Why Consciousness Cannot Just be in the Head[1]

A New Argument against Biological Theories

Adam Pautz
University of Texas at Austin

Qualia ain't in the head – Michael Tye

It is common ground that experiences-types are in some sense realized by neural states. On *biological theories* of consciousness, experience-types are *necessarily* identical with them. Biological approaches appear to be popular among neuroscientists. Consider, for instance, accounts of consciousness in terms of neural synchrony, re-entrant processing, and so on. And some philosophers have favored biological theories, including Bickle, Block, the Churchlands, Flanagan, Hardin, Hill, McLaughlin, and Polger. But among philosophers biological theorists are in the minority. There has long been a movement away from biological theories and towards functionalist-externalist theories. Of course, the first and most famous argument was based on multiple realizability. More recently, Michael Tye and others have invoked the "transparency observation". Finally, even more recently, Alva Noë and other "vehicle externalists" have wielded still other arguments.[2]

But I think that biological theories deserve to be taken very seriously. Many are beginning to appreciate that the multiple realizability argument is far from decisive. And I am not persuaded by the other arguments (e. g. the transparency argument) that have been brought against biological theories to date, for reasons I cannot explain here (but see note 2). I also think that biological theories enjoy some measure of empirical support. Against certain radical externalists and in agreement with biological theorists, I think internal neural factors play a serious role in shaping the character of phenomenal consciousness. In fact, Daniel Dennett (though himself a functionalist) has conceded to biological theorists that "the recent history of neuroscience can be seen as a series of triumphs for the lovers of detail". Ned Block has gone so far as to say *"mind-body identity is back"* (although, as we shall see in §5, interestingly, Block's response to the argument of this paper suggests that his actual view may also have significant functionalist elements).[3]

---

[1] ***Note to readers: reading – or even skimming - §§1-3 (i. e. only up to p.15) would be enough to get the gist*** My thanks to David Chalmers, Susanna Schellenberg, and Michael Tye for helpful comments or discussion. I am especially indebted to Ned Block for comments which (I hope) have led to many improvements. This remains very much a rough draft – comments would be much appreciated!

[2] For defenses of biological theories among neuroscientists, see Crick and Koch 1990, Driver and Vuilleumier 2001, and Kanwisher 2001. For defenses of biological theories among philosophers, see Bickle 2003; Block 1980, 2009; P. M. Churchland 2005; P. S. Churchland 1986; Flanagan 1992; Hardin 1987; Hill 1991 (however Hill has recently converted to externalist intentionalism); McLaughlin 2003; 2007; Polger 2004. I think that these philosophers have adequately answered the much-vaunted 'multiple realizability' argument. For the transparency argument, see Tye 1995. Tye says 'qualia are not to be found in the head'. I raise some objections to this argument elsewhere (Pautz 2007); see also foot-note 15 the present paper. For Noë's arguments, see his 2004, 2007, 2009. I think that these arguments fail for reasons pointed out by Aizawa (2007), Block (2005), Clark (2006) and Prinz (2006).

[3] The Dennett quote is from his 2001. The Block quote is from his 2009 Jack Smart lecture "Why Consciousness Does Not Extend Outside The Brain" (podcast available at the ANU website).

However, I will here develop a new argument against biological theories which I think succeeds where others fail. Like early arguments against biological type-type theories, it will take the form of a simple application of Leibniz's Law. Many biological theorists would reject "strong" intentionalism. For instance, Block argues that some phenomenal differences among experiences (involving visual blur, phosphenes, attention shifts, and so on) are differences in "non-intentional qualia". So in his view the "transparency thesis" is mistaken. But even if we grant to biological theorists that the transparency thesis is false, we should insist on something else: that experience-types have *some* "externally-directed" properties necessarily built-in. In particular, I will argue that *some* experience-types are at the very least necessarily directed at the following:

- spatial properties and relations
- dynamic/temporal properties

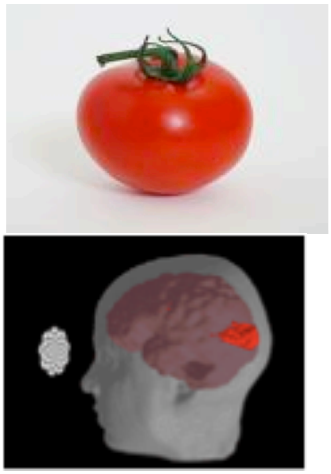
But, I will argue, neural-types do not have such externally-directed properties necessarily built-in: like the sentence 'a round thing is in front of me', they have all their externally-directed properties only *contingently*. So experience-types cannot be neural-types. Biological theories are not merely wrong; they are wrong in principle. *Neural-types are just not the sort of things that could be experience-types could be.*[4]

My plan is as follows. In §1 I explain biological theories and the structure of the argument. In §§2-3 I defend its key premises. In §4 I explain how the argument differs from, and is superior to, an argument against biological theories suggested by Hilary Putnam. In §5 I address objections to my argument, including an objection due to Ned Block which has some functionalist elements. In §6 I suggest that the argument supports an intentional theory of phenomenal consciousness. Finally, in §7, I conclude by describing an overlooked puzzle arising out of the discussion. I think the externally-directed character of consciousness rules out the biological theory. But, as I said, I think biological theories also contain an element of truth: consciousness is internally-dependent. However it is very hard to see how consciousness might be both externally-directed and internally-dependent. I use Shoemaker's influential internalist intentionalist theory of sensory consciousness to illustrate the puzzle.

1 Preliminaries

Let '*R*' rigidly denote the maximally-specific *experience property* or *phenomenal-type* someone has when he looks at a tomato in good light on a particular occasion, and that he would have in any phenomenally identical situation (see figure to be reminded of what a tomato looks like).

[4] One way of thinking of the argument is as a kind of *transcendental argument:* a thread running through much recent work on perception is an emphasis on role of experience in making possible, and justifying, external thought. My argument will show that experience could not play this role if experiences were necessarily identical with purely internal brain states.



By *biological theories* of experience, I mean theories according to which *R* is *necessarily* identical with some neural property *N*, and in general, that every experience property is *necessarily* identical with some neural property. By a neural property, I mean a physical-computational property whose definition only involves matters inside the head. An alternative, rough way of formulating biological theories is as physicalist theories which deny any claim of the form 'necessarily, *x* has *R* iff *x* has a state that plays functional role *F*, where *functional role* is understood to include any facts that reach outside the head: facts about behavioral dispositions, facts about causal-covariation with external properties under historically-determined optimal conditions, facts about what external properties our inner states have the function of indicating, facts about sensorimotor profiles, and so on. This makes biological theories interestingly different from alternative physicalist theories.

It is worth pointing out that nothing in my argument will assume that the biological theorist holds that *R* is necessarily identical with a 'local' neural property. In Shoemaker's (1981; 2007, 21) terms, I am not assuming that he holds that *R* is necessarily identical merely with its *core realizer*: roughly, the neural property that comes and goes as the instantiation of *R* comes and goes. Instead, I will assume that he holds that *R* is necessarily identical with a 'global' neural property involving more than its core realizer: as it might be, the core realizer *plus* re-entrant processing *plus* a bit of neural surround (this issue is discussed again in connection with the second objection addressed in §5).

Note that my target is the claim that every experience *property* is necessarily identical with some neural *property*. My target, then, is necessary *type-type* identity, not token-token identity. It would be conceivable to accept my arguments against type-type identity (say, embracing externalist intentionalism about phenomenal types) and go on to advocate token-token identity about particular experience-tokens.[5]

[5] Biological theories must also be distinguished from David Lewis's (1994) *realizer functionalism*. It is true that Lewis held that names for experiences in English like 'the experience as of a red and round thing' contingently refer to certain neural realizers. But he denied the distinct claim of biological theories that experience properties like *R* are necessarily co-extensive with these neural realizers. Lewis regarded himself as providing a broadly functionalist theory of the mind and, like functionalists, he claimed that experience

Note also that my target is not internalism about phenomenal consciousness, which I formulate (following Lewis, Jackson and others) as kind of holistic sufficiency claim to the effect that *total* neural duplicates (duplicates with respect to the state of the peripheral input-output systems as well as more central brain states) that *live under the same laws of nature* have the same experiences. Such duplicates will not only agree in neural states; they will also agree in behavioral dispositions and in many functional respects. So internalism can be accepted by functionalists like Lewis and Jackson and Shoemaker. Nothing in what I will say in what follows cast doubt on this claim. Again, my target is the only claim that every individual experience property is necessarily identical with an individual internal neural property that is both sufficient and necessary for the experience property.

I will focus throughout on *R*. My *Leibniz's Law argument-template* is as follows:

1 Experience property *R* has externally-directed property *P* necessarily
2 No neural property *N* has externally-directed property *P* necessarily
3 Therefore *R* is not identical with any neural property *N*

2 The First Premise

I will argue that premise 1 holds for *four* values of '*P*'. My case is based on four *externally-directed intuitions* concerning *space*.

Some preliminary remarks. First, I regard these intuitions as sacrosanct. But I realize that many will say that the appeal to intuition is just a *prima facie* starting point. The intuitions must be tested against various bizarre scenarios: inversion scenarios, brains in vats, swampmen, Brad Thompson's double earth case, and so on. I believe that the intuitions easily withstand consideration of such scenarios. However, rather than get bogged down in a discussion of such cases at the start, I will in the present section merely introduce the intuitions, saving general objections about such cases for the objection-reply section §4. (I will, however, address some more specific objections that might occur to the reader in foot-notes.)

A second preliminary point. It might be wondered why I focus on four externally-directed intuitions, if just one would do. I focus on more than one externally-directed intuition because it will make answering the argument difficult for the biological theorist. He would have to find, for each of the intuitions, a good reason to reject that intuition. This is a tall order to fill, since the intuitions are importantly different. Further, the intuitions are also interesting in their own right. They not only show that the biological theories are mistaken; each yields a unique desideratum that any alternative theory of phenomenal consciousness will have to accommodate. After developing in the argument, in §6, I will address the issue of what such a theory might look like.

---

properties are necessarily co-extensive with input-output functional properties that reach outside the head. Therefore his theory of experience delivers different verdicts than biological theories in certain possible cases. In fact, Lewis claimed that gerunds like 'having *R*' (as distinct from names like 'the experience as of a red and round thing') can be taken, "at least on one good disambiguation", as *rigid designators* of such functional properties as opposed to neural properties (Lewis 1994). Since Lewis's theory of experience was functionalist, it is invulnerable to my argument from the externally-directed properties of experience.

Now for the intuitions. First, the *externality intuition*. Imagine looking at the (real) tomato depicted in figure 1. Now consider any possible situation in which you have the property of having an experience with this very phenomenal character – that is, the property *R*. Intuitively, in all of them you have an experience as of *round* thing at a certain viewer-relative distance *d* and position *p*. This is so even in hallucinatory cases. How could there be counterexamples to this claim? For instance, even if there are possible situations in which *R* is normally caused by grey squares (something denied by certain externalists), *R* would intuitively still be an experience *as of* a round thing.[6]

Since *'R'* is a rigid designator, the *de dicto* externality intuition entails the *de re* claim that *R* has the following (second-order) property *necessarily:*

> *The externality property:* being a property such that, if anyone has that property, then he has an experience as of a round thing at a certain viewer-relative distance *d* and position *p*.[7]

The externality intuition, and all of the other intuitions to be introduced below, concern *R*'s built-in directedness at *spatial properties and relations*. This will be the principal focus in what follows. However, I will (in §3) briefly use experience's built-in directedness at *temporal properties* to illustrate the argument.

Although it is in no way required by my argument, I believe that the externality intuition, and all of the other intuitions to be discussed, generalize to color: necessarily, if one has *R*, then one an experience as of a red thing at a certain viewer-relative distance and position. Many would reject this on the basis of spectrum inversion and the like. But they would grant that, necessarily, if one has *R*, then one has an experience as of a red_p thing at a certain viewer-relative distance and position, where it is left open

---

[6] *Objection:* The externality intuition is false. For the externality intuition requires that, necessarily, if some one has *R*, then it seems to them (they have an inclination to *believe*) that there exists a *mind-independent* round object at a certain distance from him. But small children and animals provide a counterexample. So does someone who has *R* while knowingly having a hallucination.

Reply: This objection rests on a misunderstanding of the externality intuition: it reads too much into the intuition. The intuition does *not* require that, necessarily, if some one has *R*, then he has an inclination to *believe* that the there exists a *mind-independent* round object at a certain distance from him. Indeed, 'mind-independent' and 'belief' do not occur in the formulation of the intuition at all. Small children, animals, and tipped-off hallucinators who have *R* would not be counterexamples to the externality intuition, because, intuitively, their *experience* would be as of a round thing of *some sort*, even if *they* do not *believe* that there is a round, *mind-independent* thing there. Likewise for the other externally-directed intuitions to be introduced below. This 'mind-independence' claim should not be read into my talk of 'external directedness'. All I mean by 'external-directedness' is what is asserted in the formulation of the intuitions.

