

Stained Glass as a Model for Consciousness

Mihnea DI Capraru*

Humanities Center and Department of Philosophy,

Syracuse University, Syracuse, NY, USA

Abstract Contemporary phenomenal externalists are motivated to a large extent by the transparency of experience and by the related doctrine of representationalism. On their own, however, transparency and representationalism do not suffice to establish externalism. Hence we should hesitate to dismiss phenomenal internalism, a view shared by many generations of competent philosophers. Rather, we should keep both our options open, internalism and externalism. It is hard, however, to see how to keep open the internalist option, for although transparency and representationalism have not yet definitively established externalism, they have indeed made it quite intuitive. Internalism, by comparison, comes across at first sight as antiquated and ridden with difficulties. This is why I propose the Stained Glass model of consciousness. I do so with two aims: first, to make internalism intuitive in the age of transparency, and second, to show how to resist the many recent anti-internalist arguments. In particular, I argue

*<http://mihnea.capraru.org>

that phenomenal internalism need not be epistemically worrisome, that it is compatible at once with transparency, representationalism, and content externalism, and that although it requires an error theory, this error theory is a harmless one.

Keywords: transparency of experience; phenomenal externalism; internalism; phenomenal consciousness; representationalism

1 Introduction

Uninformed common sense recommends phenomenal externalism, the view that phenomenal properties are properties of non-mental, external objects. Yet when informed by physics and physiology, common sense appears to embrace phenomenal internalism, the view that phenomenal properties are properties of internal, mental entities.¹ Phenomenal internalists have played a prominent role in philosophy since Galilei and Locke and until recently. Nowadays, however, many philosophers have adopted externalism; contemporary phenomenal externalists include the likes of Gilbert Harman (1990); Michael Tye (1992; 1995; 2000; 2002a; 2002b); John McDowell (1994); Fred Dretske (1995; 1996); William Lycan (1996; 2001); Alex Byrne and Michael Tye (2006); Christopher Hill (2009); and arguably Sydney Shoemaker (1994a; 1994b; 2000; 2003) and Uriah Kriegel (2009).

Contemporary phenomenal externalists are motivated to a large extent by the transparency of experience, a phenomenological thesis brought to prominence by Gilbert Harman in 1990:

(Transparency) When we introspect phenomenal properties, we are aware of them only as properties of the objects of our awareness.

Many have inferred from transparency that phenomenal properties not only seem to be but *are* properties of the objects of our awareness. And since nowadays we find it implausible that these objects could be inner entities such as sense data, externalists have concluded that

uninformed common sense was right all along and that phenomenal properties are properties of non-mental, external objects.

Externalists usually pair their view with representationalism; so often do we encounter the two views together that it is easy to casually conflate them. Yet in this paper we shall keep them apart:

(Phenomenal Externalism) Phenomenal properties are properties of non-mental, external objects.²

(Representationalism) Phenomenal properties are part of the contents of our representations.³

We have a compelling reason in transparency to adopt representationalism. Yet we have no direct reason to adopt externalism, for we can accept that we represent phenomenal properties *as* properties of external objects, yet maintain that we are misled by appearances. This is why externalists have supplemented transparency with additional premises; I will discuss and criticize a number of their arguments in Section 3.

If we had in transparency or in representationalism a direct reason to dismiss phenomenal internalism, then perhaps we should do so. Yet we seem to have no such direct reason, and I will argue in Section 3 that we have no compelling indirect reasons either. Again, if externalists could credibly elucidate the nature of phenomenal properties, then we would have a good reason to dismiss internalism. Yet at present externalists and internalists alike are struggling with the issue. It seems therefore that externalists are too quick to dismiss a view that has been shared by many generations of competent philosophers. Given how little we know about consciousness, if we dismiss either internalism or externalism then we restrict our options too early, and in our haste we may rule out the very solution we are searching for.

This, then, is the end I pursue: not to prove internalism or to disprove it, but rather, to show that internalism remains a plausible open option. At first this may seem difficult, because

externalism has come to be perceived as quite attractive, whereas internalism, by comparison, comes across initially as antiquated, ridden with difficulties, and fraught with sceptical consequences. I will argue, nevertheless, that even if we adopt transparency and representationalism, internalism remains plausible. To this end I will introduce a vivid way of thinking about phenomenal internalism, the Stained Glass model of consciousness. I propose the Stained Glass model with two aims: to show how to resist the many recent anti-internalist arguments, and to make internalism intuitive and attractive in the age of transparency.

