then how come many of these represented properties – viz. the spatial properties and relations we phenomenally represent - happen to coincide with the actual properties of objects, as the realist says? Is this just some kind of lucky coincidence? Of course, an *irrealist* internalist like Chalmers avoids this explanatory challenge because he denies that there is such a coincidence to be explained: on his quite radical "Edenic" view, *none* of the properties we phenomenally represent is really out there. But for the realist internalist like Horgan (2014) there is a real question here. Elsewhere (2011, 418) I briefly offer a solution on behalf of the realist internalist, but I think it faces various problems and more needs to be said.

[9] For other empirical arguments against the externalist view that the experience of spatial properties can be explained in terms of tracking objective properties of objects, see Pautz (2011b, 394) and Masrour (2015).

[10] Block (2007a, 498) even suggests that experience may be present in an individual (like the patient he discusses identified as 'GK') *in the total absence of cognitive accessibility*. But it is a fact that many find this counterintuitive; even if the concept of phenomenal consciousness is not reductively analyzable in terms of cognitive accessibility, many think that there is *some* necessary connection here. Nevertheless, I think that there is a potential argument for Block's suggestion based on a "naturalness-plus-use" theory of reference, together with the conjecture that the *most natural or simple* general physical property (call it '*P*') that more or less fits our use of 'experience' across individuals is one which *excludes* the physical-functional machinery for cognitive access, so that it can be present without cognitive accessibility. In that case, considerations of naturalness or simplicity might support the claim that 'experience' refers to physical property *P* (and hence supports the identity 'experience = physical property *P*'), even if this violates our pretheoretical intuition or opinion that there can be no experience in the absence of cognitive accessibility (compare Weatherson 2003 and Sider 2011, 32). Analogy: the most natural candidate that fits our use of 'water' is simply H_20. A consequence is that water is in the air, even though we cannot detect it there (it is "inaccessible") by ordinary methods and so are pretheoretically disinclined to say it is there. However, I think that my BIV argument is neutral on whether experience is separable from cognitive accessibility. If you

think experience *does* require cognitive accessibility, then you can assume that BIV has the neural machinery underlying cognitive access.

[11] In fact, Tye (2009, 196) even suggests that he has *a priori* justification for thinking that, if the physical facts were as described in the BIV scenario, then experience would be absent – in which case internalist theories could be swiftly ruled out from the armchair. I disagree with Tye's claim, because it goes against the generally accepted point that there are simply *no a priori* links between non-experiential, physical conditions and experiential conditions (positive or negative). This point means that, contrary to many arguments in the philosophy of mind, you cannot describe a case (e. g. the BIV case, or the China-body system case) in purely physical, non-experiential terms, and then insist we have an *a priori* justification in favor of thinking that experience is present or absent in that case.

[12] For these options, see Tye (2000, 79) and Lycan (1996, 158). They advocate different versions of the idea that visual experiences have two levels of spatial content.

[13] Still, Chalmers holds that when here on Earth I *say* 'tomatoes are round', and when my twin on Twin Earth says 'tomatoes are round' (speaking of the physically different tomatoes Twin Earthians perceive), we both speak truly – our statements are "imperfectly veridical" (2006, 107; 2012, 331, 439). The reason is that Chalmers thinks the occurrences 'round' in our mouths do not refer to the uninstantiated "perfect" roundness we both phenomenally represent (roundness-as-we-see-it) in having our identical experiences of the tomatoes. Rather, my Earthian term 'roundness' refers to the quantum-mechanical "imperfect roundness" instantiated by our Earthly tomatoes, whereas my twin's term "roundness" refers to the different quantum-mechanical property instantiated by tomatoes on Twin Earth. (It follows that, on Chalmers's view, the ordinary English term 'round' is Twin-Earthable.)

[14] Nevertheless, in comments on an earlier version of this paper, David Papineau (who accepts some form of identity theory) said that his favored response to the BIV argument I'm now developing for the irreducibility of phenomenal representation *is* to deny my initial premise of External Directedness, despite its strong pretheoretical pull. (Formerly, he had been more ambivalent about External Directedness - see Papineau 2014, 30.) However, he has recently suggested a replacement claim, which we might call 'Internal Directedness': *Necessarily, if an individual (e. g. BIV) has tomato-like experience R (that is, on his identity theory, the intrinsic neural-computational property N), then there is an internal "sensory item" in the individual's experience that has a certain "visual shape", namely roundness*,* where he says that roundness* is "an intrinsic feature of the experience" (Papineau 2016, sect. 15). Now, by a "sensory item", Papineau must mean something like a population of neurons (or microchips, or whatever) within the individual's head (as a physicalist, he cannot say that it is a non-physical sense datum). And roundness* must be something like a neuro-computational property, *P,* of this population of neurons. So Internal Directedness just amounts to the claim that, necessarily, if you have the tomato-like experience *R,* then there is an item in your head with a certain neural-computational property *P* (which Papineau somewhat misleadingly calls "roundness*" *even though it is nothing like roundness*). In my view, this falls well short of accommodating what is immediately plausible on reflection. What is immediately plausible on reflection is External Directedness: necessarily, if one has the tomato-like experience *R*, then one has an experience as of a *round* item of some kind, which matches the world only if some item is present that *is round*, that is, *has edges roughly equidistant from a common point*. Evidently, the property of being round, unlike the neuro-computation property *P*, need not be instantiated by one's neural state when one has the tomato-like experience *R*. If External Directedness is false, as Papineau thinks, then why do we all agree that BIV's experience is *non-veridical*, that is, fails to match its environment?

