

Brains in Vats and the Spatial Content of Experience: Response to Sydney Shoemaker

Adam Pautz

University of Texas, Austin

I am very grateful to Sydney Shoemaker for taking the time to write his comments, which will be invaluable to me in revising the paper. Shoemaker helpfully suggests that my “externally-directed” argument against biological theories be slightly modified in the light of brains in vats (BIVs). I was pleased to find that he “leans towards” accepting the modified version. (Although, given his intentionalism, I would have thought Shoemaker is outright committed to some version of the argument – more on this under pt. 2 below.) My version relies on a number of intuitions, but here I will simplify, and say it rests on the *roundness claim*: R necessarily represents roundness at p. (Contrary to what Shoemaker implies, I do not say R necessarily represents redness.) Shoemaker replaces this with the *determinable claim* (1 below) and the rest of the argument proceeds as before:

1 Experience property R necessarily represents d-roundness at p, where d-roundness is a determinable with ordinary roundness and roundness* (the relevant computer state in the BIV scenario) as determinates

2 Biological property N doesn't necessarily represent d-roundness at p

C So R is not N

Shoemaker accepts 2 on the basis of the same sort of considerations I develop in the paper and says he “leans towards” 1. Now if either my original version or Shoemaker’s helpful alternative version is correct, then my general strategy for refuting biological theories (which is distinct from the usual multiple realization argument) is vindicated. But I think the issue of which is to be preferred raises interesting questions about shape representation (about which Shoemaker has not said much in his work, having focused mainly on color representation) and the physical basis of consciousness. For this reason, in a future version of the paper, I would like to provide some reasons to think the roundness claim is to preferred to the determinable claim. I will try them out below (under 1). Then I will and then raise some general

queries (2-3). Finally (4) I will answer a question Shoemaker asked: why is the multiple realization argument not enough to refute the biological theory? I will also offer some remarks about the possibility of indeterminacy in our concepts of experience and about the 'primitivist' view I favor. I would be grateful for any feedback on any of these matters.

1. Do BIVs Undermine the Roundness Claim?

First, a clarification. Contrary to Shoemaker, my main response to the BIV argument against the roundness claim was not that in every possible BIV scenario R's representing roundness is constituted by the truth of a counterfactual along the lines of: if the BIV were "appropriately" connected to its environment, then the relevant state would be caused by round objects and cause roundness-appropriate behavior. Although I should have been more clear, I only meant that this is *one* possible response to *some* elaborations of the BIV. In fact, in the paper, I express skepticism about all such reductive theories of sensory intentionality. And I agree with Shoemaker that in any case this response may not apply to some particularly bad-off BIVs (e. g. ones with no evolutionary history, no human input-output systems, etc.). In what follows, for the sake of argument, I will assume that we are dealing with such a bad off BIV (= a brain in a vat to which the counterfactual strategy does not apply).

As I said in the paper, my response to the BIV argument against the roundness claim is that rejecting one of its premises is more reasonable than rejecting the roundness claim. To see this, let me first reiterate how plausible the shape claim is, and then turn to the premises of the BIV argument.

Imagine looking at a tomato. On Shoemaker's suggestion, if a BIV has that very experience-type, R, then its experience might be wholly veridical, even though no round, tomato-like thing is before the BIV. On his suggestion, in the BIV, R merely accurately represents the computer state, roundness*. (Or perhaps the suggestion is that in the first instance R represents d-roundness, which in this case is realized by roundness*.) Against this, it is intuitively necessary that, if an individual (even a BIV) has R, then that individual is in a state that matches the world only if a round (not merely d-round, not merely round*) object is before one: an object whose edges are in the ordinary sense roughly equidistant from a common point. (Such an individual could learn Euclidean Geometry, not Euclidean Geometry*.) In

accordance with this, most everyone would agree that, *if* a BIV might have R, then its experience is at some level non-veridical, since no such object is before him.