2 The stained glass model of consciousness

2.1 The model

In this section I describe the Stained Glass model of consciousness. Stained Glass is a model and not a theory, a model by which I intend not to elucidate the make-up of consciousness but to provide a useful way to think about it. In this, again, I aim to prove not that internalism is true, but rather, that it remains plausible even when we accept transparency and representationalism. The Stained Glass model will help us in two ways: On one hand, it will help resist the recent anti-internalist arguments motivated by transparency; and on the other hand, it will vividly and intuitively show how phenomenal internalism is compatible with representationalism and with transparency.

Before I introduce the model let us engage in a couple of preliminary thought experiments.

First thought experiment: I am in a room I have never seen before. Dividing the room, a squeaky clean colourless glass wall. The glass is so flawless and immaculate that I fail to see it. Indeed, so thoroughly do I ignore the glass wall that I walk straight into it.

Second thought experiment: I am at a sculpture exhibition. The curators have placed a white marble statue behind a transparent red glass wall. As before, I fail to see the glass wall; instead, I see the statue as having the glass wall's colour. This causes me to believe that the

sculpture is made of red marble.

Here, now, is the Stained Glass model of consciousness. Picture a spectator under a glass dome. The glass dome is stained – different patches of it have different colours. While regular stained glass is merely translucent, our glass dome is perfectly transparent. And just like the people who sometimes walk straight into squeaky clean glass walls, the spectator cannot see that the glass dome is there.

When we look at objects through transparent coloured glass, the objects take on the glass's colour; in Fred Dretske's terms (1995, 67–68), the objects look to us *phenomenally* like the glass. Ordinarily we do not 'believe our eyes' and the objects do not look to us *doxastically* like the glass. But picture the spectator looking at colourless objects through the stained glass dome. Since the spectator is unaware of the glass, the objects will look to her like the glass both phenomenally and doxastically – that is, the spectator will believe that external objects have the colours of the dome.

Picture now that the dome has this feature: to help the spectator keep track of external objects, the dome tags them with distinct colours and continuously updates its stain structure to track the objects' motions. Suppose, for instance, that two similar, colourless butterflies are flying toward the dome. If the butterflies appeared colourless to the spectator, then she would be unable to tell diachronically which one is which. Yet if the dome tags the butterflies red and yellow, then the spectator will be able to recognize them by their apparent colours.

Just as it is useful for the spectator to keep track of the identities of external objects, it is also useful to determine their kinds. For instance, it is usually far more important to tell that a certain object is a skunk than it is to tell which skunk it is. This is why the dome assigns colour tags not only to external objects but also to their kinds; in particular, the dome tags skunks with their characteristic black-and-white pattern.

Let us use the spectator and the dome – not the spectator alone – to model the phenomenal mind and its experience of the outer world. The spectator models that part of the human

mind which *represents* phenomenal properties, whereas the dome models the phenomenal properties themselves. Although the dome lies outside the spectator's body, let me clarify that I am *not* suggesting that our minds extend beyond our bodies. I might perhaps give this impression because it may seem that the spectator models the entire human being. But this is not so; rather, it is the spectator and the dome together that model the human mind.

As explained on page 3, I call phenomenal externalism the view on which phenomenal properties belong to non-mental, external objects. By phenomenal internalism, on the other hand, I mean the opposite view, on which phenomenal properties belong to mental entities. There is, in the literature, a tendency to conflate internalism in this sense with the stronger doctrine that phenomenal properties are intrinsic features of experience. As regards this latter doctrine I am neutral: If we so desire, we *may* interpret the dome as modelling experience; yet we can also think of the dome as external to experience itself, yet still internal to the mind.

At this point it may seem that we are begging the question on behalf of internalism. We use the dome to model part of the human mind, and we identify phenomenal properties with properties of the dome, hence with mental properties. However, when the dome stains itself to track external objects, it does so by monitoring these objects' optical features, and it is among these external optical features that externalists may wish to identify phenomenal properties. One may worry, therefore, that when we locate phenomenal properties within the dome's glass rather than beyond, we do not prove internalism true but merely declare it so.

Indeed, we do not prove internalism merely by giving a model for it; but neither do we beg the question on its behalf, for we never conclude its truth. Instead, we aim to show that even if we adopt representationalism and transparency, we can still uphold internalism plausibly and intuitively. It is to this latter end that we introduce the Stained Glass model; hence we may stipulate our model any way we want, as long as we conclude not that internalism is

true, but rather, that it is plausible and intuitive. In particular, we may stipulate that it is the stained glass dome that models phenomenal properties, because from here we conclude nothing but this: that *if* phenomenal properties are internal, then it can still be natural for the human mind transparently to misrepresent them as belonging to external objects, just as it mistakenly attributes the red colour of the museum window to the white marble statue beyond it.

