

Qualia ain't in the head

Alex Byrne

Massachusetts Institute of Technology

and

Michael Tye

The University of Texas at Austin

Qualia internalism is the thesis that qualia are intrinsic to their subjects: the experiences of intrinsic duplicates (in the same or different metaphysically possible worlds) have the same qualia. *Content externalism* is the thesis that mental representation is an extrinsic matter, partly depending on what happens outside the head.¹ *Intentionalism* (or representationalism) comes in strong and weak forms. In its weakest formulation, it is the thesis that representationally identical experiences of subjects (in the same or different metaphysically possible worlds) have the same qualia.²

These three theses are widely held—especially the first two. But with the addition of some relatively innocuous assumptions, they are inconsistent. Take color as an example. Consider Bill and Ben, ordinary humans who are enjoying color experiences with different qualia. Let x be a (possible) duplicate of Bill, and let y be a (possible) duplicate of Ben. Given a specific externalist theory of content (which need not be reductive), with some ingenuity we can plausibly construct different environments for each, such that the theory predicts that x and y 's color experiences have the same content; so, by (weak) intentionalism, they have the same qualia. By qualia internalism, x 's experience has the same qualia as Bill's, and y 's experience has the same qualia as Ben's, so x 's and y 's experiences differ in qualia; contradiction. Alternatively, since an intentionalist about color qualia will typically endorse the converse thesis that the color content of an experience supervenes on its color qualia, we can start with a pair of duplicates x^* and y^* in different environments and use content externalism to argue that their experiences differ in content. Since x^* and y^* are duplicates, their experiences have the same qualia; by the converse intentionalist thesis, their experiences have the same content.

So: content externalism and intentionalism (jointly, “externalist intentionalism”) naturally lead to qualia *externalism*. And what's wrong with that? Isn't the doctrine of qualia internalism the last bastion of a widely discredited Cartesian conception of the mind?

¹ As the interaction between the organism and its environment becomes more complicated, it becomes more contentious to suppose that we can keep the organism intrinsically constant while radically varying its environment. However, to the extent that this is so, it counts against the *internalist* (see Williamson forthcoming).

² Strong intentionalism identifies qualia with representational contents that meet further conditions. See here Tye 2000.

Not according to many philosophers, who view qualia externalism with the same incredulity that greeted Churchland-style eliminativism. Qualia externalism, they think, is an absurd thesis, accepted by a handful of philosophers with too much respect for philosophical theory and not enough common sense.

To his credit, Adam Pautz does not rest his opposition to qualia externalism on this kind of “intuition”. He attempts to provide an argument against the principal motivation for it, namely externalist intentionalism. Moreover, the argument purports to be in significant degree *empirical*, drawing on results from a variety of disciplines, including psychophysics and neuroscience.

The orthodox response to our quasi-inconsistent triad is to deny intentionalism, not content externalism. Interestingly, Pautz takes the other option, and embraces content *internalism*.

So far, we have not mentioned the issue of reductive physicalism, which looms large in Pautz’s presentation. In our view, bringing in inevitably controversial reductive theses of the “awareness relation” at the start just makes it harder to see what is going on. Accordingly, we will initially set out Pautz’s argument against externalist intentionalism while ignoring the various reductive proposals that Pautz discusses. After having explained why Pautz’s argument fails, we then turn (in section 2) to the entirely separate issue of whether there is some relatively compact wide physicalistic account of the awareness relation.

1. Pautz’s argument

Pautz’s argument purports to establish that there is a pair of possible subjects x and y meeting the following two conditions: (a) x and y are having “different experiences” (that is, experiences with different qualia; see 3–4³), and (b) externalist intentionalism predicts that x and y are having the “same experiences” (that is, experiences with the same qualia). His main example involves color perception: Maxwell is a normal human perceiver in the actual world; Twin Maxwell lives in another possible world, in which the evolution of the human color vision system has gone slightly differently, resulting in “different postreceptoral wiring” (7), but leaving the photoreceptors unchanged.

The argument has two stages, corresponding to (a) and (b) above. In stage one, Pautz argues that, when Maxwell and Twin Maxwell are viewing an orange square, they have different experiences. In stage two, Pautz argues that the externalist intentionalist must say that Maxwell and Twin Maxwell have the same experiences.