Now, Shoemaker grants the pull of the intuition but objects that it makes a questionable use of the first person perspective, and that it at best shows that the roundness claim applies to human beings. I have two replies. (i) The intuition is that any individual x whatever (whether human or no) that has R represents roundness at p. (ii) Everyone must grant that the first person perspective can be used to justify *some* unrestricted modal claims about experience properties: for instance, necessarily, if x has a bluish experience, a purplish one, and a greenish one, then x's first experience resembles his second more than his third. Indeed, Shoemaker himself seems to grant this, since, on the basis of the transparency observation, he has long held that (as he puts it in his recent 'A Case for Qualia, at p. 325) "each quale has, constitutively and independently of context, a representational content of its own" (and presumably not just in humans but in all possible individuals). He also "leans towards" the determinable claim (understood as a thesis applying to all individuals, not just humans), so he must think it enjoys some justification. Presumably, the first person perspective has something to do with that justification. Why then cannot the first person perspective be used to provide an initial, defeasible justification for my shape claim?

It seems to me that, by contrast to the shape claim, the premises of Shoemaker's BIV argument against it do not enjoy very strong support. The first (implicit) premise is that a bad off BIV might have R. The second is that in the BIV R represents the computer state roundness*, not roundness. Accordingly, a BIV who has R might enjoy a perfectly accurate experience, even if no round thing is before him.

An aside: Shoemaker's main point is that the BIV's *beliefs* represent roundness*, not roundness. Even if this is right, it does not without some argument undermine my shape claim that in the BIV the *experience property* R represents roundness, so that it is inaccurate at some level. Likewise, it does not undermine my externality and matching intuitions, since these only concern the representational properties of the experience property R. In fact, in his 'Matrix' paper and elsewhere, David Chalmers combines something like Shoemaker's claim about belief with my roundness claim about (the Edenic content of) experience. For reasons I would be interested to hear, Shoemaker himself says he is "not sure whether [R in the BIV] would count as

having the matching property [and hence count as inaccurately representing roundness]”. So it looks like he is actually *open to* something in the vicinity of my roundness claim. However, since at the start of his comments Shoemaker says BIVs undermine all my claims, I will provisionally assume that Shoemaker’s suggestion is that in the BIV R (not just the BIV’s beliefs) accurately represents the computer state roundness*, rather than inaccurately representing roundness.

What might be the argument for the first premise, that a bad off BIV might have R? After all, many externalists-functionalists would deny that this is metaphysically possible. Shoemaker does not provide an argument in his comments. One argument is that it is conceivable, and therefore possible. But no *physicalist* could comfortably put forward this argument, since Zombies, etc. are equally conceivable. A second argument is based on empirical investigation into the neural basis of experience. But empirical investigation only shows at most that the neural plays a role in determining the phenomenal; this falls short of showing that a bad off BIV might have R (for instance, maybe there is a general externalist condition on experience that is not met by such a BIV). A third argument is this follows from the right theory of R - e. g. the biological theory, or some non-standard ‘narrow’ functionalist theory. But of course this is hard to establish. Indeed, for reasons I will explain (see point 3 below), I myself am confused about how it is consistent with Shoemaker’s own (partially) functionalist theory.

Granting the first premise for the sake of argument, what might be the argument for second premise? The second premise was that in the BIV R represents the computer state roundness*, not roundness, and accordingly that a BIV who has R might enjoy a perfectly accurate experience, this despite the fact that no round thing is before him. Again, Shoemaker does not give an argument. One argument might rest on the assumption that some kind of causal constraint applies to all representation, including the sensory representation of roundness. But (as David Lewis pointed out in his response to Putnam), even if a causal constraint is right for names and natural kind terms, it is very implausible that it is right for all representation. Indeed, against the second premise, Horgan, Tienson, Chalmers and others would say that even in a BIV R represents roundness; they therefore deny that a causal constraint applies here. (Horgan-Tienson would also say the BIV’s *beliefs* represent roundness, in accordance with my ‘grounding

intuition', despite the absence of a causal connection with round objects.) And Colin McGinn's much less extreme Percy-Twin Percy case (from his **Mental Content**) strongly suggests that a causal constraint does not apply to the representation of shape. I am not here defending a kind of radical internalism about sensory representation on which the BIV might sensorily represent roundness. I am only saying that we have not been presented with any strong reason why the BIV could not sensorily represent roundness. Indeed, it seems to me equally hard to explain how a BIV might represent d-roundness, as on the determinable claim suggested by Shoemaker (the last section of my paper is relevant here).