[7] *The externality intuition does not amount to the claim that R necessarily has a certain "intentional content" to the effect that a red and round object is present. This claim about 'intentional content' is controversial theoretical claim that would be rejected by certain disjunctivists and sense datum theorists, so it is not a good starting point.* By contrast, the externality intuition is pretheoretical and can be formulated in ordinary language. It is compatible with an intentionalist theory but does not amount to such a theory or indeed any other controversial theory. The 'as of' does not automatically indicate intentional content. For instance, a *sense datum theorist* (at least one who thinks sense data are literally colored and shaped) could say that *R* has the externality property necessarily because it necessarily consists in the awareness of a mental *red* and *round* sense datum – a item that is not located in the brain but causally dependent on the brain. A *disjunctivist* could say that *R* has the externally-directed property necessarily because having *R* necessarily consists either seeing the *redness* and *roundness* of something outside the brain or being in a state that cannot be discriminated by reflection from such seeing.

whether red_p is the color red (Tye), a perfect color (Chalmers), qualitative character (Shoemaker), or whatever. I believe that this would be enough to rule out the biological theory, but since it raises complex issues about the nature of red_p, I will focus mainly on shape.

The next intuition is the *matching intuition*. If one has *R* while seeing a tomato, then one's experience *matches the world* or *corresponds to how things are*. Likewise, if one has *R* while hallucinating, and there happens to be a round object present, then one's experience matches the world. If, by contrast, one has *R* while a square thing is present, then one's experience does not match the world. The matching intuition about *R* says: *necessarily,* if one has *R*, then one is in a state that matches the world only if a round object is present. To see this, look at Figure 1 again. Now consider any possible situation in which you have the property of having an experience with *this very phenomenal character* – that is, the property *R*. Intuitively, this state *corresponds to the way things are* only if a round object is before you. Likewise, if a brain in a vat might have *R*, or if someone might invariably have *R* as a result of the presence of squares, they too are in states that match the world only if a round object is present. This is why we regard their experiences as somehow defective. Several philosophers have persuasively defended the point, although they have not used it against biological theories.[8]

The matching intuition entails that *R* has the following (second-order) property *necessarily*:

> *The matching property:* being a property which is such that, if anyone has it, then he is in a state that matches the world only if a round object is present.

Next, there is the *grounding intuition* about *R*. One of the most important facts about experience is that it makes externally-directed thought possible.[9] Intuitively, it is necessary that, if an individual *with the general capacity to have belief* has *R* (for a sufficient period of time), then he will thereby have the additional capacity to have a *general* belief that is true only if something

[8] See Horgan and Tienson 2002, Loar 2003, and Siewert 1998. There are some potential points of difference between the matching intuition as I understand it and the claim made by these philosophers. First, these philosophers formulate their claim by saying that experiences are essentially assessable for *accuracy* by virtue of their phenomenology. Their use of 'accuracy' suggests that it is part of their claim that experiences have a *mind-to-world direction of fit* or *purport to represent the world* in the manner of beliefs – a very controversial claim that would be rejected by disjunctivists and sense datum theorists. This controversial is not part of matching intuition as I have formulated it, so the matching intuition could in principle be accepted by disjunctivists and sense datum theorists. (Analogy: a picture made at random by a computer might 'fit' a round thing, without having a mind-to-world direction of fit or purporting to represent.) Second, these seem to think that their claim supports a broadly intentionalist theory according to which there is some deep explanatory link between experience and intentional content. My own view is that the matching intuition does not require the truth of these controversial claims, since it is compatible with sense datum theories and disjunctive theories. A disjunctivist could say that the pretheoretical sense in which *R* essentially matches a round object is simply that in having *R* one either sees a round object or is in a state indiscriminable from such seeing.

[9] Since the grounding intuition is a conditional claim applying only to believers, it is quite compatible with this intuition that a dog lacking the capacity for conceptual thought altogether might have *R* and yet lack the capacity to have such beliefs. And, since the grounding intuition is formulated in terms of *general beliefs*, it is quite compatible with the grounding intuition that having *R* does not necessarily endow individuals with the capacity to have *singular* beliefs about particular objects. For arguably in hallucinatory cases one's apparent singular thoughts are merely "mock thoughts", there being nothing for them to be about. See Evans 1982, 29-30.

or other is present that is round. *If* a brain in a vat might have R, or if someone might invariably have R as a result of the presence of squares, they too would intuitively have the capacity to such (false) beliefs. Intuitively, the contents of these beliefs would somehow derive from R, rather than the external world.[10] To see the pull of the intuition, suppose someone has R as part of a lifelong series of hallucinations that match your actual experiences. Intuitively, he would have the capacity to have all the same shape-thoughts as you have. For instance, like you, *he would understand Euclidean geometry*. Therefore his experiences would ground the capacity to have some thoughts about *shapes*, not only thoughts about internal "shape qualia".[11]

Finally, there is the *justification intuition* about R. Recently, many philosophers have rightly stressed the rational role of experience.

---

[10] *Objection:* The grounding intuition requires the implausible claim that *R* necessarily endows believers with the general concept *round*. This is implausible because possessing this general concept requires being presented with many round objects. And it may require implicitly believing that roundness is a mind-independent property of objects that are invariant with changes in the viewing conditions. Arguably, one cannot satisfy these requirements merely by having any single experience such as *R*.

Reply: This objection rests on a misunderstanding of the grounding intuition. It does not require the admittedly implausible claim that having *R* alone necessarily brings with it possession of the general concept *round*. Rather, it says that having *R* endows an individual with the capacity to have a belief that is true only if a round object is present. Here *round* is used outside any 'believes that ___' context to characterize the truth-conditions of the individual's belief. Therefore the grounding intuition does not require that the subject himself has that concept. It may be that he only has a short-lived *demonstrative* concept, so that his belief is of the form *something is that way*. Or maybe he has a belief with these truth-conditions by merely *endorsing his experience*.

[11] *Objection:* The grounding intuition would only be endorsed by proponents of accessibility theories of consciousness. So it is not a really a pretheoretical intuition, but amounts to a controversial functionalist theory of consciousness. Of course, biological theorists like Ned Block would reject such a theory.

Reply: It is important that the grounding is much weaker than those theories, and much less vulnerable to attack. Consider, for instance, Michael Tye's accessibility theory. On this type of view, the neural machinery realizing experience might be distinct from that realizing cognitive access; but it only realizes experience when it is at least potentially linked to the machinery realizing cognitive access. In particular, one has an experience with content *p* if, and only if, one is in a state that appropriately represents *p* and that is poised to produce the belief (or desire) that *p*. This is quite strong. It is a universal claim applying to all possible subjects, experiences, and contents. And it is provides sufficient as well as necessary conditions for having an experience in non-phenomenal terms.

By contrast, the grounding intuition only applies to believers who have *R* for a certain non-trivial period of time and the single content *a round object is present*. And it does not provide necessary and sufficient conditions for having *R* in non-phenomenal terms. This means that the grounding intuition avoids putative counterexamples to accessibility theories. It avoids counterexamples concerning babies and animals who allegedly have the capacity for experience but not belief, because it is only meant to apply to believers. It avoids alleged counterexamples involving states that are suitably poised but do not realize experiences, because it does not aspire to identify a *sufficient* condition for having an experience in non-phenomenal terms. Finally, Sperling-type brief-presentation experiments, which are often thought to be problematic accessibility theories, are not a problem for the grounding intuition. Such experiments clearly refute the claim that, if an individual has a certain experience *for an extremely short period of time* (e. g. 500 ms), then he might *retain* in working memory *all* the contents of that experience *immediately after he has had the experience*. But the grounding intuition differs in several respects from this claim: it is synchronic rather than diachronic, it only concerns having an experience for a non-trivial duration, and it only concerns a single content. In particular, it only says that, if believer has *R* for a *non-trivial period of time*, then he will have the capacity to believe the single content that something red, and round is present *during that time*. This is not undermined by Sperling-type experiments.

Nor do I think any pathological cases (Block's GK, Nakamura and Mishkin's monkeys, visual agnosics) casts doubt on the grounding intuition; the interpretation of such cases is always up for grabs. So, not only does the grounding intuition have considerable pull; *biological theorists cannot find a reason, independent of their theory, for rejecting the grounding intuition*. This means that it can safely be used in an argument against such theories.

Experiences do not merely cause our beliefs; they justify those beliefs.[12] Imagine looking at the (real) tomato depicted in Figure 1. Now consider any possible situation in which you have the property of having an experience with this very phenomenal character – that is, the property *R*. Intuitively, in every such situation, provided that you have no reason to suspect that any "undercutting" defeaters obtain (e. g. that you are hallucinating), you have at least some *prima facie* justification for believing that a round object is before you.[13] This entails that *R* has the following (second-order) property necessarily:

> *The justification property:* being a property which is such that, if a believer has it in the absence of undercutting defeaters, then he has a justification for believing that a round thing is present.[14]

None of the externally-directed intuitions presupposes the controversial transparency thesis: for every experience, we know every element of its phenomenology by attending to external objects and properties, and we never attend to anything but external properties. So my argument has the advantage of being invulnerable to objections to this thesis.[15]

---

[12] See for instance McDowell 1994, Peacocke 2004, and Pryor 2000.

[13] Roughly, by an *undercutting defeater* here I mean a hypothesis which is such that, if one has a reason to suspect it, then one has reason to believe that one's having *R* is not good evidence for the proposition that a red and round object is present.

[14] *Objection:* The justification intuition is not consistent with *simple reliabilist theories*. On such a theory, if a brain in a vat might have *R*, then, even if it has no reason to suspect this, it does not thereby have *any* justification for believing that a round object is present (Goldman 1979).

Reply: It seems to me that in response we should reject simple reliabilism rather than the justification intuition. The intuition that such a brain in a vat would have a justification for believing that a round object is present is extremely strong. And most reliabilists themselves agree. In order to accommodate the intuition that the brain in a vat has a justification, they have devised more sophisticated forms of reliabilism, such as "normal worlds" reliabilism (Goldman 1986).

It is worth mentioning that, aside from extremely simple reliabilist theories that most would reject, the justification intuition is compatible with a wide variety of epistemic theories. One great divide is between liberal and conservative theories. (For liberalism, see Pryor 2000. For conservativism, see Wright 2002.) On liberal theories, *R* might necessarily provide believers with a *prima facie* justification for believing that a round object is present, even if they lack an antecedent justification to believe that things are as they appear. Thus, *R* might provide "immediate" justification for believing this proposition. This type of view is clearly compatible with the grounding intuition. By contrast, on conservative theories, *R* only provides believers with a *prima facie* justification for believing that a round object is present, if they have antecedent justification (or perhaps "warrant") for believing that things are as they appear. In this sense, *R* can only provide "mediate" justification for believing this proposition. This view, too, is compatible with the justification intuition, for conservatives often say that believers who lack a reason to doubt that things are as they appear automatically possess the required antecedent justification for believing that things are as they appear. (For the point that conservativism and the justification intuition are compatible, see Silins 2008.)

The justification intuition is also consistent with a variety of different theories of experience. It is often combined with an intentionalist theory. But it is also compatible with a disjunctivist theory. On some such theories, having *R* is a matter of being in a state that cannot be discriminated by reflection from seeing the redness$_P$ and roundness of something. And it is perhaps intelligible being in a state that cannot be discriminated by reflection from seeing the redness$_P$ and roundness of something might necessarily provide a justification for believing that a round object is present. Indeed, the justification intuition is consistent with indirect realist theories of experience and traditional foundationalism. On an indirect realist theory, having *R* is a matter of sensing the redness$_P$ and roundness of a mental object – a sense datum. This necessarily provides believers with a justification for believing that such an object is round, and hence for believing that *something* is round, in accordance with the justification intuition.

[15] To see this, consider some objections to transparency. First, the transparency observation faces a worry about what we might be attending to in hallucinatory cases (Pautz 2007). Meinongian objects? Uninstantiated universals? These options are not very plausible. The externally-directed intuitions have no commitment to the claim in hallucinatory cases we are attending to anything, so do not face this worry.

3 The Second Premise

Recall the template for the externally-directed argument against biological theories. The first premise is that the experience property *R* has externally-directed property *P* necessarily. The second premise is that no neural property *N* has externally-directed property *P* necessarily. I have argued that the first premise holds where *P* is the externality property, the matching property, the grounding property, and the justification property. In the present section I will argue that second premise holds in each of these cases. The argument will be based on certain hypothetical cases I call *separation cases*. So I call it the *separation argument*.

Before I present the argument, some background. Let *N* any neural property with which the biological theorist might identify *R*. Then the biological theorist can certainly say that *N* (that is, on his view, *R*) has the four externally-directed properties *in actual humans*. (This is why the Leibniz's Law argument must rely on the intuition that *R* has these *necessarily* while *N* does not.) In other words, *in fact*, if someone has *N*, then he has an experience as of a round object, is in a state that matches the world only if a round object is present, and so on. On standard physicalist theories, this is because *N* in fact plays a certain functional role. Here I understand *functional role* extremely broadly to include any facts that reach outside the head:

- facts about what types of behavior *N* is apt to produce (e. g. 'round-appropriate' behavior)
- facts about what external properties cause *N* under optimal condition (e. g. being round).
- facts about sensorimotor profiles

And so on.[16] Of course there are disputes about just which aspect of functional role determine the externally-directed properties. Evans famously suggested an *output-based* approach:

Second, the transparency observation is a *universal claim*. So it is open to apparent counterexamples involving blurry experiences, phosphenes, and other visual oddities. *The externally-directed intuitions, by contrast, are specific, single-case intuitions about R. The argument I will give do not require that they extend to all experience properties*. Third, the transparency thesis sometimes a displaced perception model of introspection, which is controversial and problematic. (For the claim that transparency renders introspection problematic, see Dretske 2003. For transparency as amounting to a displaced perception model, see Tye 2002. For problems with this model, see Lycan 1999, especially note 4.) By contrast, the externally-directed intuitions take on no commitments about introspection and so are not open to these problems.