1.1 Stage one

Pautz offers two independent arguments for the conclusion that Maxwell and Twin Maxwell are having different experiences.

³ page refs to the ms.

1.1.1. The argument from C-Dependence

Pautz's first argument is advertised as relying on various empirical results from color science, specifically those connected with opponent process theory. The basic idea is that opponent process theory has made it very plausible that if two subjects are in suitably different "postreceptoral" neural states, as are Maxwell and Twin Maxwell, then their experiences will have different qualia. According to Pautz, "[t]he . . . **opponent process theory** of color vision shows us that the best explanation of the character of the quality space for color experience is to be found *in the brain*" (4, our emphasis).

This is seriously misleading. So far, nothing has been discovered "in the brain" that explains the character of color quality space. J. D. Mollon summed up the situation in 1997 as follows:

We still believe today that there are chromatically antagonistic channels in the early visual system—that is, channels that draw inputs of opposite sign from different classes of cone, but these channels simply do not correspond to the phenomenologically defined channels of Hering . . . In fact, no one has found a site in the visual system where colour appears to be represented according to Opponent Colour theory—that is, a site where the cells might be held to secrete redness and greenness or yellowness and blueness. Cells are found in the cortex that respond to restricted regions of chromaticity space, but they are by no means confined to the loci of pure hues. Thirty years ago we thought we understood the existence of four unique hues, hues that are phenomenally unmixed. *Today this is perhaps the major unsolved problem of colour vision* (Mollon & Jordan, 1997). *If we understood it, we should probably be much closer to understanding the general relationship between neural activity and qualia.* (Mollon 1997, 870-2, our emphasis)⁴

As far as we can tell, these remarks still apply.

With this corrective in mind, let us turn to the details of Pautz's argument. It proceeds from this premise:

C-Dependence: Opponent channel activity plays a direct role in determining the character of color experience. By virtue of [its] reflectance [profile], an object reflects certain light and sets up certain opponent channel activity in us. In turn, that activity **directly** determines the character of the resulting color experience.
(6)

What does 'determines' mean in the statement of C-Dependence? Pautz tell us that "[t]he modally weakest interpretation of C-Dependence is a Dualist one . . . [on which] C-Dependence is supported by a brute psychophysical law directly linking opponent activity with color experience" (6). On the other, "Physicalist",

⁴ Thanks to David Hilbert for this reference.

interpretation of C-Dependence, “color experiences are somehow constituted by opponent channel states” (10). All this strongly suggests that ‘determines’ means *either nomologically or metaphysically necessitates*.

What about ‘directly’, which modifies ‘determines’ in the statement of C-Dependence? (Opponent channel activity is also said to “play a direct role in determining the character of color experience”, but presumably this is intended to be equivalent to the activity “directly determining” the character of experience.) Although Pautz bold-faces the word, and so presumably thinks it cannot be removed without loss, it is unclear what he has in mind. One possibility is this: proposition P *directly* determines proposition Q iff P determines Q and P does not determine any other proposition that determines Q. But this can hardly be right, because if P and Q are contingent and distinct (as they will be in cases of interest), then P will *not* directly determine Q: if P determines Q, then the distinct proposition $P \vee Q$ will also determine Q. In any case, as far as we can see, ‘directly’ is not doing any work in Pautz’s argument, so we will ignore it in what follows.

Call the resulting interpretation of C-Dependence, *Strong C-Dependence*:

Strong C-Dependence: internal neural states and processes (specifically opponent channel internal states and processes) either nomologically or metaphysically necessitate the character (qualia) of color experience.

According to Strong C-Dependence, if x and y are alike neurally, and live in worlds governed by the same laws, then their color experiences have the same qualia.