So, the shape claim enjoys stronger support than either one of the premises of Shoemaker's BIV argument against it. I conclude that rejecting one or another of these premises is more plausible than rejecting the shape claim.

Let us now look at Shoemaker's determinable claim, which he says he 'leans towards', and so thinks enjoys some justification. Presumably, that justification is that from the first person perspective R intuitively represents some shape-like property, and the BIV argument means that it cannot be roundness and suggests that it might be something like d-roundness. (Further, at the end of his comments, Shoemaker suggests that the determinable claim is compatible with something very close to the internalism he favors.) Since I do not myself find the BIV argument very persuasive, as I have just explained, I do not find this justification for the determinable claim very persuasive. Further, I have questions about the determinable claim. Is it possible to describe in English the relevant property, d-roundness, of which the radically different properties roundness and roundness* are determinates? Is it something like a Shoemakerian appearance property or qualitative character, a disposition to produce certain responses in individuals? What psychosemantic theory explains how it is represented (an issue briefly addressed in the final section of my paper)? I am very curious about how the determinable claim might be developed, since Shoemaker has not said much about the content of spatial experience, having focused mostly on the content of color experience. I worry that perhaps it cannot be adequately developed (for reasons related to what I say in my paper's final section). So, not only is there no strong argument for the determinable claim; there may also be an argument against it. By contrast, as we saw above, the overall case for the roundness claim is very strong, notwithstanding BIVs. This concludes my discussion of my

reasons for preferring my original roundness claim to the determinable claim that Shoemaker offered in its place.

A final point. The roundness claim (and likewise all my externally-directed intuitions) is consistent with a variety of different views about the metaphysics of the shape property that R necessarily represents: it might necessarily represent round-from-here (Tye), a Thouless property (Hill-Bennett), 'perfect roundness' (Chalmers). There are also difficult and fascinating questions arising out of Brad Thompson's 'double earth cases' and his arguments based on them. I do not deny this. But, whatever the correct view, I claim that there will be some sense in which R necessarily represents roundness at place p. (There are purely 'Fregean' views which reject this roundness claim; but I think they should be rejected, not only because they violate the intuitive shape claim, but for other reasons that Chalmers gives in his 'Eden' paper – for instance, they do not accommodate certain 'internal connections' among experiences in different modalities.) The roundness claim, however it is to be explained, is enough for my argument against biological theories. Indeed, as Shoemaker explains, even his determinable claim would be enough.

2. Doesn't Shoemaker's Intentionalism Imply External-Directedness?

As I say at the end of my paper, I am very interested in Shoemaker's view of experience. Since it is intentionalist, it would appear to accommodate 'external-directedness'. Since it is internalist, it also accommodates 'internal-dependence', which I'm attracted to for empirical reasons. But I did not understand some things Shoemaker said in his comments. So I close with some questions.

My first question is this. Shoemaker says he **leans towards** the determinable claim, the claim that (in his words) "what is necessarily true of R is that if someone has it then he has an experience of something having properties that are in certain ways functionally like the properties red and round". He goes on to say that this would rule out biological theories. So he says he merely leans towards something in the vicinity as the externally-directed intuitions on which my anti-biological argument relies. This confused me, because since for the last 15 years or so he appears to have been **strongly committing** himself to intentionalism and hence to the truth of certain externally-directed claims, on the basis of what he calls "the Moorean transparency intuition". For instance, as noted above, he has recently written that

“each quale has, constitutively and independently of context, a representational content of its own”. And he explicitly says that this is so even in a BIV. Now Shoemaker mostly discusses color experience, whereas the focus of my paper is externally-directed intuitions about shape experience. But transparency applies equally to shape experience. Why do Shoemaker’s papers feature unqualified endorsement of something in the vicinity of my externally-directed intuitions, whereas his comments he says he only leans towards something along these lines?