[16] For these functionalist-externalist factors, see Tye 1995, Dretske 1995, and Noë 2004. In my broad sense 'functional', the externalist intentionalist theories of Tye and Dretske count as functionalist theories. Tye (2006) also understands functionalism in this broad manner. By appealing to functional role, I think that the biological theorist can answer Jackson's (1977) *common term* problem. The problem is why do we use the same terms (for instance 'red', 'round') to characterize experience that we use to characterize external objects. The biological theorist can answer that when we say that an experience is an experience as of a red and round object, we mean that it is an experience as of the phenomenal type typically produced by red and round objects. Similarly, biological theorists can solve Jackson's "many-property" problem – a problem that has recently been revisited by Byrne (forthcoming) and Tye (forthcoming). The problem is: how do 'has an experience as of a red object above a green one' and 'has an experience as of a green object above a red one' differ in their semantic values? The identity theorist can say that the first predicate expresses the property of having an experience that is normally caused by a red object above a green one, while the second predicate expresses the property of having an experience that is normally caused by a green object above a red one. Of course, the biological theorist also faces the

> the complex property of auditory input which codes the direction of the sound acquires a spatial *content* for an organism by being linked with behavioral output [under normal conditions]. (Evans 1985, 385) [17]

Others (for instance Tye 2000) hold that the spatial content of an inner state is determined by what optimally causes it on the input-side. But these philosophers agree that the externally-directed properties of a neural property *N* derive from functional role, broadly understood. For convenience, I will often put the standard view by saying that *N* has the externally-directed properties with respect to roundness by virtue of playing:

> *The round-role:* being apt to be caused by round object at viewer relative place *d* and *p* and apt to cause behavior that is "appropriate to" a round object at *d* and *p*.

Both *externalists and internalists about intentionality* would agree that *N* has the four externally-directed properties with respect to roundness only if it plays a certain functional role. The argument to be presented for premise 2 applies equally on 'internalism' about externally-directed intentionality. For standard internalism like that defended by Lewis and Jackson (and Shoemaker – see §7) is *qualified:* the idea is that *N* has the relevant content, only provided that it is in ensconced in a certain type of system, and certain laws of nature prevail, so that it plays a certain "narrow" functional role with respect to other internal states and behavioral outputs (narrowly characterized).[18]

---

question: what makes it the case that these experiences differ in phenomenal character? What is the ground of the ostensible difference in spatial arrangement? The biological theorist can answer that they differ in some neurobioloical respects that can only be uncovered empirically. So, although I think the biological theory faces decisive problems, the many-property problem is not among them.

[17] Likewise, discussing Macbeth and his hallucinatory dagger, Lewis (1983, note 2) writes "the right assignment of content to Macbeth's states will be the one given by the best general rule of assignment", which "will be the one that does the best at assigning contents that rationalize behavior, according to the principles of common sense psychology".

[18] For instance, the internalist David Lewis would say that in humans *N* has the proposition *there is a round thing present* as its "narrow content". He would agree with externalists that *N* has this content by virtue of playing a certain environment-involving functional role: very roughly, being such that under appropriate circumstances it would be caused by the presence of round things and would cause behavior appropriate to round things. What made him an internalist was that he thought that the relevant functional role does not bring in facts about the subject's actual environment or history. Further, he thought that the relevant functional role concerns *potential* interactions with the environment, so that a state of a brain in a vat might count as instantiating it. In consequence, Lewis held that the relevant input-output functional role of *N* in humans is determined by the characteristics of total internal system in which *N* is ensconced together with the laws of nature (as well as facts about 'the appropriate population'). In this sense, Lewis maintained that the relevant environment-involving functional role is *narrow* as opposed to *wide*. For the point that on Lewis-style functionalism content is determined by environment-involving functional role, and yields a qualified internalism about content, see Lewis 1994 and Jackson and Braddon-Mitchell 2007, 240-2. The qualified nature of internalism about content is implicit in the writings of other internalists. For instance, when Fodor was an internalist, he said that he had in mind internal duplicates *alike in their powers to produce behavior, non-intentionally characterized* (1987, chapter 2).

Davies (1993) and Fischer (2007) develop arguments against internalism about intentionality. But their arguments only succeed against a radical, unqualified form of internalism which rejects what I will call the *functionalist claim* below. It is unclear that anyone accepts this radical form of internalism about intentionality. For more on their arguments, see foot-note 26.

On standard physicalist theories, then, *N* has the four externally-directed properties with respect to roundness only if it plays a certain functional role, understood broadly as indicated above. Call this *the functionalist account of the externally-directed properties*. Even philosophers (like biological theorists) who are skeptical of functionalist theories of *experience* typically accept functionalist theories of the externally-directed properties.

Now, my case for premise 2 does not *depend on* the functionalist account. For premise 2 only says that *N* has the externally-directed properties contingently. This is true on any theory of the externally-directed properties. For instance, on a sense datum theory, *N* has the externally-directed properties contingently, because it contingently realizes the awareness of round sense datum, or a round visual field region (Peacocke). On a disjunctivist theory, *N* has the externally-directed properties contingently, because it contingently realizes seeing the roundness of an object (or being in a state indiscriminable by reflection from this). On a "primitivist" version of intentionalism about experience (Chalmers 2006), *N* has the externally-directed properties contingently, because it contingently realizes standing in a primitive intentional relation to a content involving roundness.

The separation argument is meant to provide a general argument, independent of consideration of any particular theory, for the conclusion that *N*, by contrast to *R*, has its externally-directed properties only contingently, in accordance with premise 2. It begins by pointing out that there are possible *separation cases* in which *N* is stripped of its relevant (wide *and narrow*) functional role, and more generally is stripped of all relevant relations roundness. Then it says that in these cases *N is also stripped of the four externally-directed properties*, thus establishing premise 2. This follows immediately from the standard functionalist account, but (to repeat) is plausible on any reasonable theory of the externally-directed properties.

In more detail, the argument is as follows. First, consider some separation cases:

> *Case 1: Slug* Imagine an alien creature, Slug, who has no visual system, but who has a taste system and an auditory system. Slug has the general capacity to have beliefs, for instance beliefs about tastes and sounds. In fact, he is very intelligent. Although Slug does not have a visual system, for some reason he does have whatever neuroanatomy is involved in having (possibly non-local) human neural property *N*, and indeed occasionally has *N*. Now, in actual humans, *N* is apt to be caused by a round object at *d* and *p* and apt to cause behavior appropriate to such an object. In other words, in actual humans, *N* realizes a certain functional property. But, we may suppose, this is not so in Slug. In Slug, *N* is entirely functionally isolated: it is not even *apt* to be caused by round objects or *apt* to cause round-appropriate behavior. (One might wonder how evolution or some other process might produce Slug, but this does not matter to my argument: all that matters is that Slug is possible.) Finally, imagine that the physical facts of the case are the only facts. If all physicalist-functionalist theories of the externally-directed properties are mistaken, and these properties are grounded in primitive relations to roundness, then Slug bears no such relations

> to roundness when he has *N*. For instance, he is not aware of a round sense datum or visual field region (Peacocke); nor does he bear a primitive 'sensory representation relation' to a content involving roundness.
>
> *Case 2: Blurg.* Blurg is the lone creature in some world. He has the general capacity for belief. Further, he has *N*. Whereas in the case of Slug *N* plays no interesting functional role, in the case of Blurg it does play any interesting functional role, only one quite different from the role it plays in actual humans, due to different wiring: it is apt to be caused by bodily damage and apt to cause avoidance-behavior. So when Blurg has *N* he bears no interesting physical or functional relations to roundness. In fact, we may suppose he never does so. Imagine that the physical facts of the case are the only facts.
>
> *Case 3: the simple system.* Imagine that we excise all the neuroanatomy from an individual's brain except what is necessary for the tokening of *N*, where *N* is the (local or global) neural property that the identity theorist would identify *R* with. Now imagine that, in a world containing no sentient species, this simple system forms by chance. In this world, *N* has no evolutionary history and the system does not belong to the species *homo sapiens* or any other species. Therefore, in this world, it does not have the function of indicating round objects, and it does not play any functional role among the members of any species. Imagine that the physical facts of the case are the only facts..[19]

Now recall (from §1) that the biological theorist will probably not say that *R* is necessarily identical with some very local neural property. Instead, he might say that *R* is necessarily identical with some more global neural property *N* involving activity in the visual cortex, re-entrant processing, and perhaps a bit of surrounding neural machinery. That means that *these cases might be somewhat ungainly;* but they are metaphysically possible, and this is all I need.

It is worth emphasizing that I am not supposing that in these cases *N* is *apt* to be caused by round object and *apt* to cause round-appropriate behavior, but fails to do so because of a statistical fluke (to use a distinction in Shoemaker 1981); rather, I am supposing that *N* is *not even apt* to play this causal role.

The separation argument for premise 2 of the externally-directed argument is (schematically) now as follows:

1 For any possible case *C*, if *N* has externally-directed property *P* with respect to a round thing at viewer-relative distance *d* and place *p* in case C, then

[19] Notice that I stipulated that both Slug and Blurg have the general capacity for belief. That is because the grounding property and the justification property are conditional properties of *N* of the form: if a *believer* has *N*, then the believer is such that *p*. Therefore, to show that *N* does not possess these properties necessarily, I must show that there are possible cases in which a *believer* has *N*, but in which the believer is not such that *p*. The cases of Slug and Blurg are such cases. By contrast, the externality property and the matching property are not defined in terms of believers. Therefore cases in which a strange non-believing system has *N* would suffice to show that *N* does not possess these properties necessarily.

there must be something that *makes it the case* that *N* has this externally-directed property *P* in case C. (Trivial)

2 There is *nothing that could make it the case* that *N* has externally-directed property *P* in possible separation cases. For in such cases *N* is stripped of the relevant functional role which respect to the environment and action; and in such cases there is *nothing else* that could make this the case.

3 Therefore *N* does not possess externally-directed property *P* in separation cases.

In short, any neural property *N* is akin to the sentence 'a round thing is present'. It might have "meant" something different or nothing at all. It this way it couldn't be more different from *R*, which has built-in external-directedness.

Note that my description of separation cases is neutral on the issue of whether the systems involved have *R* as well as *N*. The reason is that my purpose is merely to justify premise 2 of the Leibniz's Law arguments, asserting that *N* (unlike *R*) does not possess any of the four externally-directed properties necessarily. For this purpose, it is not required to build into the cases any assumption about whether or not Slug and Blurg and the simple system have *R* in addition to *N*.[20]

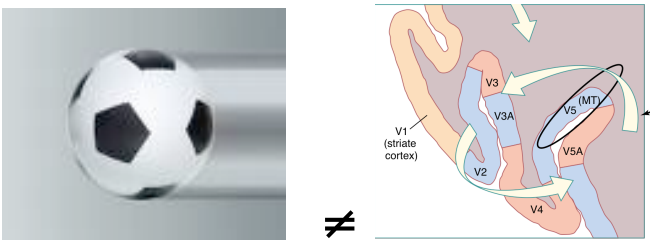
Of course, if one comes to the table defending a biological theory according to which *N* is necessarily identical with *R*, and concedes that *R* has certain externally-directed properties with respect to roundness necessarily, then one will of course want to say that *N* too necessarily has these externally-directed properties, even in separation cases. But wanting to say something is very different from being able to say it. If the biological theorist says that *even in separation cases N* is directed at a round thing at viewer relative distance *d* and place *p*, he owes us some plausible account of what could make this the case. In separation cases, what could constitute *N*'s presenting a round thing at distance *d* and position *p*, *rather than some other shape at some other distance and position?* There is no good answer to this question. But if the biological theorist does not answer this question then he has a *magical theory of intentionality* according to which *N* necessarily represents roundness at *d* and *p just by virtue of being the neural property that it is,* even though the connection between this neural property and roundness at *d* and *p* appears completely arbitrary.[21]

---

[20] From my point of view, it is good that, for the purposes of my argument, it is not required to build into the cases any assumption about whether or not Slug and Blurg and the simple system have *R* in addition to *N*. For it distinguishes the externally-directed argument against biological theories from a more suspect argument suggested by some remarks of Hilary Putman (1999), which is addressed in the following section.

[21] *Objection:* The biological theorist should accept premise 1 of the Leibniz's Law arguments: intuitively, *R* does have certain externally-directed properties necessarily. But the biological theorist reject premise 2. Sydney Shoemaker is not himself a biological theorist but he has an influential view on the nature of properties that one might think could help the biological theorist here. On Shoemaker's view, *N* has its conditional powers necessarily. So it has the same conditional powers in separation cases that it has in the actual world. Now in each separation case *N* (which, recall, includes a bit of surround) is in fact ensconced in a strange neural surround S in which it is not apt to be caused by round objects and is not apt to cause behavior appropriate to such objects. But, on Shoemaker's view, even in separation cases it still has the following conditional power: being such that, if it *were* ensconced in its actual surround A, and instantiated with other neural states *N1, N2, N3,* etc., then it would be apt to be caused by a round object at distance *d* and position *p* and apt to cause behavior appropriate to such an object. Could the biological theorist say

The four Leibniz's Law arguments are now complete. R is *necessarily* as of a round object at a certain viewer-relative distance *d* and position *p*. R *necessarily* matches the world only if a round object is present at *d* and *p*. By contrary, N is *not* necessarily as of a round object at *d* and *p*. R does *not* necessarily match the world only if a round object is present, as the separation argument show. So, by Leibniz's Law, R simply cannot be identical with N. Likewise, R necessarily gives believers the capacity to believe that a round object is present at *d* and *p*, and a justification for forming such a belief. But N does not. For Slug and are believers who have N, but they do not thereby have the capacity to form such a belief, nor do they have a justification for forming such a belief. Again, by Leibniz's Law, R simply cannot be identified with N.



