Response to Pautz, "Why Consciousness Cannot Just Be in the Head"

Douglas Keaton
Thomas Polger
Department of Philosophy
University of Cincinnati*

Adam Pautz argues that biological theories of consciousness—roughly speaking, type-identity theories—cannot be correct. As he recognizes, his arguments follow the standard Leibniz' Law based argument schema against identities (what else could it be?), for which he supplies four substitution instances.

Readers will probably find the argument familiarly Cartesian: Pautz aims to show that conscious experiences have some properties essentially that nothing exclusively "in the head" has essentially, and thus conscious experiences cannot be identified with events or states that are exclusively in the head. Although there is a great deal of supplementary argumentation in his paper, largely aimed at preemptively combating anticipated objections, the main question is clear: Has Pautz shown that consciousness experiences do and brain events or states do not have the "externally-directed" properties in question essentially?

We think that there are a number of aspects of the Pautz argument that can be resisted by identity theorists. Here we'll focus on a neat half dozen.

1. Biological theories and all the others

First, we want to register some concerns about how Pautz characterizes the target theories against which he argues, and whether he succeeds at picking out theories that anyone holds. Pautz contrasts "biological" theories with "functionalist-externalist" theories, and it is clear that the externalist element is the key. According to Pautz's biological theorist, experience types are necessarily identical to neural property types, where a neural property type is "a physical-computational property whose definition only involves matters inside the head" (3).

For clarification, we will assume that "inside the head" is an idiomatic way of talking about "inside the body." No contemporary identity theorists would deny that the spinal and peripheral nervous systems are part of the neural bases of experiences. And many of us think that the enervated sensory systems are intricately involved in experience, or at least that it is an open empirical question to what extent they are. So we really cannot take Pautz's "inside the head" idiom too literally. If there is any relevant boundary at all, surely it must be between the interior and exterior of the "skin-bag," as Andy Clark sometimes calls it.

* Authors are listed alphabetically.

Even so, two questions arise: One about the scope of the competing theories, and one about whether anyone holds the theory that Pautz attacks in the form that he attacks. Let's formulate three versions of the so-called biological theory:

> A. Every experience property type is necessarily identical to a physical-computational property type whose definition only involves matters inside the skin-bag.
>
> B. Many experience property types are necessarily identical to physical-computational property types whose definition only involves matters inside the skin-bag.
>
> C. Some experience property types are necessarily identical to physical-computational property types whose definition only involves matters inside the skin-bag.

Now, the biological theories that Pautz targets are clearly supposed to have universal scope (A), which is why he thinks that just one counter-example will do (p.4). For that reason, he can argue that there are some experience types that have externally-directed properties essentially, no neural properties have externally-directed properties essentially, therefore biological theories are false because "neural-types are just not the sort of things that could be experience-types" (p.2).

But why should the biological theorist not prefer to formulate a theory with a narrower scope, like (B) or (C)? We don't see why not. And in that case, in order to prevail Pautz will have to argue that every experience type (i) has externally-directed properties and (ii) has them essentially. He clearly holds this view, but we don't see that he has provided an argument for it. (But see his fn.4, p.2; and our remarks on it below.)

One might wonder whether the narrower formulations, (B) and (C), amount to a surrender to the functionalist-externalist. We'll return later to the question of how to distinguish identity theories from functionalist theories, and whether Pautz has correctly done so. But on the specific question of scope, as we understand the situation it is the functionalist-externalist who has advanced the universalized theory—who claims that no sensation types can be identified with neural types. And that is the view that we find implausible. We're not aware of anyone who holds the (A)-scope version of the identity theory, but maybe there are some who do. So let's play along.

2. The "Intuitions" are not intuitions

The substitution instances of Pautz's argument schema each invoke an "intuition" to the effect that experiences have some property essentially. Pautz calls these assertions "sacrosanct" (p.4) and implies that it would be dishonest to deny them (p.29). He calls these claims "intuitions" because he thinks they are obvious and (he implies) a priori; he

suggests that they are arrived at by "transcendental" (p.2, fn. 4) reasoning about conscious experiences. But it is fairly plain in the text that the "intuitions" are either observations, or inferences about introspective observations; and they are not a priori in any strong sense. On one hand, this should not be a problem for Pautz. After all, if we can introspect that experiences have properties that neural events cannot, then the Leibniz' Law argument goes through. But on the other hand, it is hard to see how we can introspect whether our experiences have some property essentially, as opposed to whether they have it at all. And in both cases we may be prone to reasonable and explainable error. For the advocate of the biological theories can provide a theory-based deflationary explanation of each "intuition" without begging any questions. And we are going to help ourselves to that tactic.