Some textual evidence points unequivocally to this interpretation of C-Dependence. Unfortunately, other textual evidence points unequivocally against it. Immediately after stating C-Dependence, Pautz comments:

C-Dependence . . . does not entail **Internalism** about color experience: the strong thesis that internal factors *completely* determine color experience, so that neurobiological duplicates living under the same laws have the same color experiences. (6)

And earlier, when explaining Dependence, the generic thesis of which C-Dependence is a species, he writes, apropos of cases “in which two possible individuals are in different internal neural states”, that:

Dependence is *no stronger* than the claim that, *at least in certain such cases*, the correct verdict is Different Experiences: the individuals involved have different experiences. (3, our emphasis)

This is particularly mystifying, because the argument from C-Dependence is supposed to show that “in certain such cases [involving color experience], the correct verdict is Different Experiences”. Moreover, the argument is not trivial, occupying three paragraphs in Pautz’s paper. Yet the above quotation implies that C-Dependence is no stronger than the claim that, in certain cases (e.g. that of Maxwell and Twin Maxwell), the correct verdict is Different Experiences. And since the argument does not seem to appeal to any other premise, C-Dependence is apparently no *weaker* than the verdict of Different Experiences, in which case

the two are straightforwardly equivalent.⁵

In any event, these other passages suggest a quite different interpretation of C-Dependence. Since Pautz presumably does not intend C-Dependence to be equivalent to the conclusion it is supposed to establish, probably the best interpretation of C-Dependence that fits the above and some other passages is along these lines:

Weak C-Dependence: color experiences are not just *correlated* with internal, neural states and processes, they *counterfactually depend* on such states and processes (specifically opponent channel internal states and processes). Suppose that opponent channel state O is correlated with quale Q: a perceiver's opponent channels are in O iff she is having an experience with quale Q. Then: if a perceiver's opponent channels *had been* in state O, the perceiver *would have* had an experience with quale Q.

Thus we have two “arguments from C-Dependence” for the conclusion that Maxwell and Twin Maxwell are having different experiences—the argument from strong C-Dependence, and the argument from weak C-Dependence. Let us take them in turn.

In fact, if strong C-Dependence is true, the falsity of externalist intentionalism follows immediately, without a detour through the hypothetical case of Maxwell and Twin Maxwell. Strong C-Dependence implies that it is either metaphysically impossible that physical duplicates have different color experiences, or else that it is nomologically impossible (and dualism is true). Since the externalist intentionalists that Pautz is principally targeting are not dualists, and they hold that it is metaphysically possible for duplicates to have different experiences, they will reject strong C-Dependence. Further, any externalist intentionalist who is also a dualist will also reject strong C-Dependence. So, once Pautz has established strong C-Dependence, the game is over.

Pautz gives two arguments for C-Dependence. The first is this:

Nothing in the outside world can explain the unitary-binary character of color experience, the comparative resemblance relations among color experiences, or the psychophysical phenomena listed above. So, the explanation must lie in the brain.
(6)

Interpreted as an argument for strong C-Dependence, Pautz's first argument may be set out as follows:

P1. Nothing external explains the unitary-binary character of color experience (etc.).

⁵ This is also suggested by the next sentence, which begins “I do not accept this claim on the basis of intuition; rather, as we shall see, I accept it . . .”, where “this claim” is clearly Dependence (or, restricting attention to the color case, C-Dependence). But then, since “this claim” is naturally taken to be “the claim” of the preceding sentence, “the claim” and (C-)Dependence are the same.

Hence:

C1. Something internal (namely, opponent channel internal states and processes) explains the unitary-binary character of color experience.

Hence:

C2. Something internal (namely, opponent channel internal states and processes) *nomologically or metaphysically necessitates* the unitary-binary character of color experience. That is, strong C-Dependence is true.

Forget about P1 for the moment, and concentrate on the step from C1 to C2. C1 is true, on its most natural reading. But on this reading of C1, the step to C2 is invalid: if P explains Q, it does *not* follow that P determines Q. P usually explains Q relative to (often unstated) background facts R. In a typical case, there are possible situations in which the background facts fail to obtain, P is true, and Q is false. Hence, despite explaining Q, P does not determine it: what (at least sometimes) determines Q is P *and* R. That the match was struck explains why it lit, but the striking does not determine the lighting: the lighting is determined by the striking *and* the presence of oxygen, the dryness of the match, etc. Notice that these background facts partly concern matters *extrinsic* to the match. And for the externalist, the relevant background facts relative to which the internal neural facts explain the unitary-binary character of color experience include, of course, facts about the perceiver's *environment*.