Notice, now, how the Stained Glass model conjoins internalism, representationalism, and transparency. The model is internalist, as it locates phenomenal properties in the stained glass dome, and therefore inside the mind. It is, moreover, representationalist, because the spectator *represents* phenomenal properties as properties of external objects. Finally, the model renders experience transparent, because the spectator is aware of phenomenal properties *only* as properties of external objects. The Stained Glass model thus achieves our purpose, to make internalism intuitive in the age of transparency.⁴

2.2 No sense data (and no other epistemically worrisome forms of indirect perception)

To some, it may seem that the Stained Glass model commits us to sense data or to other forms of indirect perception that are epistemically worrisome, i. e., that lend themselves to arguments for scepticism about the external world: ‘Isn’t the doctrine of qualia internalism the last bastion of a widely discredited Cartesian conception of the mind?’ (Byrne and Tye, 2006, 242). I will argue, however, that Stained Glass does not commit us to epistemically worrisome indirect perception – indeed, Stained Glass is *incompatible* with indirect perception in the primary sense; moreover, I will argue that although Stained Glass may allow indirect perception in a secondary sense, this is epistemically harmless.

In the primary sense, indirect perception occurs when perception of one *object* mediates perception of another. In this primary sense, the Stained Glass model is incompatible with

indirect perception. This is because the spectator does not perceive the stained patches of the dome *as objects*, but merely perceives their properties. Therefore, in the primary sense, the spectator perceives external objects directly through the glass. And notice, in particular, that since sense data are supposed to be internal *objects*, they are a special case of indirect perception in the primary sense. Hence Stained Glass is incompatible with sense data.

When we speak of indirect perception, we usually mean indirect perception in the primary sense. However, we may also want to call perception indirect in a secondary sense: this is when perception of a *property* is mediated by perception of another. In this secondary sense, Stained Glass is perhaps compatible with indirect perception. Here is why: The stained glass dome assigns colour tags based partly on the reflectance properties of external objects. This makes it plausible that when the spectator perceives dome colours directly, she also perceives indirectly the correlative non-phenomenal properties of external objects.

If Stained Glass is compatible with indirect perception in this secondary sense, let us see whether we have reason to worry. Indirect perception is worrisome when it allows the sceptic to draw a Cartesian line between our mind and external reality. Once the sceptic draws this line, he or she can claim that we know nothing directly but our own mind, and that we do not know directly but rather need to prove the external world. But the sceptic can only draw the line if there is something distinct on both sides of it: on the right side, our direct perception of mental entities, and on the wrong side, our indirect perception of things in the world. Hence the sceptic needs us to concede that direct and indirect perception constitute distinct, separate entities; furthermore, the sceptic needs us to concede that indirect perception does not confer knowledge (or justify belief), except on some further epistemic condition such as valid inference from direct perception. Only if we concede all of this can the sceptic argue that indirect perception does not confer knowledge. But this means that we can foil the sceptic if we maintain that direct and indirect perception constitute the same and *not* separate entities; if we do so, then the sceptic cannot draw the line and our kind of indirect

perception is epistemically harmless.

It is difficult for the sceptic to draw the Cartesian line for the kind of indirect property perception allowed by the Stained Glass model. When the spectator indirectly perceives an external non-phenomenal correlate of a dome colour, she does so *not by inference* from direct perception, but rather by *the same perceptual act* by which she directly perceives the dome colour itself. Hence the sceptic cannot draw the line between two distinct perceptual acts; instead, the sceptic must draw the line through the heart of one and the same act. This means that the sceptic must overcome three obstacles. First, the sceptic must distinguish between a direct and an indirect aspect of the spectator's perceptual act. Second, the sceptic must argue that, under its indirect aspect, the spectator's perception does not confer knowledge (or justify belief) except on some further epistemic condition, one that can no longer consist in inference. And third, the sceptic must argue that this non-inferential condition fails. The first of these steps is perhaps easy; the next two, however, look rather more difficult.

Moreover, even if the sceptic can do all this, he or she only gets to cast doubt on our perceptual knowledge of the non-phenomenal correlates of phenomenal properties. And this is awfully little doubt: For one thing, we can presumably still acquire knowledge of these non-phenomenal correlates by other, non-perceptual means; indeed, we seem to have far better knowledge of them from science than we could ever get from perceiving them. And for another thing, non-phenomenal correlates are limited and superficial properties; they include, for instance, surface reflectance properties or the concentrations of certain chemical substances. Hence even if the sceptic managed to cast doubt on our perceptual knowledge of these limited, superficial properties, he or she would be far from calling into question our knowledge of the external world.