3. How Is Shoemaker’s Partial Intentionalism Consistent with his BIV Claim?

Shoemaker says that a bad off BIV might have R. This is crucial to his argument against my roundness claim. Indeed, he says that a bad off BIV might be a phenomenal duplicate of a human being, even though (he thinks) the BIV has no shape beliefs (but only shape* beliefs). Call this the *the BIV claim*. I do not see how Shoemaker’s BIV claim is consistent with his partial functionalism – I am curious to know what I am missing. (I can see how a functionalist account is consistent with a non-bad off BIV having experiences, since a non-bad off BIV belongs to a typical population and so on; but here we are considering a bad off BIV.) While he thinks certain qualia (viz. color qualia) cannot be functionally defined because of spectrum inversion, Shoemaker is a functionalist about phenomenal similarity. He holds that (intrasubjective) phenomenal similarity can be defined in terms of experiences’ effects on (i) beliefs (or ‘representations’) about objective similarities in the environment, and (ii) their effects on discriminatory behavior. I *think* he also holds that these definitions will be ‘analytic’, and so available on reflection. (Two small points. (i) There are interesting questions about how the details might go. It is well-known that ‘experience x resembles experience y more than z’ cannot be straightforwardly defined in terms of the relative numbers of jnd’s between the experiences. But I will set this issue aside. (ii) Since it is much less plausible that behaviorally undetectable inversion of ‘shape qualia’ is possible than it is than behaviorally undetectable inversion of ‘color qualia’ is possible, for all I know, Shoemaker holds that individual shape qualia are completely functionally definable. He has not addressed this issue in his work, as far as I know. However, my points below are independent of this issue, and only require that Shoemaker accepts a functionalist account of phenomenal similarity.)

There are two reasons why I wonder about how Shoemaker's functionalist account of phenomenal similarity is consistent with his BIV claim.

First, on Shoemaker's functionalist account, if two shape experiences are similar, then they are apt to give rise to a belief (or representation) that there is an 'objective' similarity among objects. Now presumably the belief is a belief to the effect that there is an objective similarity *in some respect*. Further, Shoemaker holds that the right functionalist account of phenomenal similarity is available on reflection by conceptual analysis. It seems to me implausible that conceptual analysis should reveal that the relevant respect is something so arcane as d-shape (which can be realized by shape and by the computer-state shape*). If the account is meant to be a bit of conceptual analysis, it seems more plausible that relevant beliefs concern similarity in (ordinary) shape. So, on the most plausible way of elaborating Shoemaker's analytical functionalist account of phenomenal similarity, if two shape experiences are similar, they are apt to give rise to a belief concerning an 'objective' similarity with *respect to (ordinary) shape*. But this contradicts Shoemaker's BIV claim. On his BIV claim, the BIV has all the same shape experiences as an ordinary individual, and they stand in the same similarity relations as they do in an ordinary individual; but, on his view, contrary to the functionalist account (as I think it must be elaborated), they are *not* apt to give rise to a belief concerning an objective similarity with respect to shape, but rather (at best) a belief concerning a similarity (or similarity*) with respect to shape*, where this is a type of computer state. Evidently, I am missing something – but I am curious to know what it is.