So, if C2 is to be a defensible consequence of C1, C1 must be interpreted as saying that the unitary-binary character of experience can be *completely* explained by internal facts. And if P1 is interpreted as saying that external facts are *no part* of the explanation, then it supports C1. But P1, thus interpreted, is groundless, since Pautz has said nothing at all to exclude the possibility that external factors are *part* of the explanation of the unitary-binary character of color experience. Of course, Pautz is entitled to P1 if it is interpreted as saying that external factors cannot be the *sole* explanation—internal factors are also relevant. But then—depending on the interpretation of C1—either P1 fails to support C1, or C1 fails to support C2.

Pautz's second argument for C-Dependence is stated very briefly:

Why accept C-Dependence?...Systems of neurons whose activity approximates the hypothesized opponent channels have been discovered in the early visual system. (6)

But this is not much of an argument for anything—let alone strong C-Dependence—since, as Pautz notes earlier, these (LGN) neurons “cannot constitute” the hypothesized opponent channels (5). More importantly, even if the opponent channels *had* been neurally identified, this would not support strong C-dependence in the slightest. An analogy: the “hypothesized internal symbols” in the telephone directory that encode information about people's phone numbers have been typographically identified, but typography does not determine semantics.

Strong C-Dependence, then, is not supported by an argument that can be extracted from Pautz's paper. What about weak C-Dependence? Here we need not bother to examine whether either of the two arguments for C-Dependence discussed above can be turned into an argument for weak C-Dependence, because (modulo any concerns about opponent process theory), weak C-Dependence is entirely uncontroversial. Even ignoring sophisticated evidence from macaque LGN single-cell recordings and the like, it is not in serious dispute that mental states counterfactually depend on neural states. Weak C-Dependence will be granted on all sides. But does this thesis show that Maxwell and Twin Maxwell have different experiences?

No, it does not. The argument from weak C-Dependence for Different Experiences may be set out as follows:

P1 (from weak C-Dependence). Maxwell is looking at an orange square. His opponent channels are in state O, and he is having an experience with quale Q. If his opponent channels had been in a different state O*, he would have had an experience with a different quale Q*.

P2. Twin Maxwell (who, for simplicity, can be identified with Maxwell himself) is in a "nearby counterfactual situation" (7) in which evolution goes slightly differently, and in which the causal connections between external conditions and Twin Maxwell's internal states are slightly different. Twin Maxwell is looking at a square of the very same color, and his opponent channels are in state O*.

Hence (appealing to implicit background details about the actual and counterfactual situations):

C. Twin Maxwell's experience (in the counterfactual situation) has Q*, and so Maxwell (in the actual situation) and Twin Maxwell (in the counterfactual situation) are having different experiences.

This argument is invalid. For illustration, assume a standard (and simple) possible-worlds semantics for counterfactuals (and that Maxwell = Twin Maxwell). Then P1 can be rewritten as follows:

P1*. Maxwell is looking at an orange square. His opponent channels are in state O, and he is having an experience with quale Q. In the closest world to the actual world in which Maxwell's opponent channels are in state O*, he is having an experience with quale Q*.

In order to derive C, we need to establish (by appeal to implicit background details):

A. The closest world to the actual world in which Maxwell's opponent channels are in state O* is one in which he is looking at a square of the very same color and *in which human evolution goes slightly differently and the causal connections between external conditions and Maxwell's internal states are slightly different.*

But there is no prospect of doing this because—at least on the orthodox view—

the closest world will keep evolution fixed. Therefore A is false, and the argument fails.

1.1.2. The argument from behavior

The argument from C-Dependence is supposed to rely on *recherché* facts from color science. (As we have seen, such facts are only relevant if C-Dependence is construed as strong C-Dependence.) Pautz does not make a similar claim for his second argument: on the contrary, he says it is “relatively *a priori*” (11).

This argument may be set out as follows:

P1. Maxwell and Twin Maxwell have different color-related behavioral dispositions.

P2 (the “**Experience-Behavior Link**”). “If two actual or possible individuals have qualitatively identical color experiences, then they have the same color-related behavioral dispositions” (11).