My Leibniz's Law arguments generalize. Block (e. g. 2009, p. 1112) often illustrates the biological theory with motion experience. Thus, let L be the experience property you in fact get on viewing an object move from right to left. Block points out that evidence suggests that visual area MT+ is implicated in motion experience (see figure above taken from Block 2005). Block's view is that L is simply necessarily identical with M, where M is *something like* [MT+ activity plus recurrent loop plus thalamic switch]. But, L has a built-in connection to movement to the left: it is necessarily as of *movement to the left*. However, M (setting aside a magical theory) clearly is not. For, in possible separation cases, M is not connected with movement in any direction, and does not produce behavior appropriate to movement in any direction. So, by Leibniz's Law, M is simply the *wrong sort of property* with which to identify L.

I think my Leibniz's Law arguments also generalize to non-visual experiences. Let P, T, S be pain, taste, and sound experience properties, respectively. Maybe these experience properties, unlike R, do not necessarily present an *object*. But P and T necessarily possess analogous externally-

---

that *this* is what makes it the case that even in these strange scenarios *N* is as of round thing at distance *d* and position *p*, *rather than some other shape at some other distance and position?*

Reply: No. The reason is that in separation cases (indeed, on Shoemaker's view, in all cases) *N* has indefinitely many conditional powers of this kind. For instance, in separation cases, *N* has the following conditional power: if it *were* ensconced in surround *E*, and instantiated with other neural states *N1*, *N2*, *N3*, it would be apt to be caused by elephant-shapes. The claim that in separation cases the 'roundness-involving' conditional powers of *N*, as opposed to all of its other conditional powers, determines N's externally-directed properties in those cases would be totally arbitrary (for a similar point against the appeal to rigidification, see note 29). Indeed, for the functionalist, the only reasonable view is that in separation cases the externally-directed properties of *N* are determined by *its neural surround in those cases*. So, in the case of Slug and the simple system, it has *no* externally-directed properties; and in the case of Blurg it has externally-directed properties with respect to a *certain bodily property in a certain bodily region*. This agrees with premise 2 of my Leibniz's Law arguments.

directed properties with respect to certain *qualities* and a *place of the body* (in the case of *T*, the tongue). And *S* necessarily possesses these properties with respect to certain events, qualities like loudness and pitch, and temporal and spatial relations and properties. Since the connection between the experience properties and a certain ostensible bodily or external spatiotemporal region is necessary, but the connection between any neural property *N* and such a region is only contingent (deriving from *N*'s contingent relations to the region), these conscious properties cannot be necessarily identical with any neural property *N*.

These arguments show that a biological theory of experience is no more plausible than a purely biological theory of belief. Let *B* be the property of believing that a round thing is present. Many would say that *tokens* of *B* are neural *tokens*. And many would say that the property *B* is *realized by* some neural property *N*. But even biological theorists would say that *B* is not necessarily identical with any neural property *N*. While they would say that the experience property *R* is necessarily identical with a neural property, they would say that the intentional property *B* is necessarily identical with some functionalist-externalist property that is merely realized by a certain neural property. Why? One argument is that *B* necessarily possesses certain second-order *externally-directed properties*. For instance, necessarily, if one has this property, then one is in a state that is about a round object, and that "matches the world" only if a round object is present. This is obvious. But no neural property *N* has these externally-directed properties necessarily. On any theory of intentionality (even an 'internalist' theory), *N* has its externally-directed properties contingently, owing to its functional role with respect to other inner states, inputs, and behavioral outputs. So *B* cannot be necessarily identical with *N*. Rather, it must be identified with some "bigger" functionalist-externalist property which (like *B*) might possess the relevant externally-directed properties necessarily. But we have seen that, just like *B*, *R* has some externally-directed properties necessarily, so that *R* too cannot be identified with any mere neural property *N*.

In short, biological theorists fail to appreciate that some of the very same considerations that show that belief properties are not necessarily identical with neural properties also show that visual experience properties are not necessarily identical with neural properties. Since visual experience properties have built-in external-directedness, but neural properties do not, *neural properties are just not the sort of things visual experience properties could be.*

4 Putnam on Petri Dishes: A Simpler Argument?

Before I consider objections, I want to distance my argument from an argument suggested by Hilary Putman. Putnam has written:

> If . . . we say that the modules for visual "appearances" are in the visual cortex . . . . . then we run up against the fact that parts of the visual cortex . . . can be *dissociated* from the "speech areas". Are we to say that in such a case . . . there are visual sense data of which the person is not aware? And what would happen if our technology advanced to the point at which we could remove the module involved in the visual recognition of, say, chairs from the brain and keep it alive and functioning in a vat [or a Petri dish - AP] . . . Would one then have chair sense data . . .? This way madness lies. (1999: 30-31)

Although Putnam is not explicit, this passage suggests what we might call the *functionalist argument* against neurobiological theories. (I do not mean to suggest that Putnam now accepts functionalism: he does not. But a natural way of developing his remarks depends on a premise that would be accepted by functionalists.) Let the *R-functional properties* be the functional properties that typically accompany *R*, such as being a state that is apt to be caused round objects and apt to cause speech-behavior and other behavior appropriate to round object. And let a *strange system* be any system that has *N* but fails to have any state with the *R*-functional properties.[22] Then the functionalist argument might be put as follows:

1 There are possible strange systems that have *N* but do not have any state with the *R*-functional properties.
2 Intuitively, no such strange system has *R*.
3 Therefore *R* is not identical with *N*

The second premise is based on what might be called the *functionalist intuition:* the alleged intuition that having *R* requires being in a state with the *R*-functional properties. Common sense or analytic functionalists about experience like Lewis, Shoemaker and Jackson would presumably endorse this intuition and this argument.[23]

The reason we must address the functionalist argument here is that it might be thought that the externally-directed argument is in effect equivalent to it. It is important to realize that this is not the case. There are several problems with the functionalist intuition on which the functionalist argument relies. In view of these problems biological theorists might simply reject 2 and say that strange systems could have *R*. This also *seems to be* how Polger (2004, chapter 8) responds to Putnam's argument. (He says he "might not be able to rule out such a possibility" (p. 242) but does not outright say it is a possibility.) Contrary to Lewis, Shoemaker, Jackson and other analytic functionalists, I do not think that an anti-functionalist biological theory can be ruled out on the basis of reflection on our concepts. Fortunately, the externally-directed argument differs from the functionalist argument and is not open to any of these problems. This requires some explanation.

To begin with, let me explain the problems with the functionalist intuition. The first problem is simply that on reflection it is not very compelling. In fact, the implausibility of functionalist-externalist theories is part of the case for biological theories. (See for instance Hill's 1991, pp. 61-5, 75 *absent role argument*, which is endorsed by McLaughlin 2003, p. 181.) The *input-oriented* functional role of *R* does not seem a necessary feature of *R*: it does not seem to be a necessary truth that *R* is apt to be caused by round

[22] Previously, I stipulated that *separated systems* bear *no* relations – physical or non-physical – to roundness. This is not a stipulated feature of *strange systems*. It is only stipulated that they do not bear interesting *physical-functional* relations to these properties; they are allowed to bear *non-physical* relations to them. This will be important below when I use "primitivist intentionalism" to show that the externally-directed argument is independent of the functionalist argument.

[23] Another argument depends on the claim that if a strange system had *R* then it would not be attached to a "self", together with the intuition that *R* must be attached to a "self". (For responses to this general worry from biological theorists, see Polger 2004 and Block 2008.) I think that this argument is problematic, and I will not discuss it here.

objects.[24] Likewise, many will say that the *output-oriented* functional role of *R* does not seem to be necessary. It does not seem necessary that experience is apt to cause *beliefs* (about those experiences or the world) or *behavior*. The case of animals suggests that experience can be severed from belief. And since Putnam's attack on behaviorism, it has been standard to say that experience properties like *R* are only *contingently* connected with behavioral dispositions on the output side. This intuition finds expression in Strawson's case of the weather watchers, who have experiences but are constitutionally unable to act on them.[25] Given his well-known anti-functionalist view that 'phenomenal consciousness' is totally separable from 'access consciousness', Ned Block is committed to the possibility of cases in which experience is severed from both cognitive access and behavior.

Considering these points, one can get into a mood in which one thinks that perhaps *R* could be *entirely* stripped of its actual typical functional role: it might fail even to be *apt* to be caused by round things and *apt* to cause round-appropriate belief and behavior. In that case, contrary to the functionalist intuition, strange systems could have *R*.

Second, philosophers often say that in general there are no *a priori* links between phenomenal concepts and non-phenomenal concepts. But if one accepts this, then one must allow that there is no *a priori* guarantee that having *R* requires having the *R*-functional properties, and hence no *a priori* guarantee that strange systems could not have *R*.

Third, recall that the biological theorist will probably hold that *R* is necessarily identical with a *quite global* neural property *N*. On this view, strange systems that have the relevant neural property *N will have quite a bit of neural machinery*. Once this is realized, the intuition that they could not have *R* begins to fade.

Fourth, to some extent, our intuition that strange systems cannot have *R* might derive from a general mistaken intuition that mere patterns of neuronal firing cannot realize experiences. When we consider such systems, we only think of such patterns: we do not think of a human face or the other things we associate with consciousness. And we think: those could not realize consciousness! Leibniz's thought-experiment in which we are shrunk and look at the inner workings of a brain has a similar effect. But, of course, we know that this intuition is mistaken from our own case. In some sense, it is uncontroversial that neural patterns realize experiences.[26]

---

[24] Granted, there are *theories* of experience on which this is a necessary truth – for instance, Tye's (2000) intentionalist theory as supplemented by a tracking theory of sensory content. My point is that it does not *seem* necessary on the basis of pretheoretical intuition.

[25] Strawson 1994, chapter 9.

[26] Jackson (1993) offers a two-premise argument against biological theories of experience that is similar to the functionalist argument. The first premise is that *N* might play radically different functional roles in two individuals belonging to different kinds. An actual human and Slug would be an example of such a pair of individuals. The second premise is the broadly functionalist intuition that such individuals would not have the same experiences; in particular, they would not both have *R*, contrary to the neurobiological theory. The argument here, which concerns the possibility of the same neural properties playing different functional roles, is simply the reverse of the multiple realizability argument against type-type neurobiological theories of experience, which concerns the possibility of different neural properties playing the same functional role. (Davies (1993) and Fischer (2007) develop a similar argument against internalism about *intentionality*. But their argument only succeeds against a radical form of internalism about intentionality that lacks the usual functionalist qualifications mentioned previously (in §3 and footnote 19), and it is unclear that anyone accepts this radical form of internalism about intentionality.) This

In contrast to the functionalist argument, my Leibniz's Law argument based on external-directedness nowhere use its disputable premise 2. Nor do they use the functionalist "intuition" offered in favor of premise 2 that having *R* entails having a state with the *R*-functional properties. In particular, as already noted in the previous section, my description of separation cases is entirely *neutral* concerning whether the individuals in separation cases have *R*. Instead, my arguments only use externally-directed intuitions to the effect that having *R* entails having a state with certain externally-directed properties. These intuitions are not subject to the problems I have raised for functionalist intuition. For instance, maybe intuition does not rule out the possibility of Strawson's weather watchers having *R* without having a state with the *R*-functional properties. But, intuitively, if they did have *R*, they would necessarily thereby have an experience as of a round thing at *d* and *p*.

This is enough to show that my Leibniz's Law arguments based on the externally-directed intuitions avoid the problems with the functionalist argument. But we may say something stronger. It is not only the case that my Leibniz's Law arguments do not *use* the functionalist intuition. In addition, they are compatible with the *falsity* of this intuition. The reason is simple. The externally-directed intuitions only say that having *R* entails being in a state with the externally-directed properties. The functionalist intuition says that having *R* entails being in a state with the *R*-functional properties. These are independent. Granted, standard theories endorse what I called (in the previous section) the *functionalist account* of the externally-directed properties: having the externally-directed properties requires having the *R*-functional properties. On such an account, the externally-directed intuitions entail the truth of the functionalist intuition. But other theories would deny any such account. According to them, the externally-directed intuitions might be true even if the functionalist intuition is false.

For instance, friends of the *phenomenal intentionality program* adopt the broadly intentionalist theory that phenomenology is inextricably linked with intentionality but reject functionalist-externalist of phenomenology. Instead they favor internalist versions of intentionalism.[27] One possible version goes as follows. First, *R* is identical with standing a primitive "sensory representation" relation to a content involving the co-instantiation of $\text{redness}_{\text{p}}$ and roundness at viewer-relative distance *d* and place *p*. Second, it is contingent law of dualistic psychophysics that, if a system has *N*, then it stands in this primitive relation to $\text{redness}_{\text{p}}$ and roundness, and thereby has *R*.