Second, Shoemaker's functionalist account implies that, if two shape experiences within a certain individual are similar, then they are apt to give rise to certain behavioral outputs (e. g. sorting behaviors). How are these behavioral outputs to be characterized? (i) They might be characterized in a mentally-loaded way: e. g. as events that are caused by certain beliefs and desires, and satisfy certain beliefs and desires (e. g. the desire to sort or sort* objects or objects* together), or cause certain experiences (e. g. the appearance of sorting objects together). Then the computer events that the BIV is apt to cause might count as realizing the relevant behaviors, provided the BIV has the relevant experiences and beliefs and desires. But of course the proponent of the functionalist account cannot characterize the

relevant behaviors in this way, because the relevant functional definition of phenomenal similarity is supposed to be devoid of mental terms. (ii) The behaviors must therefore be characterized in a mind-independent way. To achieve this, the functionalists Lewis (in *Reduction of Mind*) and Jackson&Braddon Mitchell (in their intro book) suggest characterizing behavior in terms of environmental impact: for instance, causing a round thing and a oval thing to go into the same bin. But this proposal is not consistent with Shoemaker's bad off BIV claim, since the computer events that the BIV is apt to cause do not (even potentially) have such environmental impacts. Again, I'm sure that I am missing something. But I'd like to know what I'm missing, because, as I said above, I am very interested in how one might work out a view (such as Shoemaker's) on which experiences is not only 'externally-directed' but also 'internally-dependent'.

4. Multiple Realizability, Indeterminacy, Primitivism

Shoemaker asks: why do I think the multiple realization argument is not enough to refute the biological theory? The reason is that I think that the physicalist has no strong argument, based on empirical considerations or apriori reasoning (e. g. conceptual analysis), for thinking that multiple realizability (sufficient to refute the biological theory) is possible. This is related to the "harder problem" of consciousness discussed by Block and Papineau. However, even if my skepticism is misplaced and the multiple realization argument provides some reason for rejecting the biological theory, I hope my argument is still worthwhile, if it adds to the case against the biological theory. Indeed, I think that the externally-directed intuitions on which my argument relies are on firmer ground than the multiple realizability claim on which the multiple realizability argument relies.

I said that I think that it is difficult to settle, empirically or apriori (e. g. by conceptual analysis), whether multiple realizability (sufficient to refute the biological theory) is possible. In fact, I think that reductive physicalists might be forced to say that the extension of experience-terms is *radically indeterminate*, and that the extent to which multiple realization is possible is likewise indeterminate. David Papineau defends this kind of view, but the view I have in mind differs from Papineau's in form and motivation. We have, I think, two sets of intuitions that pull in opposite directions. On the one hand, we have the externally-directed intuitions. These are best satisfied by multiply-realizable, non-biological states that involve the environment. On the

other hand, as mentioned at the close of the paper, we also find compelling claims about the ‘internal-dependence’ of experience as well as the resemblance structure of experience. I think that these are best satisfied by biological properties. (I develop this ‘argument from structure’ for biological theories elsewhere: <https://webpace.utexas.edu/arp424/www/rest.pdf>.) It is hard to see what could make it the case that experience terms refer to one of these sorts of candidates to the exclusion of the other. So the reductive physicalist might have to say that experience terms are indeterminate in meaning among the ‘broad’ and ‘narrow’ candidates, with the result that their extensions are radically indeterminate. (Ted Sider defends a similar view about personal identity in his ‘Criteria of Personal Identity’.) Now, although I take it seriously and am surprised it is hardly ever discussed, I cannot myself accept this sort of ‘deflationary’ approach on which the extension of ‘has an experience of a round thing’ (etc.) is radically indeterminate. Instead, my view (which I briefly mention at the end of the paper and develop in greater detail elsewhere) is that there are primitive, relational properties which satisfy all of the relevant intuitions about external-directedness, internal-dependence, and the rest. These properties are extremely ‘natural’, and hence are ‘reference magnets’, even if their supervenience-bases are enormously complex. So experience terms (determinately) refer to them. (I realize these remarks may seem cryptic; if so, ignore them.) However, even on this view, there is room for skepticism about the extent of multiple realizability. Again, I think that the multiple realizability argument against biological theories is difficult to defend. I hope that my externally-directed arguments provide an interesting alternative case for rejecting such theories.