[15] You might think that internalists like Mendelovici and Kriegel who reject uninstantiated properties must reject my initial claim of External Directedness, which says that the tomato-like experience *R* is *necessarily* directed at a round thing. For consider a different version of the BIV case in which the property *being round* is *not* instantiated in BIV's environment, and hence (according to their anti-Platonic view of properties) does not exist in that scenario. You might think that *R* cannot be directed at a round thing in *this* case, if there are no round things

and there is not even such a property as the property of being round. But this is not obviously right. In fact, Mendelovici and Kriegel would accept that even in this case *R* is necessarily directed at a round thing. For they accept a "non-relational" view of the intentionality of experience on which it is totally neutral with respect to the existence of perceptible properties as well as individuals. (See also Crane 2014. However, Crane is mostly concerned to deny that experience is "fundamentally" a relation to *propositions*; he doesn't address the issue of whether it necessarily involves being related to perceptible *properties*.) Now, as I explained in the text, even if they are right that having *R* doesn't *necessarily* involve standing in the phenomenal representation relation to the property of being round, the second step of my argument ("phenomenal representation") goes through, for External Directedness still implies standing in this relation to the property of being round *in the cases like the BIV case where that property exists* (see also note 17). However, for the record, I myself reject this view; I think that having *R necessarily* involves standing in the phenomenal representation relation to the property of being round, even in scenarios where *nothing* has this property. This implies the possibility of uninstantiated properties. For an argument for this view, and against the Kriegel-Mendelovici view, see Pautz (2007, 525-526) and Tye (2014, 51-52).

[16] The conception of reduction requires a somewhat rich ontology of complex properties. As formulated, it also requires that necessary coextension is insufficient for property-identity (for otherwise every property would count as reducible) – but an emendation might make it compatible with an intensionalist theory of properties (King 1998, fn. 22). Other, less ontologically loaded conceptions of reduction are available: for instance, a conception invoking Sider's notion of "metaphysical semantics" (2011), or one invoking the notion of a "real definition" (*to be F is to be G*).

[17] Previously, I mentioned the anti-Platonic view of Kriegel and Mendelovici that there are no *uninstantiated* properties. If an internalist goes further and accepts an extreme form of nominalism on which there are only individuals and there no properties at all (not even instantiated ones), then he will deny that External Directedness implies that BIV bears a "phenomenal representation" to clusters of *perceptible properties* in any scenario, for the simple reason that he thinks that there are no such things as properties. So he will dodge the question of how to reduce such a relation to a dyadic physical relation. However, given External Directedness, even the internalist who is a total nominalist must at least allow that BIV stands in a mind-world relation to *concrete objects and scenes*. For instance, if there happens to be a round tomato before BIV, then BIV stands in the following relation to it: $\lambda x \lambda y$(*x* has an experience that is accurate with respect to object or scene *y*). We might call this 'the veridical representation relation' (a term suggested to me by Uriah Kriegel). The arguments I will employ below to argue that internalism implies the irreducibility of the phenomenal representation relation could be used, *mutatis mutandis*, to establish the irreducibility of this "veridical representation relation". (For instance, he could not provide a "disjunctive" reduction of this relation, for reasons I'll explain.) So, given External Directedness, even the internalist who is total nominalist cannot avoid irreducible representational relations.

[18] In response to this problem of fixing the "appropriate" population for determining phenomenal representation, the internalist tracking theorist might simply specify in his account that *actual humans* are always the "appropriate" population. That is, he might say that *x* (in this case, BIV) phenomenally represents property *y* (in this case, *being round*) iff *x* is in some internal state or other (in this case, *N*) that, in *normal humans in @*, tracks the instantiation of *y*. This might be called the *human-centered tracking theory*. But as a general account this view is a non-starter, for several reasons. For one thing, it's absurdly chauvinistic to suppose that what perceptible properties any creature (BIV, pigeon, or whatever) phenomenally represents is a matter of what perceptible properties its inner states track in *actual humans* rather than some other population. Among other things, this makes facts about phenomenally representation totally *arbitrary and insignificant*. For instance, we can equally say that BIV "phenomenally represents" nothing but some bizarre properties involving *patterns of bits*, if we choose a possible population of brains-hooked-up-to-computers, rather than actual humans,