On this view, if a strange system *S* had *N*, then *S* would also have *R*, without having a state with the *R*-functional properties. On this view, then, the functionalist intuition driving the functionalist argument is mistaken. Proponents of this view do not heed Putnam's warning 'this way madness lies'.

---

type of argument against biological theories of experience is vulnerable to the first two problems I raised for the functionalist argument.

[27] See Chalmers (2006), Horgan and Tienson (2002), Loar (2003), Pautz (2006). I do not mean to suggest that any of these philosophers accept the very radical version of internalism that I discuss here; I discuss this radical version only for the purposes of making vivid the independence of the externally-directed intuitions from the functionalist intuition.

But they do *not* violate the externally-directed intuitions. (That, I think, would be real madness.) Thanks to the fact that *R* essentially involves sensorily representing a content concerning the co-instantiation of redness_p and roundness at viewer-relative distance *d* and place *p*, *R* has the externally-directed properties necessarily, even in the strange system *S*, in agreement with intuition.[28] Further, even on this view, my Leibniz's Law arguments based on external-directedness succeed. By contrast to *R*, the underlying neural property *N* with which it is contingently associated does not have the externally-directed properties necessarily, for the separation argument shows that in other possible scenarios it does not ground any intentional relations to roundness at *d* and *p*. So, even on this view, *R* is not identical with *N*. Rather, it is identical with the emergent relational property of sensorily representing a content involving the co-instantiation of redness_p and roundness at viewer-relative distance *d* and place *p*, a property which *does* have the relevant externally-directed properties necessarily.

5 Objections: Big States, Magical Theories, and Brains in Vats

So, my Leibniz's Law arguments against biological theories of experience differ from the simpler functionalist argument. And they avoid objections to the functionalist argument. They are, however, open to other objections which must be addressed.

Recall the template of my Leibniz's Law arguments:

1 Experience property *R* has externally-directed property *P* with respect to roundness at *d* and *p necessarily*
2 No neural property *N* has externally-directed property *P* necessarily
3 Therefore *R* is not identical with any neural property *N*

The objections I will consider fall into two sorts. Some reject premise 2, claiming that some neural properties do have the relevant externally-directed properties necessarily. The first three objections I will consider are of this kind. I will argue that these objections either effectively amount to functionalist theories which forsake the biological theory, or else require a "magical theory" of intentionality. The remaining two objections use bizarre scenarios to cast doubt on premise 1. I argue that they do not cast doubt on 1.

I put each objection in the mouth of an imaginary biological theorist. I relegate technical objections (involving rigidification[29] and applying Leibniz's Law in modal contexts[30]) to foot-notes.

---

[28] If the grounding intuition is true, and if the relevant strange system counts as having the general capacity for belief, then in this strange system *R* will also ground the capacity to have *beliefs* involving roundness. The content of his belief would derive from the content of its experience, rather than its relations to external environment. In other words, given the grounding intuition, the failure of functionalist-externalist theories of *R* brings in its wake the failure of functionalist-externalist theories of belief. Similarly, many would say that, thanks to having an experience of the blue sky, one of Strawson's (1994) weather watchers could count as believing that a blue field is present, despite being unable to act on the world and lacking normal connections to the environment.

[29] *Objection:* To dodge this argument for the conclusion that *N* only has its externally-directed properties contingently, the biological theorist might appeal to *rigidified accounts* of these properties. For instance, in the case of the externality property, he might say that an experiential state is an experience as of a red_p and round thing in a world *W* just in case it is the kind of state that is typically brought about by a red_p and round object *in the actual world* (rather than in world *W*). Then he can say that *N* has the externality

First Objection: Big States. The biological theorist should grant premise 1. *R* does necessarily possess certain externally-directed properties with respect

---

property in every possible scenario - even separation scenarios. For, according to him, in separation scenarios, indeed in all scenarios, *N* is identical with an experiential state, namely *R*, that is normally brought about by a red_P and round object *in the actual world*.

Reply: But the rigidification ploy fails for two reasons. First, on the biological theory, combined with a rigidified causal account of the externality property, it might *actually* be the case that we separated systems (as it might be, brains in vats), and hence *R* might actually fail to be an experience as of a red_P and round thing. As against this, intuitively, *however* the actual world might turn out to be, *R* is an experience of a red_P and round thing. In the terminology of Evans, this is a *deep necessity* (Evans 1979). Second, there are numerous additional problems with extending the rigidification ploy to the other externally-directed properties. (i) It seems counterintuitive and arbitrary to suppose that the matching-facts, belief-facts and justification-facts concerning individuals in *other worlds* depend on facts about *our world*. Should they not depend on facts about their worlds? And, if they do depend on facts about some other world, why our world as opposed to any other? (ii) Rigidified accounts have the chauvinistic consequence that unactualized, non-human brain states could not realize experiences with the externally-directed properties, because such states do not play any causal-functional role *in actual humans*. (iii) Presumably, individuals in other worlds could have beliefs about the externally-directed properties of their own experiences. But, on rigidified accounts of these properties, it is hard to see how they might, since this would require that they have beliefs whose truth-conditions involve our world, which might be quite remote from their worlds. For these reasons, we must reject the idea that *N* might have the externally-directed properties *in separation scenarios* owing to the fact that it plays a certain causal-functional role in *our world*. Therefore the original conclusion stands: by contrast to *R*, *N* can stripped of its externally-directed properties in certain possible scenarios.

[30] *Objection:* Since premises 1 and 2 of the externally-directed argument are *de re* modal claims rather than *de dicto* ones, applying Leibniz's Law may *appear* legitimate. But there is an analysis of *de re* modality on which applying Leibniz's Law is not even legitimate in the *de re* context. Even though properties like *R* and *N* are repeatables that might exist in different worlds, it is possible (though perhaps not very natural) to provide a *counterpart analysis* of *de re* modal statements about them. Say a property *Y* instantiated at another world *W* is a *functional counterpart* of a property *X* instantiated at our world iff *Y* plays roughly the same functional role in *W* that *X* plays in the actual world. Say that *Y* is a *neural counterpart* of *X* iff *Y* has the same neural characteristics as *X*. On a counterpart analysis, the premises of the externally-directed argument might come out true and the conclusion false, if '*R*' and '*N*' co-refer, but '*R*' somehow evokes the functional counterpart relation and '*N*' evokes the neural counterpart relation. (For more on why under a counterpart analysis Leibniz's Law might fail even in *de re* modal contexts, see Lewis 1971. However, as we saw in note 5, Lewis himself would have agreed with the conclusion of the externally-directed argument: for he would have said that 'the property of having *R*' refers to a *functional property* of a certain kind, rather than to *N*.)

Replies: I myself think that this kind of counterpart analysis is deeply implausible, so I think that the externally-directed argument it is perfectly valid as I have formulated. (For one thing, since '*R*' is a technical term, it is unclear how it might evoke a "functional counterpart relation". In this regard, it is unlike (for instance) 'the statue', which is part of ordinary language and so might be conventionally associated with a certain type of counterpart relation, a statue-counterpart relation.)

In any case, the issue is really a red herring, for two reasons. (i) There is an alternative version of the externally-directed argument that does not invoke Leibniz's Law and that is therefore invulnerable to the present worry. The first premise is that, by the separation argument, separated systems have *N* but neither *N* nor any other of their states has the externally-directed properties. The second premise is that, by the externally-directed intuitions, if an individual has *R*, then the individual is in a state with the externally-directed properties. From these two premises it follows that separated systems have *N* but not *R*, contravening that neurobiological theorist's claim that these properties are *necessarily* identical and coextensive. (ii) Indeed, since the biological theory maintains that *R* and *N* are *necessarily* identical and coextensive, any counterpart analysis on which '*R*' and '*N*' somehow evoke different counterpart relations would itself *directly* entail that the biological theory is false, because it would entail that *R* is only *contingently* identical with *N*, and that they are only *contingently* coextensive. In fact, the proffered counterpart analysis entails a *functionalist theory* of *R*, according to which an individual at a world *W* has *R* iff he has a property that is a functional counterpart of *R*, that is, iff he state that plays a certain *functional role*! So to adopt it would be throw in the towel.

to roundness at distance *d* and place *p*. But he should reject premise 2, and claim that some neural property also has the externally-directed properties necessarily. Indeed, he can claim this, while accepting the standard *functionalist account* of the externally-directed properties.

Granted, if we are dealing with a *fairly local* neural property, separation cases show that it is not metaphysically necessary that it is apt to be caused by a round object at *d* and *p* and apt to cause behavior appropriate to such an object. So, by contrast to R, any fairly local neural property does not have the externally-directed properties necessarily, on a functionalist account of them. In other words, for fairly local neural properties, premise 2 is true, ruling them out as candidates to be identified with R.

But perhaps the biological theorist can identify R with a very "big" neural property N. Further, perhaps some such N will necessarily play the "round-role", and hence necessarily have the externally-directed properties with respect to roundness, just as R does. In other words, perhaps for some suitably big N, separation cases in which it fails to play the round-role are *metaphysically impossible*. For such a N, there can be no Leibniz's Law objection to the claim that R is necessarily identical with N, since both have the externally-directed properties necessarily. Call this the *big state theory*.

Reply. Before I say why this objection is unavailable to the biological theorist, let me make some remarks about what the big state theory would have to look like. The big state theory is underdescribed. It comes in different versions. In one version, R is identical with N, where N is the conjunctive property *having local neural property L and being so constituted, and living under such laws of nature, that L is apt to be caused by a round thing at distance d and place p and is apt to cause behavior appropriate to such a thing*. This has the form of what Shoemaker (2007, 21; 1981) calls a *total realizer*. On a functionalist account of the externally-directed properties, this conjunctive property is indeed necessarily directed at roundness, just as R is. So it is fit for identification with R.

In another version of the big state theory, R is necessarily identical with N, where N is the conjunctive property *having local neural property L and having a very large internal neural surround S*. Further, on the second version, the laws of nature are metaphysically necessary, so that N necessarily plays the round-role, and hence (given the functionalist account) necessarily possesses the externally-directed properties.

It is worth mentioning that there are serious problems with this second version of the big state theory. I have assumed throughout that the biological theorist will say that R is necessarily identical with N, where N is the conjunctive property *having local neural property L and having *some* neural surround S* In order to guarantee that N necessarily plays the round-role, the big state theory we are now considering would presumably have to hold that the 'surround' part of N is *extremely large*, and *specifies the presence of the input and output systems and perhaps memory system* (Shoemaker 2007, 123). But since one could presumably have R without these systems (e. g. one could hallucinate after a terrible accident that removed one's motor output system), the claim of the big state theory that R is necessarily identical

with *N* now becomes very implausible. For this and other[31] reasons, I think that the first version of the big state theory, based on Shoemaker's idea of a total realizer, is superior.

Now I come to my main point. It is not important which version of the big state theory is best. The important point is that, in either version, the big state theory is very much like functionalism. Granted, in one regard it differs from standard functionalism. On a standard functionalist theory, in order to have *R*, you just need *some or other property* that plays the round-role. By contrast, on this version of the big state theory, you need in addition to have a specific neural property, namely *L*. In other words, the big state theory rejects multiple realizability. But in another respect the big state theory is like functionalism, for it entails that, necessarily, if one has *R*, then one has a property that plays the round-role.

This leads to two points. (i) *The big state theory, in either version, is not a version of the biological theory that was my target.* If the biological theorist accepts this view, he is in effect giving up the biological theory. (ii) Of course, this is a terminological point. But it is also the case that the big state theory described in the objection, entailing as it does that *R* necessarily plays the round-role, is inconsistent with what biological theorists explicitly say. As we saw in the previous section, Block, Hill, McLaughlin and Polger all assert the possibility of "absent role" cases in which *R* does not play the round-role.

So, for these philosophers, accepting the big state theory would amount to change of view. Another point is that this type of theory is implausible. Indeed it is implausible for reasons independent of the possibility of 'absent-role' cases, as we shall see in connection with the third objection.

Second Objection: A Magical Theory of Intentionality? The biological theorist should grant premise 1. *R* does necessarily possess certain externally-directed properties with respect to roundness at distance *d* and place *p*. But he should reject premise 2, and claim that some neural property *N* also has the externally-directed properties necessarily.

Granted, if he does this by providing a functionalist account of the externally-directed properties and accepting the big state theory, his theory becomes inconsistent with his anti-functionalism. But he can avoid inconsistency by saying that *N* has the externally-directed properties necessarily while rejecting the standard functionalist account of the externally-directed properties. He can say that *R* is necessarily identical with *N* and it is metaphysically necessary that *N* results in "sensorily representing" a content involving roundness at *d* and *p*, *even in separation cases in which it is not apt to be caused by a round thing at d and p nor apt to cause behavior appropriate to such a thing*. And he can say that this is why *N*, like *R*, has

[31] (i) The second version of the big state theory requires necessitarianism about laws, which of course is very controversial. (ii) Even if necessitarianism about laws is right, it is unclear that any neural property *N*, however big, should be necessarily apt to be caused by a round thing at *d* and *p* and apt to produce behavior appropriate to such a thing, because even on necessitarianism the functional role of *N* will still depend on background conditions. For instance, consider a scenario in which *N* is instantiated in a population, but the population is constantly being bombarded by 'pulsars' from outer space (Fischer 2007), in such a way that *N* is not apt to cause, or be caused by, anything. In such a case, *N* will fail to have the externally-directed properties with respect to roundness.

certain externally-directed properties necessarily. This view combines the biological theory with a kind of minimal intentionalism.