Pautz presents his argument schema in terms of experience property R, "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" (p.2). He then constructs the Leibniz' Law argument 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

The substitution instances of (1) are the main claims that Pautz means to establish, and they will be justified by introspection. (2) is just the universal elimination on:

(2*) No neural property has any externally-directed property necessarily.

The justification for (2*) remains to be discussed. But the question will be whether the "intuitions" show that experiences have "externally-directed" properties in a univocal way that renders (1) and (2*) both true, and thereby renders the argument sound. First, let us consider the instances of premise (1).

3. The "intuitions" are ambiguous and false

Pautz calls the four claims the *externality*, *matching*, *grounding*, and *justification* intuitions. The last two are completely implausible, and no biological theorist should believe them. The first two are superficially plausible, but that is because they suffer from an ambiguity. We'll discuss the general ambiguity and then apply it to each so-called intuition.

There is a well-worn and not well understood ambiguity in the idea of the "contents" of experience. On the one hand, there is a notion of "content" that has to do with how experience presents the world to be, and that appears to be well approximated if not fully explained by notions of representation, intentionality, aboutness, and such. This kind of content is plainly "externally-directed" with respect to the experience itself, and plausibly with respect to the "head." On the other hand, there is the notion of "content" that has to

do with the character of consciousness itself and what its properties are, where the contents are the properties that experiences have rather than the properties (if any) that experiences present. This kind of "content" is not externally-directed because it is not directed at all. This is because not really any kind of "content" in the philosophers preferred sense, it is just the having of a property. Call the first kind of content *epistemic/semantic*, because it purportedly plays a role in epistemic and semantic theories.[1] Call the second kind of content *raw*, for lack of a better term. Are the contents of experience more like the contents of a sentence or the contents of a box?

As all the world knows, many philosophers believe that all facts about experience can be explained in terms of epistemic/semantic content; and in particular they claim that raw content is exhausted by epistemic/semantic content: phenomenal properties are epistemic/semantic properties. But that is a theory, and not one that the biological theorist must adopt. So we won't.

Now consider Pautz's claim that everyone's experience of a tomato (R) essentially has the following second-order property, which he calls the *externality property*:

> being a property such that, if anyone has that property [R], then he has an experience as of a round thing at a certain viewer-relative distance *d* and position *p*. (p.5)

We think the expression "as of" is just the sort of expression that is used to obscure whether the externality property is supposed to be a epistemic/semantic property of experience or a raw property of experience. So really we have two candidates:

(a) being a property such that, if anyone has that property R, then he has an experience property like that which is had by me when I see a round thing at a certain viewer-relative distance *d* and position *p*.

(b) being a property such that, if anyone has that property R, then he has an experience that presents the world as though there were a round thing at a certain viewer-relative distance *d* and position *p*.

The biological theorist should have no trouble with reading (a), for it's just an old-fashioned topic-neutral reference fixer for experiences: "*There is something going on which is like what is going on when* I have my eyes open, am awake, and there is an orange illuminated in good light in front of me, that is, when I really see an orange" (Smart 1959:149, italics original). But that property is not essentially externally directed or, as far as we know, directed at all. Reading (b), in contrast, plausibly picks out an

[1] As Pautz says, in his fn. 4 (p.2), "a thread running through much recent work on perception is an emphasis on role of experience in making possible, and justifying, external thought," and it is this characteristic of experience with which he is concerned. Biological theorists, insofar as they are a coherent group, are typically unconcerned with this alleged semantic/epistemic feature of experiences. Typically we think that theories of content and justification should conform to the ontology, rather than vice versa.

externally directed property, but not one that the biological theorist need say that experience—or anything else—has essentially.

Because the "as of" locution is so slippery, let's set it aside for a moment and examine the other externally-directed properties that Pautz finds in experience: matching, grounding, and justification. Consider matching:

> being a property which is such that, if anyone has it [R], then he is in a state that matches the world only if a round object is present. (p.6)

Pautz glosses "matches" as "corresponds to the way things are" (p.6), which is a plainly epistemic or semantic notion of matching. But according to the biological theorist, conscious mental states do not have any such properties essentially. The raw content of experience may or may not have a structure that "matches" the world in that it is isomorphic to and/or caused by the world in various ways. But plainly those would be contingent, and it would be false that those matchings would occur "only if a round object is present"—for everything is isomorphic to indefinitely other things, and thereby "matches" them in indefinitely many ways that are not epistemic/semantic.