In this picture, I can see but one ray of hope for the sceptic: Perhaps the sceptic could instigate substantive doubt if we allowed that internal phenomenal properties include not only the likes of colour, taste, scent, etc., but also more consequential things such as apparent

shapes, sizes, or locations. The latter, indeed, would have non-phenomenal correlates that are neither limited nor superficial; hence if the sceptic could cast doubt on our knowledge of the non-phenomenal correlates of internal apparent shapes, sizes, and locations, then he or she could perhaps call into question our knowledge of the external world. However, Stained Glass does not commit us to internal apparent shapes, sizes, or locations; on the contrary, Stained Glass makes it natural to regard apparent shapes, sizes, and locations as external properties that the spectator perceives directly through the glass.

(On a side note, even when we look beyond the Stained Glass model, we seem to have no compelling reason for internalism about apparent shapes, sizes, or locations. To see this, let us review some of the main reasons for phenomenal internalism, and let us evaluate whether these reasons apply to apparent shapes, sizes, and locations. First, phenomenal internalists are often moved by this phenomenon: when they attempt to ascribe colours, scents, or (phenomenal) sounds to external objects, their intuition protests: ‘how possibly?’ Notice, however, that we do not get this vehement intuitive response when we ascribe to external objects features such as apparent shapes, sizes, or locations. Instead, the most we get is puzzlement over their perspectival relativity. And indeed, phenomenal internalists are often moved by the perspectival relativity of phenomenal properties. But this is inconclusive: we can be externalist about apparent shape, size, and location, yet still account for their perspectival relativity without saying anything wildly implausible (for a recent example, see Hill 2009, 128–168). Finally, many internalists are moved by the apparent possibility of inverted spectra: we can imagine people who have their colour perceptions reversed (compared to us) yet still behave the same as everybody. However, it is quite implausible that some people could see squares as circles and circles as squares and nevertheless behave like everybody else.)

Here then is the upshot of our discussion. Stained Glass allows neither sense data nor other forms of indirect perception in the primary sense; in a secondary sense, Stained Glass is perhaps compatible with indirect perception, i. e., with the indirect perception of properties.

Yet this kind of indirect perception seems to be largely harmless, and even if we allow it, we do not seem to provide the sceptic with any dangerous weapons or implements.

2.3 Theoretical implementations of the stained glass model

Stained Glass is an intuitive and effective way of thinking about consciousness, it is a model of consciousness and not a theory. I propose this model to show how internalism relates to transparency and to representationalism, rather than to explain consciousness as such. For the latter purpose, we would need to implement the model as a theory. I will not try to implement the model here; nevertheless, let us look briefly into how to do it and let us examine two potential theoretical implementations, the visual field theory and Russellian physicalism.

First, let us notice a structural restriction on theoretical implementations. Although Stained Glass does not try to say what consciousness consists in, it does make certain assumptions. The model assumes that phenomenal consciousness involves at least two kinds of components. The glass dome models a lower-level kind of component, one that can recognize external objects by non-phenomenal features of the light they reflect, the scents they emit, etc. These low-level components can alter the representations of external objects by updating them to represent the objects as having phenomenal properties. Once updated, the representations are consumed by a higher-level kind of component that corresponds to the spectator. (There may be only one high-level component or there may be many.) The high-level components cannot recognize external objects by the same non-phenomenal features as their low-level counterparts; instead, they recognize them by their apparent phenomenal properties. In this they are like people, which is why they are modelled as human spectators; nevertheless, the high-level components need *not* be homunculi or bear any further resemblance to human beings.

The low-level components fill the theoretical role modelled by the stained glass dome.

The dome has a geometrical structure that we must decide whether to be realists about. For instance, we can be realists about the dome's structure if we implement Stained Glass within the visual field theory, whereas we are non-realists if we choose a Russellian physicalist implementation. Let us investigate the advantages and disadvantages of these two options.

In 1989, Paul Boghossian and David Velleman advocated the visual field theory. According to Boghossian and Velleman, colours are properties of the visual field, properties that we misrepresent as properties of external objects. We can implement Stained Glass within the visual field theory if we let the dome be implemented by the visual field and the spectator be implemented by whatever (mis)represents the visual field's colours.

(On a side note, if we implement Stained Glass within the visual field theory then we account only for *visual* phenomenal properties and not for phenomenal properties in general. Recall, however, that we may posit multiple low-level components to play dome-like roles; hence even if we adopt the visual field implementation we can still supplement it with further implementations for other sensory modalities.)