Hence:

C. Maxwell and Twin Maxwell are having different experiences.

The intended interpretation of P2 is not completely clear, because Pautz immediately tells us that one admissible version of that premise restricts quantification over “possible individuals” to those in “nearby worlds”, but does not elaborate further. Pautz also claims that “paraplegics and the like are not counterexamples” on the nonobvious ground that they have the appropriate color-related behavioral dispositions. And, finally, Pautz says that to suppose that the **Experience-Behavior Link** fails to hold for “normal individuals” is perhaps “inconceivable” and is “at the very least . . . counterintuitive” (11). This suggests that the **Experience-Behavior Link** is intended to be a conceptual truth. In any case—pending further clarification of the nature of the “nearby worlds”—the second premise of the argument strikes us as a suspicious behaviorist relic.

1.1.3 A simpler argument

Although Pautz’s elaborate arguments fail to show that Maxwell and Twin Maxwell are having different experiences, the conclusion can be established much more straightforwardly. Twin Maxwell’s visual system differs post-receptorally from Maxwell’s. As a result, his behavior vis-à-vis colored stimuli is slightly different. Pautz just needs *one* case where Maxwell and Twin Maxwell’s experiences differ. (Dubious *generalizations* like the **Experience-Behavior Link** are a distraction.) It would be unmotivated to suppose that there is *no* such case. Given that a dramatic post-receptor change will certainly induce dramatically different experiences, why shouldn’t a modest post-receptor change induce modestly different experiences—at least *sometimes*? And to clinch the case, there are—as we will note in the next section—plenty of *actual* examples of more-or-less the Maxwell/Twin Maxwell sort.

Thus, although stage one of Pautz’s overall argument does not succeed, at least his conclusion is correct.

1.2 Stage two

With the conclusion that Maxwell and Twin Maxwell have different experiences in hand, stage two of the argument attempts to show that the externalist intentionalist must deny it. This stage immediately becomes extremely complicated, because Pautz's strategy is to enumerate every candidate externalist theory and show in each case that the theory gives the wrong result.

We can see why the second stage of the argument fails (or, at least, doesn't add anything new) by ignoring the details of particular externalist theories, and considering Maxwell and Twin Maxwell rather abstractly.

Here are two kinds of examples of "different color experiences", both of which have received much discussion in the literature.

Example 1: shifted spectra (Block 1999)

The original formulation, using the fact that unique hue loci vary between subjects, is due to Hardin (1993). In a shifted spectra case, two human subjects, both with "normal" color vision, have slightly different color experiences (i.e. experiences with different qualia) when they look at some colored stimuli in similar conditions. (We may reasonably speculate that part of the overall explanation involves differences in postreceptoral processing.)

Example 2: "non-standard" color vision

Examples of this sort turn on color vision in non-human animals, or (less commonly) color vision in humans with various forms of so-called "color blindness". Here one subject is a normal human, and the other is either a non-human animal (a pigeon, for example) or a human with a color vision deficit (a deuteranope, for example). Although this kind of case is somewhat more speculative (at least for the pigeon), we may assume that such subjects have different experiences when they look at some stimuli in similar conditions.

Set aside content externalism for a moment, and consider what an *intentionalist* should say about these cases. Since our subjects have phenomenally different experiences, their experiences must have *different* color contents. In the shifted spectra case, it is plausible that these contents are incompatible—that is, at least one subject is misperceiving. For instance, the same square looks reddish-yellow to one subject, and yellowish-red to another, and presumably the square cannot have both colors simultaneously. By contrast, in the non-standard case, it is plausible that these contents are compatible: the normal human and the pigeon (and, arguably, the deuteranope) both see things in their true colors. And this is possible because human color space does not include *all* colors: pigeons perceive colors, but not the same ones as us.⁶ Similarly, there are many spatial patterns only perceptible to non-human

⁶ For more here, see Bradley and Tye 2001 and Byrne and Hilbert 2003.

animals.

Pautz's example of Maxwell and Twin Maxwell (who, despite both being humans, do differ in evolutionary history) could be elaborated so as to fit either one of the two above cases. In fact, it is clear that Pautz has the first sort of case in mind (more on this in Section 2). But for safety's sake, we will argue that there is no problem for the externalist intentionalist either way.