So the pattern is clear: It's not any old kind of external directedness that concerns Pautz, it's the special kind or kinds that might play a role in making possible and justifying conceptual thought. We don't need to repeat our discussion for "grounding" or "justification" for it should be obvious how those go. Those two are not even superficially plausible to the biological theorist. No biological theorist must think that experiential properties are essentially such that they can make possible or justify beliefs in the manner considered by Pautz. But nor will most functionalists. On most materialist theories, that will be a contingent feature of experiences if it is a feature of them at all.[2]

4. The Second Premise

According to Pautz's second premise, no neural property has any externally-directed property essentially. Pautz defends this substitution instances of this premise at length, through a series of thought experiments about what he calls "separation cases." We're not sure what the big deal is. Plausibly, nothing at all essentially has the kinds of externally-directed properties that concern Pautz—viz., semantic/epistemic directedness. Since nothing has those properties essentially, no neural state has them essentially. But, also, no mental state has them essentially. So there is no trouble for the biological theorist.[3]

---

[2] Despite initial appearances, it seems like Pautz's argument is more similar to the arguments of Kant and McDowell than those of Chalmers. But we're not sure why the biological theorist gets singled out: That physical properties and causes are not essentially located in the space of reasons is a standard complaint against all "naturalist" theories by those who take the neo-Kantian approach to perceptual experience.

[3] According to Pautz, philosophers who reject proffered intuitions are guilty of accepting a "magical theory of intentionality" (p.13). Pautz says this even though he denies that his externally-directed properties have anything to do with intentionality (p.5, fn. 7).

We think that Pautz has, in a manner of speaking, reinvented Black's objection to Smart's identity theory. Pautz thinks that there must be some essential property of mental state types—their externally-directedness—by which we recognize them as mental state types, and that is such that no physical state type ("defined" as such) can have that property essentially. Pautz may be right that this kind of reasoning is compatible with any theory of externally-directed properties, not just the usual functionalist theories.[4] But it is not compatible will all theories of the meanings of our mentalistic terms, or of the modal properties of the entities to which those terms refer. In short, the biological theorist who thinks that mental states and neural states are necessarily identical will typically think that they are a posteriori identical. Such a theorist will deny that there are any defining properties of mental states that they can be known to have a priori. Or, to put it in the classic form, the biological theorist will hold that mentalistic vocabulary is topic neutral, so does not refer to mental states by picking them out by means of their essentially mental properties.

5. Concerning Slug

Pautz alleges that biological theorists believe in "magical" neural states that have externally-directed properties even in "separation cases" in which those neural states are isolated from their functional roles. On Pautz's telling this is practically to believe an outright contradiction: that a state could have a property in the absence of the only kind of condition that can bestow that property. But the response should be obvious. Philosophers who don't accept functionalism are not going to accept Pautz's "separation cases" in the first place. This is because the separation cases are described in functionalist terms, specifically, in terms of total realizers.

We can expose the difficulties if we examine how Pautz muddles the notion of a total realizer in his "separation cases." Pautz asks to imagine the case of Slug, a creature with a total realizer for *R* but in whom the total realizer is functionally isolated. The total realizer is "not even apt" to be caused by the typical causes of *R*, in Slug.

But it is hard to make sense of the idea of a total realizer of an experience that is "not even apt" to be caused by the typical causes of that experience. A total realizer is a conjunctive property whose conjuncts are (at least) the core realizer (which is apt) and a property that Shoemaker (1981) designated as T(x). T(x) is the property of having a specific n-tuple of physical properties that jointly satisfy the psychological theory T. To have T(x) is to have a specific set of core realizers at the ready to play the causal roles of every psychological state mentioned in theory T. So if Slug or any other creature has a total realizer for the experience of seeing a round tomato at a certain distance d and position p, then Slug *also has* a complete set of specific core realizers, one for every

[4] But given his broad characterization of functionalism and the fact that the contrast class are "functionalist/externalist" theories, we're not sure what else is left.

psychological state; all of which are hooked up to each other and to input and output mechanisms in T-appropriate ways.

One could, we suppose, imagine a bizarre creature that is shaped like a balloon with another creature inside of it. The interior creature, Slug, has states that satisfy T (among them, *R*) but, sadly, the interior creature is forever cut-off from the outside world—cut off by the surrounding body of the host creature, Slug+. But this is a decidedly *non-normal* circumstance; so all bets are off about normal circumstances, and nothing of interest seems to follow from this sad state of affairs. If we read Pautz correctly he is insisting that the interior creature has experiential states that are *not even apt* to be caused in the *normal* ways. That's what we find puzzling, given the stipulation that Slug has the total realizer for R.