The dome is geometrically akin to the visual field; therefore, once we adopt the visual field implementation we become realists about the dome's structure. There is something we gain this way and something we lose. Here is what we gain: we can explain why colours appear to be essentially extended (see Shoemaker, 1994a, 25; Shoemaker, 1994b, 295). And here is what we lose: it is hard to tell just *what* could exemplify such a dome-like geometrical structure. We cannot attribute the dome's structure to the brain states that constitute or underlie visual experience. Some of us are dualists, but they too will find it hard to ascribe a dome-like structure to an immaterial mind. Perhaps we could make the visual field a non-existent, Meinongian entity; then we would be free to ascribe to it the structure of the dome. Granted, it is not clear that we would still count as internalists, but perhaps it is less important to be internalists than it is *not* to be externalists. – More damningly, however, we would need to burden ourselves with the implausibility of Meinongianism. To sweeten the pill, one could

point out that externalists, too, seem to need Meinongianism (e. g., Harman, 1990, 37–38; Lycan, 2001, 18–19). But we should not commit to an implausible theory just because the rival view does the same.

Let us now look at a way to implement Stained Glass without becoming realists about the geometry of the dome. Consider the view commonly referred to as Russellian monism or Russellian physicalism (Russell, 1948, 224–231; for recent formulations – not always endorsements – see Chalmers 2002, 2009; Strawson 2003; Montero 2010; Pereboom 2011, 85–122). On this view, phenomenal properties are intrinsic properties of our brain states, properties that we can only access through introspection and not through scientific knowledge (through the latter we can only access relational properties). Here is how we can implement Stained Glass within Russellian physicalism: First, we let the dome be implemented by the relevant brain states. And second, we stipulate that people can perceive the phenomenal properties of their brain states but *not* the brain states themselves. Rather, they perceive the phenomenal properties of brain states *as* properties of external objects. Since our brains have no dome-like structure, if we choose a Russellian implementation we are no longer realists about the structure of the dome, hence we no longer need to explain what exemplifies it. Yet on the downside, it becomes harder to explain why colours appear to be essentially extended.

2.4 A note on Occam's Razor

One may worry that we can only countenance the Stained Glass model if we posit the existence of a mysterious new kind of entity, of mental 'stained glass.' Perhaps that would be hard to justify. The model, however, requires no such ontological commitment, and this is because stained glass is not an entity but a theoretical role

As we saw in the previous subsection, we may fill the theoretical role of stained glass with anything in the mind to which we are willing to ascribe phenomenal properties. Thus the Stained Glass model adds no ontological requirements to those of the particular kind of

phenomenal internalism at issue.

What Stained Glass does require, instead, is not a matter of ontology but of psychosemantics. Stained Glass requires that our phenomenal representations have what we may call mixed content, i. e., that they mix together certain objects with other objects' properties. We have seen this happen in the case of the white statue and the red glass, where our representations mix together the statue with the colour of the glass. In the Stained Glass model, phenomenal representations mix together external objects with properties of objects internal to the mind.

3 Five arguments against internalism and how the stained glass model resists them

I aim to show not that phenomenal internalism is true, but rather, that it remains plausible and intuitive in the age of transparency. To this end I need not give positive arguments for internalism; I do need, however, to defend it from the arguments brought against it, which I shall do in what follows.

3.1 The argument from content externalism

Phenomenal externalists often argue that their view follows from representationalism and content externalism (e. g., Dretske 1996, 143–145; Lycan 2001, 21; Byrne and Tye 2006, 241–242; see also Kriegel, 2009, 64, who does not endorse the argument). Here is their argument in outline:

1. The phenomenal properties we experience are determined by the contents of our perceptual representations. (Representationalism⁵)
2. The contents of our perceptual representations depend on external factors. (Content Externalism⁶)

3. The phenomenal properties we experience depend on external factors. (from 1, 2)

∴ Phenomenal properties are external to the mind. (from 3)

I contend that this argument is mistaken; the argument equivocates between what is external (or extrinsic) to representation and what is external to (or outside of) the mind.

Before I elaborate on the distinction, let me illustrate it. Picture the spectator seeing a colourless butterfly through a red glass patch that tracks its motion. Recall that we model the mind not by the spectator alone, but by the spectator together with the dome: the spectator models representation, the dome models representation's phenomenal content. Assume, now, content externalism; i. e., assume that when the spectator represents an occurrence of the property red, this is in virtue of relations among the spectator, the red glass patch, and perhaps other things. Even so, notice that the phenomenal property red occurs within the dome, hence within the mind. Thus we may at once be content externalists yet phenomenal internalists.