This view goes beyond any usual internalism. Typical internalism about intentionality endorses the functional account and so is *qualified:* N has the relevant content involving roundness, only provided that it is in ensconced in a certain type of system, and certain laws of nature prevail, so that it plays the round-role (something that is not true in the "separation cases"). As we saw (§3), this is so on Lewis's internalism about intentionality. And, as we shall see (§7), this is so on Shoemaker's internalism about intentionality as well. On this qualified form of internalism, my premise 2 is true. By contrast, what is now being proposed is a radical *unqualified* form of internalism.

Reply: Unqualified internalism amounts to a *magical theory of intentionality*. It is like saying the sequence of marks 'a round thing is present' is somehow necessarily about a round thing, regardless of how it used in the language. This may seem too radical to be worth taking seriously. But at least one well-known philosopher appears to accepts unqualified internalism. I have in mind John Searle. As is well known, he rejects functionalism about intentionality. He holds that the neural fixes sensory intentionality independently of functional role.

There are two reasons why the biological theorist cannot accept this way out of my Leibniz's Law arguments. First, recall my separation argument. In separation cases, what could make it the case that N represents a round thing at distance d and position p, *rather than any other property at some other distance and position?* Second, unqualified internalism recognizes the existence of a *two-place sensory representation relation* between individuals (or their states) and external properties (or contents involving such properties). Further, on unqualified internalism, it is metaphysically necessary that, if one has a certain neural property, then one stands in this relation to a certain set of external properties, regardless of the functional role of the neural property. Now, on standard theories, the sensory representation relation is identical with some kind of relation involving functional role: an 'indicator relation', or a 'tracking relation' (more on this in the next section). But of course these theories yield externalism about sensory intentionality and are inconsistent with unqualified internalism. In general, there is arguably no naturalistic account of the sensory representation relation compatible with unqualified internalism.[32] This

[32] One response the magical internalist might offer appeals to an analogy with the relation *x has mass-in-grams y*, a relation between objects and numbers. There is a reductive account of this relation in terms of conventions and the isomorphism between the structure of numbers and the structure of masses. This account has the consequence that having a certain non-relational mass property necessarily results in standing in the mass-in-grams relation to a certain number. The magical internalism might hope that there is an analogous reductive theory of the sensory representation relation, one that is compatible with his claim that merely having N entails representing the instantiation of redness_p and roundness at a certain viewer-relative place.

The trouble with this response is that no such analogous radically internalist reductive account of the sensory representation relation is possible. In some cases, there is modest congruence between our neural states and sensory quality. But this is not enough for a reductive account of the sensory representation relation that is compatible with magical internalism. Such an account would require a single, general modality-independent algorithm for going from intrinsic neural properties to what properties we sensorily represent. There is no such thing. For instance, there is no single algorithm for going from intrinsic neural properties to what shapes and other spatial properties we sensorily represent, just as there is no algorithm

means that the proponent of unqualified internalism would have to take this relation to be a *primitive relation* which is somehow necessarily connected with internal neural states. Perhaps this view is open to John Searle, who has no reductive aspirations. But it is not open to the biological theorist, who wants a reductive theory of consciousness.

Third Objection: A Mixed Theory. (An objection along the following lines was suggested to me by Ned Block.) The biological theorist should grant premise 1. *R* does necessarily possess certain externally-directed properties with respect to *primary qualities* like (apparent) roundness. In fact, he should also grant that *R* necessarily possesses certain externally-directed properties with respect to *secondary qualities* like redness$_P$ (as briefly discussed at the start of §2). But this does nothing to cast doubt on the biological theorist's claim that *R* is necessarily identical with some neural property *N*. For the biological theorist should reject premise 2, and claim that some neural property *N* also has these externally-directed properties necessarily.

This need not amount to the "magical theory of intentionality" contemplated in the previous objection. The biological theorist can provide a demystifying account of how *N* necessarily represents certain primary qualities, and a different demystifying account of how it necessarily represents the secondary quality redness$_P$.

In particular, he can provide a *functionalist account* of how *N* necessarily represents (apparent) roundness. On this account, by contrast to the magical theory of intentionality previously discussed, it is not the case that *N* necessarily represents roundness at *d* and *p regardless of its functional role*. Rather, the biological theorist should concede a point to functionalists: *R is necessarily connected with the round-role*. Since the biological theorist holds that *R* is necessarily identical with *N*, he should also say that *N* is necessarily connected with the round-role.[33] In this way, given a functionalist account of the externally-directed properties, *N* (and hence *R*) necessarily represents roundness at *d* and *p*. This is not a magical theory of intentionality, because *N's* necessarily representing roundness at *d* and *p* is grounded in its necessary connection with the round-role. On this view, separation cases fail to show that *N* only contingently represents roundness at *d* and *p*. The reason is that, since on this view *N* is necessarily connected with the round-role, there are difficulties concerning whether separation cases are even metaphysically possible.

---

for going from the shapes of words in a language to what they represent. *Again, the only reasonable reductive accounts of the sensory representation relation are the standard functionalist-externalist accounts, which are unavailable to the unqualified internalist.*

[33] What is the 'round-role'? What functional role might *R* be necessarily connected with? In his comments in which he raised the objection now under discussion (according to which *R* necessarily plays a certain functional role), Block pointed out to me a passage his from his (1995) to his BBS article in which he implies that having R 'allows one to see' that round shapes are 'not packable' and 'have a large number of axes of symmetry'. (He is replying to Katz.) This suggests that the functional role which he thinks necessarily belongs to R is *allowing one to see these things*. But this conflicts with Block's well-known view that phenomenology is separable from cognitive access. Further, the claim is problematic. When my 3 year-old daughter has *R* she is not allowed to see these things. So I am not sure what functional role Block thinks necessarily belongs to *R* (and hence *N*). I will assume it is not something so intellectually demanding: something like being apt to be caused by a round thing at d and p, and being apt to cause behavior appropriate to such a thing.

Now, for reasons discussed in connection with the first objection, it seems that *N* could necessarily connected with the round-role just in case it is a "big state". So this account of how *N* necessarily represents roundness at *d* and *p* seems to amount to one of the two versions of the big state theory discussed in connection with the first objection. The best version of this theory said that *R* is necessarily identical with *N*, where *N* is the conjunctive property *having neural property L and being so constituted that (given the laws of nature) L is apt to be caused by a round thing at distance d and place p and is apt to cause behavior appropriate to such a thing*.

The biological theorist can provide a different, *non-functionalist account* of how *N* (and hence *R*) necessarily represents the redness_p at *d* and *p*, where redness_p is taken to be a Shoemakerian appearance property. On this account, *N* (and hence *R*) *does* necessarily represent redness_p at *d* and *p regardless of its functional role with respect to the environment and color-related behavior*. Hence, it represents redness_p at *d* and *p* even in separation cases in which it is not apt (even potentially) to produce color-related discriminatory or sorting behavior. (As we shall see in §7, since his view of color experience has functionalist elements, Shoemaker himself denies this.) Now, this also does not have to amount to a magical theory of intentionality. For the biological theorist can provide a two-part answer to the question: "what makes it the case that *N* necessarily represents redness_p at *d* and *p,* even in separation cases in which it is stripped of its functional role with respect to color-related behavior?" The two-part answer is as follows:

(i) *N* is necessarily identical with *R*, and
(ii) *R* necessarily represents redness_p at *d* and *p*.

Reply. The mixed theory described in the objection provides an account of shape experience that is very much like functionalism. In fact, it requires the 'big state' theory discussed in connection with the first objection. This leads me to reiterate points I made in connection with the first objection. (i) *This is not the biological theory that was my target.* As discussed in connection with the first objection, this is basically a kind of functionalism. So if the biological theorist accepts this view, he is in effect giving up the biological theory. (ii) Of course, this is a terminological point. But it is also the case that the mixed theory described in the objection, entailing as it does that *R* necessarily plays the round-role, is inconsistent with what biological theorists explicitly say. As we saw in the previous section, Block, Hill, McLaughlin and Polger all assert the possibility of "absent role" cases in which *R* does not play the round-role.

These are my main points. But even if for these philosophers accepting the mixed theory would involve a change of view, it is worth considering whether it is a plausible theory of sensory consciousness and its intentionality. I think it is not.

To begin with, the objection does not answer the sort of concerns I raised about 'magical theories of intentionality' in response to the previous objection. It provides an inadequate answer to the question "what makes it the case that *N* necessarily represents redness_p at *d* and *p,* even in separation cases in which it is stripped of its functional role with respect to color-

related behavior?". To see this, consider an analogy. Suppose we are Platonists about numbers, and we are wondering what makes it the case that Mark Twain is thinking of the number 197 rather than some other number. It is no answer to say that he is Samuel Clemens, and Samuel Clemens is thinking of the number 197 rather than some other number.

In addition, the objection fails to provide a reductive account of the two-place sensory representation relation. To suppose that we must provide a reductive account of monadic properties like R, and but do not have to provide a reductive account of relations like the sensory representation relation, would be to accept an unjustified double standard. If the proponent of the mixed theory does not provide a reductive account of this relation, he is in effect maintaining that this is a primitive relation. And that is inconsistent with his reductive aspirations.

In fairness, I think it is possible to construct a kind of disjunctive reductive theory of the sensory representation relation that is consistent with the mixed theory described in the objection. The trouble is that it implausible.

Let us first see how such a disjunctive theory might go. According to the objection, N necessarily represents red_p. How is this? One idea is that redness_p is itself definable in terms of N: it is something like the property *normally causing* N. And, in general, *what it is* to sensorily represent such a 'secondary quality' is just to have the neural property which that secondary quality normally manifests in. This account would explain why N necessarily represents redness_p at d and p, even in separation cases in which it is stripped of its functional role with respect to color-related behavior.[34] The objection provides a different, more functionalist account of the representation of primary qualities. On this account, N necessarily represents roundness at d and p by virtue of necessarily playing a functional role, presumably something along the lines of being apt to produce round-appropriate behavior. Now the relevant functional roles are not *identical with* primary qualities like roundness. So, to get a reductive account of the sensory representation relation, we need some kind of mapping from them onto the primary qualities represented. Maybe there is some kind of mapping, g, from the functional roles onto experiences of primary qualities. And maybe one might say that what it is to sensorily represent a primary quality P is just to have a neural property such that g maps the functional role of that neural property onto the experience of P. The relevant mapping might be provided by 'common sense psychology' which allegedly associates every primary quality experience with a distinct functional role (more on this issue below).

Putting all this together, the 'mixed theory' of sensory representation suggested in the objection might naturally be combined with the following *disjunctive theory* of the sensory representation relation:

> Individual X sensorily represents Y iff (Y is a secondary quality and X has the neural property that Y normally manifests in) *or* (Y is a primary quality and X has an appropriate neural property such that

[34] I propose and then criticize this idea in Pautz (forthcoming). https://webspace.utexas.edu/arp424/www/simple.pdf

> *g* maps the functional role of that neural property onto the experience of *Y*).

The trouble, as I said, is that is that this theory is implausible. First, it is implausible just by virtue of being disjunctive. Second, there are problems with specifying the mapping function *g*. David Lewis, one of the main proponents of a functionalist account of intentional relations, holds that a brain state with a certain functional role is mapped onto an experience as of an *F* just in case the "best interpretation" of the population assigns to the brain state the experience as of an *F*, based on its functional role and certain constitutive constraints on interpretation. But there are problems about 'deviant interpretations'. And what does it even mean to say one interpretation is 'best'?[35] Third, this account implies that we sensorily represent secondary qualities like $\text{redness}_{\text{p}}$ *simpliciter*. In fact, we represent them at certain viewer-relative distances and places. It is difficult to see how the first disjunct might be revised in order to accommodate this.

I suppose the proponent of the mixed theory might not accept my gift of a 'disjunctive account' of the sensory representation relation. But then he must provide some other account of this relation. And there are a number of problems with the mixed theory that arise even if this disjunctive account is rejected. First, on the mixed theory, *R* and hence *N* plays the round-role. As already noted, various absent-role cases appear to be possible, and it is because biological theorists accept the possibility of such cases that they cannot accept the mixed theory. Second, the mixed theory requires a very strong claim. Let *E1*, *E2*, *E3*, . . . be all possible experiences that represent different primary qualities: for instance, shapes, distances, orientations, velocities, and so on. Let *B1*, *B2*, *B3*, . . . be the 'big states' with which the proponent of the mixed theory identifies them. On the mixed theory, *E1*, *E2*, *E3*, . . . and hence *B1*, *B2*, *B3*, . . . , *necessarily* represent the relevant different primary qualities. To avoid a magical theory of intentionality, the proponent of the mixed theory says that these states all necessarily play *different* functional roles *F1*, *F2*, *F3*, . . . with respect to the environment and action, which somehow ground (presumably via some kind of mapping) their representing different primary qualities. The trouble is that it is very implausible that each and every possible experience representing a different set of primary qualities plays a different functional role which belongs to it is necessarily. For instance, is it really plausible that each possible experience of a different velocity plays a different functional role which belongs to it necessarily?