as the "appropriate" population. For another thing, when an individual phenomenally represents a property, then she can predicate that property of things in thought *because* she phenomenally represents it. But it would be bizarre to suppose that BIV, which occupies a different "possible world", can predicate *being round* of things in thought *because* BIV is in a state that tracks the instantiation of roundness *in humans and actual world @* (a species and a world that are totally remote from BIV). Finally, the human-centered tracking account has the mistaken implication that, if it turns out to be the case that we are *all* brains in vats (i. e. it is an illusion that there is such a natural kind as *humans* that we belong to), then "we sometimes have experiences *as of* round things" is false - since in that case it is false that we are in states that *track* round things in *actual humans*. There are yet other problems of a more technical nature with the kind of "actuality-based" maneuver employed by the "human-centered tracking theory" (Soames 1998, 15).

[19] These properties *do* exist in the actual world, if a version of Platonism about properties is true on which, necessarily, every property is such that necessarily it exists. But, needless to say, a cost of this solution to the modal problem would be a hyper-abundant ontology of uninstantiated properties.

[20] It might be the thought that the disjunctivist can say I achieve determinate reference to one of these disjunctive relations to the exclusion of all of the others because I think of it *by means of the description* of *the one that is the phenomenal representation relation*. But this response is obviously totally wrongheaded. For one thing, it presupposes what the disjunctivist needs to explain, namely, our ability to think of the phenomenal representation relation. For another, on disjunctivism, the phenomenal representation relation *just is* one of the disjunctive relations, which means that disjunctivists cannot sensibly accept this "by virtue of" claim.

[21] There is yet a third problem with disjunctivism. Consider again all the distinct, but very similar disjunctive relations I gestured at previously, whose extensions agree for humans but whose extensions diverge for non-human creatures. Since they are *very similar* to one another, it would be arbitrary to suppose that *one* of them, to the exclusion of all the others, has a special explanatory significance in enabling us to think about certain properties, and in providing immediate justification to our perceptual and introspective beliefs. Why *that* one? So disjunctivism is in tension with the common idea (mentioned in §3) that phenomenal representation has a unique explanatory and epistemic significance. Hawthorne 2006, pp. 108-9 and Lee 2013 discuss a similar issue. However their point is that "significance" is in tension with *unnaturalness*, whereas my point is only that it is implausible to suppose that a relation *R* but not a relation *R** has explanatory significance if *R* and *R** are objectively very similar (a principle that is entirely neutral on their levels of naturalness).

[22] There is a difference between the case of measurement and the case of phenomenal representation. While the mass-in-grams relation admits of a reduction in measurement-theoretic terms, the phenomenal representation relation cannot likewise be reduced. See Field (2001, 69-72) for the same point about representational relations more generally.

[23] Declan Smithies pointed out to me that proponent of the internal grounding view might reply that there *is* a kind of explanation of the fact that the neural property *N* always grounds the distinct property of standing in the irreducible phenomenally representation relation to *being round*. The explanation is simply that it "lies in the essence" of the neural property *N* that it always grounds the distinct property of standing in the irreducible phenomenally representation relation to *being round*. This fits with the idea that all modal facts derive from essence (see Rosen 2010 for discussion). However, in my view, this non-reductive view doesn't make any real progress in reducing complexity because it simply replaces brute grounding connections with brute essentialist connections between distinct properties. In fact, it is no less complicated than the initial internal grounding view. It also has the drawback of non-uniformity: it requires "special" essentialist connections concerning phenomenal representation of a kind we do not encounter elsewhere in nature. As I am about to explain in the text, non-uniformity is a drawback of any non-reductive view.

[24] I should say that, while I think that certain internalists (viz. identity theorists) ought to accept the "internal grounding view" of phenomenal representation that I have described, I do not

myself accept all elements of that view. (For one thing, I am skeptical about physicalism.) But I do accept a non-reductive, internalist view of phenomenal representation along broadly the same lines (Pautz 2010a, 2010b). David Chalmers (2006, pp. 83-84; 2012, pp. 342-344) accepts such a view as well, although his arguments are more *a priori* than mine (Pautz 2013b).

[25] Earlier versions of this paper were presented at Oxford University, the University of Southern California, and Brown University. I thank the audiences on those occasions for very helpful discussions. I would also like to thank Brian Cutter, Uriah Kriegel, Angela Mendelovici, David Papineau, Jeff Speaks, and Daniel Stoljar for helpful comments or discussion. Finally, I am conscious of a considerable philosophical debt to Ned Block, which I hope is evident in these pages.