Let us now explicitly distinguish the two senses of 'external' at issue. First, take 'external' as in 'content externalism.' We often summarize content externalism under Hilary Putnam's pithy slogan 'meanings ain't in the head' (paraphrased from 1975, 227). But while we can use this slogan to approximate content externalism about representations of external objects, we must not confuse the slogan with the view. Content externalism does not *strictly speaking* maintain that meanings are not in the head. Rather, it maintains that meanings are not *in the representations*; in yet stricter terms, content externalism maintains that the contents of representations depend on relations between the representations and their extensions:

A psychosemantic feature of a representation is external_{CE} iff it depends on relations between the representation and its extension.⁷

This, however, is not how we read 'external' in 'phenomenal externalism.' Instead, we read it thus:

A phenomenal property is external_{PE} iff it occurs outside the mind.⁸

Notice, now, that we cannot infer from representationalism and content externalism that phenomenal properties are external_{PE}. All we know from representationalism is that phenomenal properties are part of the contents of representations, and all we know from content externalism is that the contents of representations depend on relations between the representations and their extensions. And while their extensions often lie in the environment, in the history, or elsewhere outside the mind, it is perfectly possible that sometimes they lie within the mind itself.

Finally, let me illustrate, on the case of Byrne and Tye 2006, how the argument from content externalism equivocates between the two senses of ‘external:’

[Phenomenal internalism, content externalism, and representationalism] are inconsistent. ... 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 ..., 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.
(Byrne and Tye, 2006, 241)

To show that phenomenal internalism is incompatible with the conjunction of content externalism and representationalism, Byrne and Tye claim to derive a contradiction from the conjunction of the three. In this, however, they equivocate between the two senses of ‘external’: ‘Given a specific externalist theory of content ... we can plausibly construct different environments for each [of x and y], such that the theory predicts that x and y ’s colour experiences have the same content’ (Byrne and Tye, 2006, 241). According to this crucial step in

Byrne and Tye's argument, we can conclude, from content externalism, that content supervenes on the relation between the experiencers – x and y – and their external environments. But as we have seen, we cannot conclude this much; instead, all we can conclude is that content supervenes on the relation between x and *yes representations* and their extensions. And since these extensions could be mind-internal, Byrne and Tye cannot infer, from content externalism and representationalism, that we can determine x and y to experience the same qualia if only we suitably manipulate their external environments.

3.2 The massive error argument

If internalism is right that phenomenal properties are mental, then in this respect experience is misleading. Proponents of the Massive Error argument contend that it is unacceptable to attribute this mistake to experience (e. g., Shoemaker, 1994a, 25–26; Tye, 2000, 46; Tye, 2002a, 139). According to Michael Tye, it is 'just not credible' and 'totally implausible to hold that visual experience is systematically misleading' (2000, 46). I will try to make it credible that experience could be systematically misleading with respect to the location of phenomenal properties.

Consider, for one thing, that experience belongs to the same epistemic family as uninformed common sense: both are simple, natural, and uneducated. And as we already know from physics and other sciences, uninformed common sense is often wrong in systematic ways. For instance, uninformed common sense thinks moving objects have an inherent tendency to slow down; since Descartes and Newton we understand that this is not so. Again, uninformed common sense thinks heavier objects fall more swiftly; since Galilei we know this is not so. Experience, as I said, is epistemically akin to uninformed common sense, and since uninformed common sense can be wrong in systematic ways, it is rather plausible that experience too can be systematically wrong about some things.

And indeed, here is one way in which experience is systematically wrong. On windless

days, experience indicates that the space surrounding us is empty. Moreover, when the air pressure is close to one standard atmosphere, experience indicates that we aren't acted on by any force from this apparently empty space. Yet the space is not empty but filled with transparent air. Even more strikingly, the air is acting on us with a force so powerful that it quite literally keeps us whole: if the air ceased to exert pressure on us, we would explode. Since experience can be wrong about something as ubiquitous and consequential as air pressure, we cannot *a priori* rule out that it might be wrong about the location of phenomenal properties.

Here is a third argument. Suppose the externalists are right. Then many since Galilei and Locke were wrong: Leibniz, Hume, Russell, etc. But none died of it, nor did they suffer any significant non-epistemic damage. Or suppose, instead, that Galilei and Locke were right: then the externalists are wrong, but they won't die of it either. In Stained Glass terms, as long as the dome does its job, the spectator will be fine. If the spectator were colour-blind, then she would have a practical problem; but she has no practical problem just because she cannot tell that phenomenal properties are in her head and not out there. And since this epistemic ability does not help in practice, we have no reason to expect it to have evolved such as to be embedded in experience.