Fourth Objection: Strange Cases According to the first three objections, the biological theorist should accept premise 1 of the Leibniz's Law arguments but reject premise 2. Granted, this strategy is not promising. But there is another strategy available. The biological theorist should instead accept 2 but reject 1. In accordance with 2, he should accept that, for any neural property *N* that is a reasonable candidate for identification with *R*, there are possible strange cases in which *N* is stripped of its externally-directed properties. But, contrary to 1, the same is true of *R*, he should insist, so there

[35] See Pautz (forthcoming). https://webspace.utexas.edu/arp424/www/simple.pdf

is no problem with identifying *R* with *N*. In a brain in a vat scenario, or an inversion scenario, *R* loses its actual externally-directed properties. For instance, in the brain in the vat scenario, an individual has *R*, but *R* does not count as an experience as of a round thing at distance *d* and place *p*, because in that scenario *R* is *not normally caused by* round objects at *d* and *p*. In an inversion scenario, an individual has *R*, but *R* is normally caused by (e. g.) squares. In that case, again, *R* does not count as an experience as of a round object, but as an experience as of a square object.

Reply: There are two possible replies.

First, many who accept the externally-directed intuitions will be *tracking intentionalists* about *R* who will deny that these scenarios are even metaphysically possible (see the next section).

Second, even if we accept that these cases (on some ways of elaborating them) are possible, we can still say that *R* retains the externally-directed properties in them. The case for thinking *R* loses the externally-directed properties in such cases relies on a controversial, simple pure input-based causal theory of the externally-directed properties on which (for instance) *R* is an experience as of a round thing only if it is normally caused by red and round things. But, as we saw in §3, Evans and others have proposed more output-based accounts. If we accept a more output-based theory, we can say that even in these scenarios *R* counts as an experience as of round thing, on the grounds that having *R* involves being in a state that would lead to actions appropriate to a round object at *d* and *p*. (In the case of the brain in the vat, *R* would lead to such actions were the brain embodied.)[36] Alternatively, we might say that even in these scenarios *R* counts as an experience as of a round thing, on the grounds that having *R* involves bearing a *primitive* intentional relation to roundness.[37]

It seems to me that accepting one or another of these replies is much more plausible than rejecting the externally-directed intuitions.

Fifth Objection: Modus Tollens. Granted, brains in vats and inversion scenarios do not provide decisive reasons to reject the externally-directed intuitions. But the biological theorist might simply reject these intuitions on the basis of his theory. He might say that the best overall theory has it that *R* is necessarily identical with some internal neural property *N*. Since *N* has its externally-directed properties only contingently, so does *R*. For instance, in separation cases, individuals have *N*. If *N* is a 'global' neural property, such cases will be ungainly; but they will still be possible. Hence, on the biological theory, such individuals have *R*, in spite of their different functional organizations. But, when they have *R* (that is, *N*), they do not have an experience as of a round object at *d* and *p*; they are not in a state that matches the world only if a round object is present at *d* and *p*; they do not have the capacity to have beliefs that are true on this condition; and they do not have a justification for accepting such beliefs. For in separation cases there is nothing that could make these things the case – just as the separation argument in §3 says.

[36] For this view, see Lewis 1994.
[37] Chalmers 2006.

Reply: This objection fails for two reasons. First, imagine looking at the tomato depicted in Figure 1. According to the objection, an individual could have an experience *just like this* (that is, have *R*) *without having an experience as of a (roughly) round thing*. Indeed, according to this objection, in the Blurg case in §3, when Blurg has *R*, *R* counts as an experience *as of bodily damage in his foot!* But we must be honest. This is just not plausible. Evidently, *R* has built-in directedness towards roundness. Since no neural property *N* has built-in directedness towards roundness, *R* is simply not a neural property.

Second, I disagree when the objector says that the biological theory is the best overall theory. As we shall see next, we can and should adopt an alternative theory of visual phenomenology which easily accommodates the externally-directed intuitions.

6 How Might Phenomenal Consciousness be Essentially Externally-Directed?

I have argued that the externally-directed intuitions rule out the biological theory. However, one might wonder just how a theory might accommodate these intuitions. Isn't experience an interior phenomenon? How then can it be 'essentially externally-directed'? I think that the best way of accommodating the externally-directed intuitions is to accept an intentional theory that forsakes the biological theory.

On an intentional theory, *R* is necessarily identical with the relational property of standing in the "sensory representation relation" to the co-instantiation of redness_p and roundness at *d* and *p*. In general, every experience property is a matter of standing in the sensory representation relation to a certain content.

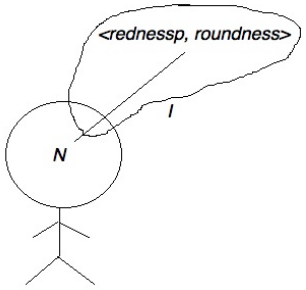
The intentional theory comes in a bewildering variety of forms. I will focus on Michael Tye's (2000) well-known version, which I will call *tracking intentionalism*. It has attracted a lot of interest, because it provides what may be the simplest route to developing intentionalism in a reductive form. For reasons that I will explain in the next section, I believe that there is a serious problem with this type of theory. I focus on it now only purposes of explaining how the intentional theory might accommodate the externally-directed intuitions.

Tracking intentionalism has two parts. The first is response-independent reductionism about 'secondary qualities'. So, for instance, redness_p is a reflectance property; felt pains are types of bodily disturbance; smells are chemical properties and so on. The second part is a tracking theory of the "sensory representation relation". On this theory, the sensory representation relation is identical with:

> *The tracking relation: individual X* is some state or other that is poised to ground the formation of beliefs and desire and that would be cause by the instantiation of cluster-of-properties *Y* were optimal conditions to obtain.

Previously, we saw that standard functionalist theories of sensory intentionality can be either input-oriented or output-oriented (§3). Tracking intentionalism is obviously input-oriented.

On tracking intentionalism, then, *R* is identical with the relational property of standing in the "sensory representation relation" to the co-instantiation of redness_p and roundness at *d* and *p*. And this property in turn reduces to the relational property of standing in the tracking relation to the property of being round and having so-and-so reflectance at *d* and *p*. Call this *I*. See the illuminating figure below:



Unlike the biological theory, the tracking theory is not subject to my Leibniz's Law arguments. The biological theorist's claim that *R* is necessarily identical with *N* is subject to my Leibniz's Law argument, because it is impossible to see how the inner neural property *N* might have the externally-directed properties necessarily. It is especially obvious that *N* cannot have the externally-directed properties necessarily on a standard functionalist account of these properties. For *N* does not necessarily play the role of being apt to track a round object at *d* and *p* and being apt to produce behavior appropriate to such an object, as witness separation cases. The tracking theorist accepts the functionalist account of the externally-directed properties. But he avoids my arguments, because, instead of identifying *R* with *N*, he identifies *R* with the bigger, world-involving property *I* which *builds in* the kind of functional role which might necessarily ground the externally-directed properties. For *I*, by contrast to *N*, *separation cases are not possible*. Since having *I* essentially involves tracking the instantiation of roundness at *d* and *p*, it is essentially directed at roundness at *d* and *p*, just as *R* is. Evidently, *I* necessarily possesses the externality property and the matching property. And *I* trivially necessarily possesses the grounding property, because it is partly *defined* in terms of having a state that is "poised to ground the formation of beliefs and desire" with the appropriate content. Finally, on a suitable epistemological story (e. g. a dogmatist (Pryor 2000) story), *I* necessarily possesses the justification property. So, unlike *N*, *I* might have all four externally-directed properties necessarily, just as *R* does. In other words, *R* and *I* *mesh* in their essential externally-directed features. This is why, in contrast to *N*, the relational property *E* is a potential candidate for identification with *R*.

Here is an analogy. The content-carrier 'something round is present' is only contingently directed at roundness. But the "bigger" property of being a sentence that is used to track the presence of a round thing is necessarily

directed at roundness. This property builds in the relevant functional role. Likewise, the content-carrier *N* is only contingently directed at roundness. So *N is just not the sort of property with which R might be identified*. But the "bigger" property *I* – roughly the property of being in a state that tracks roundness – is necessarily directed at roundness. This property builds in the relevant functional role. So *I* is fit for identification with *R*.

I have focused on tracking intentionalism for the purposes of illustration. But I think that all version of intentionalism accommodate the externally-directed intuitions – with the exception of certain *Fregean versions* of intentionalism of the kind defended by Brad Thompson and (the erstwhile) David Chalmers.[38] On all such theories, *R* is identical with an intentional property which involves a 'built-in' intentional relation to roundness, and which therefore has certain externally-directed properties necessarily. This is even so on 'primitivist' versions of intentionalism which reject the functionalist account of the externally-directed properties, as we saw at the close of our discussion of Putnam's Petri dish case.

Now I do not mean to suggest that intentional theories are the only theories consistent with the externally-directed intuitions. Appealing to intentional content is one way to explain external-directedness, but not the only way. I also think sense datum theories, disjunctive theories, and sensorimotor theories are consistent with the externally-directed intuitions (cf. note 7). All of these theories hold that phenomenal consciousness extends beyond the brain. However, I think that these theories fails on other grounds.

Therefore, the externally-directed intuitions provide the starting point for an argument for an intentional theory, one different from the much-criticized 'transparency' argument.

7 An Overlooked Puzzle: How Might Phenomenal Consciousness be Both Externally-Directed and Internally-Dependent?

I have argued that biological theories fail to accommodate the essential "external directedness" of phenomenal consciousness. But I also think that they contain an element of truth: phenomenal consciousness is indeed "internally-dependent", in a sense to be explained. In closing, I would like to briefly describe an overlooked puzzle about how consciousness might be both externally-directed and internally-dependent. To illustrate the puzzle,

[38] On one version of *Fregean intentionalism* (Chalmers 2004, Thompson forthcoming), *R* is identical with sensorily representing a content of the form *something has the property normally causing color experience C and the property normally causing shape experience S*. On this version of Fregean intentionalism, the externally-directed intuitions are false. For instance, on this version of Fregean intentionalism, if one is a brain in a vat receiving artificial stimulation from a computer, then the descriptive concept *the property normally causing shape experience S* picks out a certain computational property of the relevant computer program, or something along these lines. Therefore, on this version of Fregean intentionalism, in this scenario, one has *R*, but one's experience does not match the world only if something *round* is present; rather, it matches the world only if something is present with the relevant *computational property*. For this reason, we should reject this version of Fregean intentionalism. (Chalmers (2006) now rejects this version of Fregean intentionalism on the basis of considerations within the same vicinity.) I do, however, think that the externally-directed intuition are quite consistent with *another* version of Fregean intentionalism: one which maintains that the content of *R* is built up out of Fregean concepts, but which maintains that these are *non-descriptive, rigid* concepts that pick out redness$_P$ and roundness *in all possible scenarios*.

I will assume that external-directedness is best accommodated by an intentional theory.

If we adopt the tracking intentionalism discussed in the previous section, we accommodate external-directedness. But tracking intentionalism does not accommodate internal-dependence. To show this, I must of course first say what I mean by 'internal-dependence'.[39]

Decades of psychophysical research have shown that phenomenal character is poorly correlated with the external properties tracked, even under optimal conditions: the reflectances of objects, chemical properties of foodstuff, types of bodily disturbance. By contrast, neuroscience has shown that phenomenology is often very well correlated with internal factors: opponent channel states of the color system, across-fiber patterns in the taste system, somatosensory firing rates.

Now two creatures from different species might optimally track the very same reflectance properties of objects under optimal conditions, but differ in downstream opponent processing and in color-related behavior. Similarly, if a foodstuff is poisonous to one creature but an important food source to another, then they might optimally track the same chemical property of the foodstuff, but differ in profoundly in their inner "across-fiber patterns" and their innate behavioral responses to the foodstuff. And the same bodily disturbance might normally produce radically different somatosensory firing rates and behavioral responses in different creatures. Call these *coincidental variation cases*, because in them the properties optimally tracked coincide but there is also variation in neural processing and functional organization.

To say that experience is *internally-dependent* is merely to say that in *some* cases of this kind the individuals involved have different experiences, in spite of the fact that they optimally track the same external physical properties. The case for this claim is based on the fact that those individuals exhibit vast neural differences (differences which in actual creatures are very well correlated with phenomenal differences) as well as sensorimotor differences. Note that internal-dependence differs from internalism, because it only requires that internal factors play *some* role in determining phenomenology.

In short: biological theories evidently accommodate internal-dependence. But they do not accommodate external-dependence. By contrast, standard, tracking forms of intentionalism err on the other side. They accommodate external-dependence, but they do not accommodate internal-dependence. This empirical argument against tracking intentionalism seems to me more convincing than the usual arguments about inverted earth, inverted spectrum, and shift spectrum, which have not caused tracking intentionalists to budge.[40] So what we need is a middle way.

---

[39] I develop the argument below, as well as other arguments against tracking intentionalism, in Pautz (forthcoming). https://webspace.utexas.edu/arp424/www/rest.pdf

[40] For inverted earth, see Block 1990. For the alleged possibility of inversion without misrepresentation, see Shoemaker 1994. My own view is that spectrum inversion without misrepresentation is intuitively *impossible*, since the inverted color experiences intuitively would have incompatible contents. So I do not myself endorse this argument against tracking intentionalism. For shifted spectra, see Block 1999 and Shoemaker 2007, p. 126. In these cases, due to individual differences, two experiences (e. g. pain or color experiences) are caused by the same stimulus. Elsewhere (Pautz forthcoming a, sect. 2 and sect. 4) I suggest that such a case is no problem for the tracking intentionalist. He can say that, while the actual

We need a theory that accommodates both external-directedness and internal-dependence.