Here is the crucial question: is there any problem adding *content externalism* to the mix? Of course, there *is* a problem adding various simple-minded reductive versions of content externalism—for instance, a condition-independent causal covariational account, which would predict that the two subjects' experiences have the *same* content. But content externalism—any more than content *internalism*—should not be tied to a reductive program, let alone a very simplistic one. For example: content externalism of the sort argued for by Putnam and Burge-style thought experiments is not thought to be in trouble because every psychosemantics proposed to date (arguably) makes incorrect predictions. Hence there is no obvious reason why a content externalist should not accept the above intentionalist descriptions of shifted spectra and non-standard cases.

Further, if any problems are posed for content externalism by these two sorts of examples, the connection with intentionalism is somewhat tenuous. Independently of any doctrines about qualia, it is plausible that subjects in a shifted spectrum case have experiences with different contents (they call different things 'reddish-yellow' and 'yellowish-red'). And likewise for pigeons, whose color matching behavior is quite different from ours.

Suppose for the sake of the argument that Pautz has shown that every halfway-defensible wide reductive-physicalist account of the "awareness relation" runs into trouble. The proper conclusion—*at best*—is that the reductive physicalism program cannot work; it would be an error to conclude that content externalism is false. The failure of reductive physicalism is not even a good reason to reject *physicalism*, considered simply as a supervenience thesis. (Note that Pautz himself defines 'physical' very liberally—see 2.)

Pautz cannot move from the alleged falsity of physicalist accounts of the awareness relation to the conclusion that content externalism is false. Yet that is exactly what he does in the first paragraph of section five. Somehow, the preceding is supposed to show that "what properties we sensorily represent" (colors, in particular) "depends on what happens in the head" (the clear implication being that it depends *only* on what happens in the head). So what is going on? Why does Pautz think that his argument establishes (color) content *internalism*?

The answer may be this: Pautz *doesn't* think his official argument involving Maxwell and Twin Maxwell establishes content internalism. Instead, he rests the falsity of content externalism on strong C-Dependence. If strong C-Dependence is true, then the color experiences of neural duplicates have the same qualia. Assuming the converse of intentionalism for color experiences (i.e. that color content supervenes on color qualia), content internalism follows.

2. Pautz's argument against wide physicalist accounts of the awareness relation

We will now briefly comment on Pautz's specific charge that existing externalist physicalist accounts of the "awareness relation" are mistaken.

Let us first take Maxwell and Twin Maxwell to be a pair of ordinary perceivers in the actual world who have slightly different experiences when they look at an orange square—it looks reddish-yellow to Maxwell, and yellowish-red to Twin Maxwell, say. If anything, this simplification of the case should *help* Pautz. Since Maxwell and Twin Maxwell are both ordinary humans living (we may suppose) in the same location, their evolutionary histories and environments are automatically equalized.

A point made in the previous section is worth elaborating. Given the fantastic complexity of color vision, the fact that there are huge gaps in our knowledge of how colors are represented in the brain (and of mental representation generally), and of the selection pressures driving the evolution of color vision, nothing exciting will follow from the failure of existing reductive physicalist accounts of the awareness relation. It would be absurd to think that such failures teaches us anything other than the lesson that mental representation is a very difficult subject.^{7,8} Still, has Pautz shown that these accounts fail?

One thing to bear in mind at the start is that psychosemantic theories are typically underspecified, somewhat vague, and anyway not really 100% physicalistically kosher. For example, note that Pautz's "S-role" (the "functional role characteristic of experiences" (12)), which figures in all six theories he considers, is specified in terms of beliefs and desires (Tye), or "cognitive systems" (Dretske). Of course, Tye and Dretske both think that this mentalistic vocabulary could itself be explained (at least "in principle") in broadly functional terms, but neither pretends to spell out such an account in any detail.

More relevantly, Tye's notion of a state causally covarying with the fact that *p* in "optimal conditions" and Dretske's related notion of a state having the "function of indicating" that *p* (to take two of many examples), are not sufficiently

⁷ If these accounts were all intended to be "conceptual analyses" of the awareness relation then the proper moral might be that the project of conceptual analysis is doomed. However, these accounts are typically not so intended.