Finally, some may worry that if we accept that experience mislocates phenomenal properties then we open the door to scepticism. But if experience mislocates phenomenal properties then this is a well-contained error that won't spill over into the rest of knowledge. For suppose that the externalists are right: Even so, internalist physicists won't get their physics wrong, nor will internalist pilots crash their aeroplanes. Or suppose that the externalists are wrong: Even so, externalist physicists will do just fine, and so will externalist pilots. Whether or not the externalists are wrong, whether or not experience mislocates phenomenal properties, little of our knowledge hangs on it.

3.3 The argument from physicalism

Phenomenal consciousness poses a notorious challenge to physicalism, and some physicalists have become externalists because they believe they can only account for consciousness if phenomenal properties are properties of external, physical objects. According to William Lycan, phenomenal externalism 'is the only very promising way to preserve materialism while accommodating qualia. For the only viable alternative exegesis ... seems to be belief in actual Russellian sense-data or at least in immaterial properties' (Lycan, 2001, 19). Yet this is not so. As we have seen in 2.3, we can implement the Stained Glass model within the framework of Russellian physicalism, hence we can be internalists and physicalists at once.⁹

3.4 The object–property awareness argument

In 2009, Christopher Hill argues that

... if appearances are properties that we are aware of in visual experience, and the only objects of visual awareness are ordinary physical objects, then it must be true that appearances are properties of ordinary objects. Surely it could not be true that we are visually aware of properties without being aware of the objects that instantiate them. (Hill, 2009, 144)

I dispute the key premise of this argument. To see that we *can* be visually aware of properties without being aware of the objects that bear them, recall the squeaky clean red glass wall and the white marble statue. Although unaware of the glass wall, we are visually aware of its property red, which we mistakenly attribute to the statue. Hence *pace* Hill we can be visually aware of properties while unaware of the objects that bear them.

3.5 The argument from puzzling arbitrariness

In 1992 and 2002b, Michael Tye argues that if internalism is true, then it is ‘an impenetrable mystery’ that colours feel as they do (1992, 175): ‘Why, for example, does having a visual experience of blue “feel” the way it does and not some other way?’ (1992, 175). Armed with the Stained Glass model, though, we can dispel the mystery.

Recall that the dome ascribes colours to external objects in order to *tag* them. If so, then it is no mystery why it feels one way rather than another to see blue-looking objects. This is not because we can answer the question informatively; on the contrary, it is *of course* arbitrary to tag blue-looking objects with one feeling and not another. If colours are tags, then as such they are inherently and unmysteriously arbitrary.

4 Conclusion

The Stained Glass model of consciousness intuitively conjoins phenomenal internalism, representationalism, and the transparency of experience. While the model is not obviously accurate, neither is it obviously inaccurate. This allows us to keep open our internalist option on solving the problem of consciousness: On one hand, the model makes internalism plausible and intuitive in the age of transparency, and on the other, the model can resist the recent onslaught of anti-internalist arguments. Moreover, it turns out that we can keep open the internalist option without committing to sense data in particular or to epistemically worrisome indirect perception in general; and finally, we can do so while committed to content externalism and to physicalism.

Since Stained Glass is not a theory but a model, it is not on its own a solution to the problem of consciousness. While the model can help us better to understand the problem, in order to seek an internalist solution we must proceed by surveying the space of potential theoretical implementations of the Stained Glass model.

Acknowledgements

I am grateful to Robert Van Gulick for important feedback on several aspects of this paper, to Neil Mehta for a number of insightful observations, to Amy Kind, who commented on the paper at the Eastern APA in 2012, and to André Gallois and Clyde Hardin, who have read early drafts.

Notes

¹On another formulation, phenomenal internalism is the view that phenomenal properties supervene on the intrinsic physical properties of the perceiver. Although this overlaps largely with my formulation, it is not entirely equivalent to it.

²Among the objects we call external, let us include somatic objects in order to accommodate bodily phenomenal properties.

³Philosophers often distinguish strong representationalism from weak (Tye, 2002a, 137; Byrne and Tye, 2006, 241, 254; Kind, 2007, 405–406; Kind, 2010, 904). Weak representationalism claims merely that phenomenal properties *supervene* on the contents of our representations. In this paper, by ‘representationalism’ I mean strong representationalism.

⁴In 2009, Bernard Molyneux has proposed another thought experiment intended to model the transparency of experience. Molyneux has done so to defend a theory that I think suffers significant flaws. This, however, is not the place to criticize the theory; rather, let me criticize the thought experiment and let me show why we are better off with Stained Glass.