Of course, anyone who denies the functionalist account of mental states *tout court* will deny that mental states are identical to total realizers. For that matter, they will probably deny that there are any such things as total realizers at all. Identity theorists may fall into this camp. For this reason, Pautz's discussions concerning what biological theorists can or should say about total or core realizers is beside the point. Identity theorists don't have to say anything about core or total realizers. They are free to deny that they exist, or at least to deny that they have anything to do with mental states.[5]

6. Biological theories and all the others, again

Finally, we want to return to some concerns about how Pautz characterizes biological theories. We think that this question is worth revisiting because Pautz says that both Ned Block and John Bickle hold the sort of view he intends to criticize, and we doubt that Block and Bickle share any beliefs whatsoever, even about the weather, let alone a theory of mind.

Pautz's first and primary target seems to be the idea that experience terms rigidly designate core realizers. We are not sure who is supposed to have held such a odd view, but certainly not anyone he lists. One reason for this is that core realizers of *R* are by definition not metaphysically sufficient for *R*, so not candidates for identification with *R*. A second reason is that the very idea of a "core realizer" goes hand-in-glove with functionalism: with a theory that appeals to role-occupancy, and so to realization. But this is exactly the idea with which the biological theorists mean to contrast their view.

Eventually, as we discussed above, Pautz allows that his opponents may fall back to the view that experience terms rigidly designate *total* realizers. But that is also puzzling, and for exactly the same reason. Indeed, Pautz himself says that the notion of a total realizer is an inherently *functionalist* notion—a functionalist way of characterizing a "big" neural state. (He writes, "the big state theory is very much like functionalism" (p.22). Yes, it

[5] On the many confusions about core and total realizers, see Keaton (forthcoming-a, "Kim's Supervenience Argument and the Nature of Total Realizers," *European Journal of Philosophy*; forthcoming-b, "Two Kinds of Role Property," under review; and Ph.D. dissertation, University of Cincinnati.)

is.) It's not the case that the total realizer for R is equivalent to "whatever is metaphysically sufficient for R" or "the proper supervenience base for R." So we are somewhat at a loss as to why Pautz characterizes biological theories in terms of realizers.

In closing, we will allow our selves to speculate about the view that Pautz has in his sights. As we've said, we don't think it's the view held by either Block or Bickle, but we are ready to be corrected. It seems that Pautz has in mind a class of theories constituted by something like the following three theses:

(a) The right way metaphysically to individuate the properties that constitute experience properties is via their causal/functional roles, following David Lewis

(b) Multiple realizability is false (i.e., not functionalism); experience terms rigidly designate neural property types.

(c) Experience makes possible and justifies contentful thought.

Now, we're not sure who if anyone holds this view. Claim (c), as we have been at pains to point out, has nothing to do with the identity theory. And in recent years it has become clear that the identity theory is not just realizer functionalism, à la Lewis. That is, the identity theory is not just the theory that mental states are uniquely realized—the conjunction of (a) and (b).

At best, the identity theorist will say that functional roles provide rough-and-ready methodological clues for finding brain structures that are experiences—one might think of this as Cummins as opposed to Lewis style functionalism. Could such experiences be instantiated even if the brain structures were removed from the surrounding tissue and, say, electrically stimulated in a lab? Hard to say. But an identity theorist can say that the answer is not an *a priori* "no" as a metaphysical functionalist would have to say.

If we've come this far, then maybe if we said (against our better judgment!) that some epistemic/semantic roles are constitutive of genuine experiences, then we would have to say that it is an open question whether the Petri-dish properties are apt to make possible and justify beliefs.[6] Maybe *that* is the view that Pautz thinks is crazy and magical. And then perhaps *that* is the (very general) view that Pautz means to criticize: the view that one shouldn't give an a priori "no" to the question about the lab. Perhaps Bickle and Block would agree on that one should not give an a priori "no" to the lab question and so to that extent share view. Note, though, that this question about labs is a *different* question than the one that Pautz raises with his separation cases, which is closer to: Can experiences be necessarily identical to either core or total realizers?

[6] Pautz repeatedly brings in the possibility of disjunctivism. If "experiences" are what is had by cognitive creatures only in successful perception, then maybe the biological theorist is only giving a theory of the animal processes that underlie "experiences"—call them, for lack of a better term, *sensations*. The biological theorist would then hold, as one might say, that "sensations are brain processes."

The answer to *that* is: For core realizers, of course not. For total realizers, almost certainly not, not least of which because identity theorists have no reason to reify role properties at all. And in any case, would one ask this question, and to whom?