⁸ In the course of discussing a "two-factor theory" of the awareness relation, Pautz claims that "the semantic value of 'x is aware of y'" must be a "somewhat "natural" relation", "codifiable in a fairly simple general rule" (20). If so, then a physicalistic account of the awareness relation must be "fairly simple", which arguably implies straightaway that there is no such account. Pautz supports this claim of simplicity by appeal to "Lewis's theory of content in terms of use plus eligibility". (Pautz also alludes to another motivation.) But Pautz has misunderstood Lewis's theory. According to Lewis (1984), if two relations more-or-less equally fit our use of a predicate 'Rxy', then the better candidate for the semantic value of 'Rxy' is the more "natural" relation (ignoring complications due to the interpretation of other expressions). However, the *more* natural relation may not be particularly natural. Semantic values must only be *comparatively* natural—there is no general requirement that they can be simply expressed in some canonical physical language.

well-developed to allow anything approaching a prediction in cases like Maxwell and Twin Maxwell.

Take Tye as an example. Suppose that when Maxwell looks at the square—colored a specific shade of orange (orange_{e17}, say)—the relevant color content-bearing state of his visual system is S_M ; when Twin Maxwell looks at the square, the corresponding state of his visual system is S_{TM} . According to Tye’s externalist psychosemantics, the square will look orange_{e17} to Maxwell iff S_M causally covaries with orange_{e17} in “optimal conditions”: situations in which the “various components” of Maxwell’s visual system are “operating as they were designed to do in the sort of external environment in which they were designed to operate” (Tye 2000, 138).⁹ Similarly, mutatis mutandis, for Twin Maxwell. So what does Tye’s account predict? Does the square look orange_{e17} to both of them, or not? In an actual situation of this sort the differences between the two subjects are so intricate, and the theory so programmatic and oversimplified, that no prediction is forthcoming. Perhaps when Maxwell looks at the square, some components of his visual system are not operating exactly as they were designed to do, and it looks orange_{e18} to him. If so, Maxwell is like a slightly miscalibrated thermometer (Byrne and Hilbert 2004). There is certainly no basis for saying that both S_M and S_{TM} track orange_{e17} in optimal conditions, which is essentially Pautz’s objection.

Let us now change the description of Maxwell and Twin Maxwell to bring it more into line with what Pautz says. According to Pautz, Maxwell and Twin Maxwell are perceivers in different worlds whose color experiences are quite different: “Maxwell has a binary experience of the square while Twin Maxwell has a unitary one. In particular, Maxwell has a red-yellow experience, while Twin Maxwell has a unitary red experience” (10). Similar binary/unitary differences presumably obtain in connection with viewing objects of other colors so that the case is one involving a shift in color phenomenology that corresponds to a 45 degree shift around the hue circle.

This way of understanding the Maxwell/Twin Maxwell case is immediately problematic, because Pautz also claims that Twin Maxwell lives in “the closest possible world” in which human opponent channels are (internally) different (leaving the cones unchanged) (7). And that can’t be right. Instead, we must suppose, *contra* Pautz, that Twin Maxwell lives in a world that is fairly remote from actuality.

Consider first Maxwell’s situation. He is looking at an orange square, and that is how it looks to him. We may assume (as Pautz himself allows) that the components of Maxwell’s visual system are operating in accordance with their design, so that Tye’s account correctly predicts that Maxwell’s experiences pretty much get things right.¹⁰

⁹ There is a further asymmetric dependence condition Tye imposes (2000, 139-40). This can be ignored for present purposes.

¹⁰ Which is not to say that his experiences need be correct as to whether the square is orange_{e17} or orange_{e18}.

Now consider Twin Maxwell's situation. The orange square looks red to him and, in general, he is seriously in error about the colors of things.¹¹ But does Tye's theory incorrectly predict that the square looks *orange* to Twin Maxwell? Everything turns on whether the relevant color content-bearing state of his visual system causally covaries with orange in "optimal conditions". But why think it does? It is not even clear how Twin Maxwell could have evolved so as to systematically misrepresent the colors of things. In any event, given that there is widespread error, a defender of Tye's externalist psychosemantics may fairly insist that one component or other of Twin Maxwell's visual system must be malfunctioning.