Now I can state the puzzle. On intentionalism, we accommodate external-directedness by saying that experience involves the two-place sensory representation relation directed at external properties:

Individual X sensorily represents property Y

(Some would say that the sensory representation holds in the first instance, not between individuals and properties, but between individuals' inner states and contents involving properties. But the puzzle I will describe arises either way.)

If we also wish to accommodate internal-dependence, we must say that the individuals in coincidental variation case bear sensory representation relation to *different* external properties (phenomenal colors, felt pains, tastes) due to internal neural-functional differences, even though they bear the tracking relation, and other naturalistic relations, to the same external properties. The puzzle is: *what account of the sensory representation relation might make it intelligible how this is so?*

We have no ready model here. Our standard reductive theories of intentional relations identify such relations with tracking relations, indicator relations, asymmetric dependence relations, and so on. But all of these relations are held constant in coincidental variation cases. So, given internal-dependence, the sensory representation relation cannot be identified with any of them. Indeed, we already encountered the puzzle of how the sensory representation relation might be internally-dependent in connection with the second objection ("magical theory") and the third objection (Block's "mixed theory") in §5.

To illustrate the puzzle, consider Sydney Shoemaker's subtle and influential theory of phenomenal consciousness (1994, 2006, 2007). Although his arguments are somewhat different than mine (they are based on transparency plus the intuitive possibility of spectrum inversion without misrepresentation), Shoemaker has long defended a theory that is meant to accommodate both internal-dependence and external-directedness. Briefly, as I understand it, the theory goes as follows.

Shoemaker holds that phenomenal similarity is functionally definable. Likewise he holds that what it is to have some experience or other is functionally definable. But, as he believes in the possibility of spectrum inversion, he does not believe that individual experience properties are completely functionally definable in the usual Ramsey-Lewis way. For instance, very roughly, on his view, *R* is identical with having a certain type of physical property that plays a certain kind of functional role with respect to belief and behavior. (In Shoemaker's scheme, this is to say that *R*'s necessary causal profile includes (i) causal features that can be captured in a particular kind of Ramsey sentence for *R and* (ii) causal relations to

---

cause of the two experiences is the same, their optimal causes (and hence the represented properties) are distinct but overlapping ranges of physical properties. Thus the phenomenal difference is grounded in a representational difference, without making either experience non-veridical. The tracking intentionalist cannot likewise reply to my hypothetical coincidental variation cases, because I stipulate complete coincidence in the physical properties optimally tracked.

particular physical properties which restrict the instantiation of *R* to creatures with particular physical natures.) Shoemaker also holds that having *R* necessarily involves standing in the "sensory representation relation" to a content involving the co-instantiation of roundness and redness_p at *d* and *p*. Thus, Shoemaker would presumably agree with premise 2 of my Leibniz's Law argument as regards roundness. A neural property *N* realizes *R*, and hence realizes the sensory representation of roundness at *d* and *p*, only if it plays a certain functional role, involving effects on belief and behavior ('round-appropriate' behavior). Since it is contingent that *N* plays the relevant functional role (it depends on details of surrounding wiring, for instance), it is only contingent that *N* realizes the sensory representation of roundness at *d* and *p*. *Hence, on Shoemaker's view, a neural property N will have the externally-directed properties with respect to roundness only contingently – just as premise 2 says*. (Unless, of course, *N* is a 'total realizer' in Shoemaker's sense. But, as we saw in connection with the first objection in §5, this does not matter to my argument against biological theories. For, where *N* is a total realizer, the view that *R* is necessarily identical with *N* is not a version of the biological theory.) Finally, Shoemaker would identify red_p with an aspect of objective properties, where the having of an aspect by an objective property consists in its being such as to cause experiences of certain sorts in creatures with a certain sort of perceptual system (2007, 127). Shoemaker has not yet presented a view on the nature of the represented primary qualities like roundness.

Now there is, it seems to me, one thing missing from Shoemaker's theory: an account of the two-place sensory representation relation. As far as I know, Shoemaker's only remark on this issue is as follows:

> In virtue of what does an experience having a quale represent an object as having a particular [response-dependent] property? It cannot do so in virtue of a causal relation between the experience and the property it represents-one cannot say that the causing of A by B is the (or a) cause of A (here I am indebted to David Robb). . . I have no fully satisfactory answer to this question (which is hardly surprising, given that no one has a fully satisfactory account of how any experience has the representational content it has) (1994, note 7).

Shoemaker has not addressed this issue in subsequent work. But if we are worried about how monadic mental properties are realized by physical properties, shouldn't we be worried about how dyadic intentional relations are realized by dyadic physical relations? Otherwise, it seems, we are operating with an unjustified double standard for properties and relations. Indeed, this was a theme of Hartry Field's well-known paper 'Mental Representation'. Among other things, Field pointed out that it would not be right to say that we do not need a reductive account of intentional relations just because we bear them to *abstracta*. After all, objects bear the mass-in-grams relation to *abstracta* (numbers), but we still expect an account of this relation in physical terms (and measurement theory provides such an account). Indeed, some – including Shoemaker (2007, 2) – define Physicalism about the mind as the thesis that mental properties are physically realized. But presumably mental properties include polyadic mental properties, that is, mental relations. So on this definition of

Physicalism, Physicalism is false if there is no dyadic physical relation that realizes (or is identical with) the dyadic sensory representation relation.

Here is an idea that I think is worth pursuing. We just admit that there is *no* dyadic physical relation that realizes, or is identical with, the dyadic sensory representation relation. In that sense, it is a 'primitive relation'. This would presumably require giving up Physicalism, if Physicalism is defined in terms of realization. But it would be compatible with Physicalism, if Physicalism is a mere supervenience thesis. (Think of Moore's theory of goodness as primitive but supervenient.) For instance, one might say that the "big" monadic relational property *standing in the sensory representation relation to redness$_P$ and roundness* is realized by some physical property P, even if the dyadic sensory representation relation which is a *component* of this relational property is not realized by any dyadic physical relation. Then one could say that, because the individuals in the "coincidental variation" cases described above have relevantly different physical properties, they bear the sensory representation relation to different sensible properties. In this way, both internal-dependence and external-directedness are accommodated. The main problem here is that, if we accept this "primitivism" about the sensory representation relation, supervenience appears to amount to a kind of objectionable brute emergence.

The puzzle of how sensory intentionality might be internally-sensitive as well as externally-directed, then, is a serious one. It must be confronted by any complete theory of consciousness.

References

Bickle, J. (2003) *Philosophy and Neuroscience: A Ruthlessly Reductive Account*. Dordrecht: Kluwer Academic Publishers.

Block, N. 1978 "Troubles with Functionalism", in C. Savage (ed.) *Minnesota Studies in the Philosophy of Science*, Vol. IX: 261-325.

Block, N. 1995 BBS Replies

Block, N. (2007) "Consciousness, Accessibility, and the Mesh between Psychology and Neuroscience", *Behavioral and Brain Sciences* 30: 481–99.

Block, N. (forthcoming) "Comparing the Major Theories of Consciousness", in M. Gazzaniga (ed.) *The Cognitive Neurosciences IV*. Cambridge: MIT Press.

Block, N. and Fodor, J. (1972) "What Psychological States are Not", *The Philosophical Review* LXXXI: 159-181.

Braddon-Mitchell, D. and F. Jackson (2007) *Philosophy of Mind and Cognition*. Oxford: Blackwell.

Chalmers, D. (1997) *The Conscious Mind.* Oxford: Oxford University Press.

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

Chalmers, D. (2006) "Perception and the Fall from Eden", in T. Szabo Gendler and J. Hawthorne (eds.) *Perceptual Experience*. Oxford: Oxford University Press.

Churclhand. P. M. (2005) "Chimerical Colors: Some Phenomenological Predictions from Cognitive Neuroscience", *Philosophical Psychology*18.

Churchland, P. S. (1986) *Neurophilosophy.* Cambridge: MIT Press

Crick and Koch (1990), Crick, F., and C. Koch. 1990. "Towards a Neurobiological Theory of Consciousness", *Seminars in the Neurosciences* 2: 263–75.

Dretske, F. (1995) *Naturalizing the Mind.* Cambridge: MIT Press.

Driver, J., and P. Vuilleumier. (2001) "Perceptual Awareness and Its Loss in Unilateral Neglect and Extinction", *Cognition* 79: 39–88.

Evans, G. (1982) *Varieties of Reference*. Oxford: Oxford University Press.

Farah. M. (2004) *Visual Agnosia, 2nd Edition.* Cambridge: MIT Press.

Fisher, J. (2007). "Why Nothing Mental is Just in the Head", *Nous* 41:318-334.

Flanagan, O. (1992) *Consciousness Reconsidered.* Cambridge: MIT Press.

Flanaga, O. and Polger, T. (1999) "Natural Answers to Natural Questions", in V. Hardcastle (ed.) *Where Biology Meets Philosophy,* Cambridge: MIT Press.

Goldman A. (1979) "What is Justified Belief?", in Pappas (ed.) *Knowledge and Justification.* Dordrecht: Reidel Publishing Company: 1–24.

Goldman, A. (1986) *Epistemology and Cognition.* Cambridge: Harvard University Press.
Hardin, C. L. (1987) "Qualia and Materialism: Closing the Explanatory Gap", *Philosophy and Phenomenological Research* 48: 281-298.
Hill, C. (1991) *Sensations: A Defense of Type Materialism.* Cambridge: Cambridge University Press.
Jackson, F. (1977) *Perception: A Representative Theory* Cambridge: Cambridge University Press.
Jackson, F. (1993) "Review of Christopher S. Hill, *Sensations: A Defense of Type Materialism*", *Philosophical Review* 102: 614-16.
Jackson, F. (2004) "Representation and Experience", in H. Clapin, P. Slezack and P. Staines (eds.) *Representation in Mind: New Approaches to Mental Representation*. Elsevier.
Kanwisher, N. (2001) "Neural Events and Perceptual Awareness", *Cognition* 79: 89–113.
Lewis, D. (1968) "Counterpart Theory and Quantified Modal Logic", *Journal of Philosophy* 65: 113-26
Lewis, D. (1994) "Reduction of Mind", in S. Guttenplan (ed.) *A Companion to the Philosophy of Mind*. Oxford: Blackwell: 412-31.
Lycan, W. (1996) *Consciousness and Experience,* Cambridge: MIT Press.
Martin, M. (2004) "The Limits of Self-Awareness", *Philosophical Studies,* 120: 37-89.
McLaughlin, B. (2007) "Type Materialism for Phenomenal Consciousness", in M. Velmans and S. Schneider (eds.), *The Blackwell Companion to Consciousness*. Oxford: Blackwell Publishing.
Peacocke, C. (forthcoming) "Sensational Properties: Theses to Accept and Theses to Reject", in a special issue of the *Revue Internationale de Philosophie*.
Polger. T. (2004) *Natural Minds.* Cambridge: MIT Press.
Polger, T. and O. Flanagan. (1999) "Natural Answers to Natural Questions", in V. Hardcastle (ed.) *Where Biology Meets Psychology: Philosophical Essays.* Cambridge, MA: The MIT Press.
Pautz, A. (2007) "Intentionalism and Perceptual Presence", *Philosophical Perspectives* 21: 495-541.
Pautz, A. (forthcoming) A Simple View of Consciousness, available online. In The *Waning of Materialism*
Pautz, A. (forthcoming) Do Theories of Consciousness Rest on a Mistake? Available online In *Phil Issues 20*
Polger, T. and K. Sufka. (2006) "Closing the Gap on Pain: Mechanism, Theory, and Fit", in M. Aydede (ed.) *New Essays on the Nature of Pain and the Methodology of its Study*. Cambridge: MIT Press.
Pryor, J. (2000) "The Skeptic and the Dogmatist", *Nous* 34: 517-549.
Putnam, H. (1967) "Psychological Predicates", in W. Capitan and D. Merrill (eds.) *Art, Mind, and Religion*. Pittsburgh: University of Pittsburgh Press.
Putnam, H. (1999) *The Threefold Cord: Mind, Body, and World*. New York: Columbia University Press.
Searle, J. (1982) *Intentionality: An Essay in the Philosophy of Mind*. Cambridge: Cambridge University Press.
Shoemaker, S. (1994) "Phenomenal Character", *Nous* 28: 21-38.
Shoemaker, S. 1981 Varieties of Functionalism
Shoemaker, S. Ways Things Appear
Shoemaker, S. (2007) *Physical Realization.* Oxford: Oxford University Press.
Smith, D., John, S. and Boughter, J. (2000), "Neuronal Cell Types and Taste Quality Coding", *Physiology and Behavior* 69: 77-85.
Sperling, G. (1960) "The Information Available in Brief Visual Presentations", *Psychological Monographs: General and Applied* 74: 1-30.
Strawson, G. (1994) *Mental Reality.* Cambridge: MIT Press.
Tye, M. (1995) *Ten Problems of Consciousness.* Cambridge: MIT Press.
Tye, M. (2000) *Consciousness, Color and Content.* Cambridge: MIT Press.
White, R. (2006) "Problems for Dogmatism", *Philosophical Studies* 131: 525–57.