So, as Pautz would describe the case, there is no obvious reason to suppose it is metaphysically possible. Allegedly, Twin Maxwell is a product of natural selection, someone operating under the same laws as Maxwell with a similar kind of visual system, whose experiences represent the same range of colors as Maxwell's, and who not only has no abnormalities whatsoever in his visual system but also is subject to significant color illusions. Pautz simply *stipulates* that all these conditions can be met together. A defender of Tye's theory may reasonably deny it. Each condition is indeed metaphysically possible, but they are not all possible together.

We conclude with some more general remarks. Theories of psychosemantics (really, theory sketches) are tested by their ability to deal with clear and simple examples, where the relevant differences are comparatively well understood. Consider a familiar case: Oscar, who is looking at a horse, and Twin Oscar, who is looking at a muddy cow, which causes the same proximal input as the horse. A psychosemantics should provide an explanation (at least in outline, and granted plausible auxiliary hypotheses) of why Oscar's perceptual judgment ("That's a horse", as he would put it) is true, and Twin Oscar's is false. If a theory shows promise in this way, it is not unreasonable to conjecture that a suitably elaborate and detailed descendant of the theory will work for the more complicated

¹¹ It might be denied that, for this particular example, there is an error in Twin Maxwell's experience (and indeed one of us does deny this). For if something looks (pure) red just in case it looks reddish without looking to have a tinge of any other color, and orange is the conjunctive property of being reddish and being yellowish, then the orange square, in looking red to Twin Maxwell, looks to have a property it does have, namely reddishness. On this understanding of the example, Twin Maxwell's experience *underrepresents* the color of the object rather than *misrepresenting* it. Still, even if this is correct, it does not undermine our general point that Twin Maxwell is seriously in error about the colors of things (not only at the level of his beliefs but also at the level of his experiences).

To see this, consider another square Twin Maxwell is seeing which, in his world, is the counterpart to a *red* square that Maxwell is seeing. Given the 45 degree shift in color phenomenology, Twin-Maxwell has a "red-blue" experience. But the square he is viewing is red. So Twin Maxwell's visual experience misrepresents its color. This, at any rate, is how the intentionalist about color qualia will describe the situation (since a red-blue experience, for the intentionalist, is one that represents purple); and Pautz has no quarrel with intentionalism (see 1). Similar misrepresentations with respect to color will occur all over the place. So, even if Twin Maxwell doesn't misrepresent the color of certain orange squares, he clearly is seriously in error about the colors of things generally.

examples. We emphasize that all the leading versions of psychosemantics face extremely difficult problems, but the case of Maxwell and Twin Maxwell is not one of them.¹²

13

¹² Thanks to David Hilbert for very helpful comments on an earlier draft.

13

References

- Block, N. 1999. "Sexism, Racism, Ageism and the Nature of Consciousness". *Philosophical Topics* 26 (1&2), 39-70.
- Bradley, P., and M. Tye. 2001. "Of Colors, Kestrels, Caterpillars, and Leaves". *Journal of Philosophy* 98: 469-87.
- Byrne, A., and D. R. Hilbert. 2003. "Color Realism and Color Science". *Behavioral and Brain Sciences* 26: 3-21.
- Byrne, A., and D. R. Hilbert. 2004. "Hardin, Tye, and Color Physicalism". *Journal of Philosophy* 101: 37-43.
- Hardin, C. L. 1993. *Color for Philosophers* (expanded edition). Indianapolis: Hackett.
- Lewis, D. 1984. "Putnam's Paradox". *Australasian Journal of Philosophy* 62: 221-36.
- Mollon, J. D. 1997. "' . . . On the Basis of Velocity Clues Alone': Some Perceptual Themes 1946-1996". *Quarterly Journal of Experimental Psychology* 50A: 859-78.
- Tye, M. 2000. *Consciousness, Color, and Content*. Cambridge, MA: MIT Press.
- Williamson, T. Forthcoming. "Can Cognition be Factorised into Internal and External Components?" In R. Stainton, ed., *Contemporary Debates in Cognitive Science*, Blackwell.