Molyneux’s thought experiment features a character named Bob. Bob is fitted with imperceptible ‘televsual contact lenses’ (Molyneux, 2009, 123). These TV contacts are not quite lenses and they are not transparent. Instead, they film external reality with small cameras and they display pictures of external reality on their inner surfaces. So it comes that Bob, unbeknownst to him, perceives external reality indirectly (Molyneux, 2009, 123).

Molyneux maintains that TV contacts effectively model the transparency of experience. Let us, however, consider a Stained Glass alternative, one that I will argue is preferable. Consider stained contacts. Stained contacts work just like our transparent stained glass dome, except that they nest imperceptibly in Bob’s eyes. Compared with TV contacts, stained contacts have multiple advantages: First, since we are trying to model transparency, it helps that unlike TV contacts, stained contacts are genuinely transparent. Second, with transparent stained contacts it is obvious that Bob can perceive external objects rather than merely his own contacts;

with opaque TV contacts, this is not obvious (notwithstanding Molyneux's arguments in 2009, 124–125). Third, with stained contacts it is obvious that Bob is experiencing phenomenal properties as properties of external objects, whereas with TV contacts it is not: with TV contacts, it might be that Bob experiences phenomenal properties as properties of the TV pictures, *and* mistakes the TV pictures for regular external objects. And finally, with stained contacts Bob perceives external objects *directly*; with TV contacts, however, he perceives them indirectly in a strong sense that seems to commit Molyneux to something similar to sense data (Molyneux, 2009, 123).

⁵For this premise, notice that even weak representationalism will serve.

⁶We can distinguish shades of content externalism along at least two dimensions: First, how many – and which – representations have their contents determined extrinsically, and second, whether contents are determined entirely or only in part extrinsically. To mount the argument from content externalism, phenomenal externalists need to assume at least this much: that all representations with phenomenal content have their content determined extrinsically, and that the phenomenal component of their content is determined extrinsically at least in part.

⁷It can be tricky to formulate content externalism rigorously; my formulation draws from Robert Van Gulick (2004, 258–261).

⁸To accommodate the extended mind hypothesis, we might add: or inside the extended mind, yet outside the organism.

⁹To be sure, proponents of Russell's view do not always intend it as a form of physicalism; indeed, Russell himself intended it as a form of neutral monism. Nevertheless, authors like Montero (2010) or Pereboom (2011) have recently proposed physicalistic versions of Russellian monism.

References

- Boghossian, P. A. and Velleman, J. D. (1989). Colour as a Secondary Quality. *Mind*, 98:81–103.
- Byrne, A. and Tye, M. (2006). Qualia Ain't in the Head. *Noûs*, 40:241–255.
- Chalmers, D. J. (2002). Does Conceivability Entail Possibility? In Szabó Gendler, T. and Hawthorne, J., editors, *Conceivability and Possibility*, pages 145–200. Oxford University Press.

- Chalmers, D. J. (2009). The Two-dimensional Argument Against Materialism. In McLaughlin, B. P., editor, *The Oxford Handbook of Philosophy of Mind*, pages 313–335. Oxford University Press.
- Dretske, F. (1995). *Naturalizing the Mind*. MIT Press.
- Dretske, F. (1996). Phenomenal Externalism or If Meanings Ain't in the Head, Where Are Qualia? *Philosophical Issues*, 7:143–158.
- Harman, G. (1990). The Intrinsic Quality of Experience. *Philosophical Perspectives*, 4:31–52.
- Hill, C. S. (2009). *Consciousness*. Cambridge University Press.
- Kind, A. (2007). Restrictions on Representationalism. *Philosophical Studies*, 134:405–427.
- Kind, A. (2010). Transparency and Representationalist Theories of Consciousness. *Philosophy Compass*, 5:902–913.
- Kriegel, U. (2009). *Subjective Consciousness: A Self-Representational Theory*. Oxford University Press.
- Lycan, W. G. (1996). *Consciousness and Experience*. MIT Press.
- Lycan, W. G. (2001). The Case for Phenomenal Externalism. *Philosophical Perspectives*, 15:17–35.
- McDowell, J. (1994). The Content of Perceptual Experience. *The Philosophical Quarterly*, 44(175):190–205.
- Molyneux, B. (2009). Why Experience Told Me Nothing about Transparency. *Noûs*, 43:116–136.
- Montero, B. (2010). A Russellian Response to the Structural Argument Against Physicalism. *Journal of Consciousness Studies*, 17:70–83.

- Pereboom, D. (2011). *Consciousness and the Prospects of Physicalism*. Oxford University Press.
- Putnam, H. (1975). The Meaning of 'Meaning'. In *Mind, Language and Reality: Philosophical Papers, Volume 2*, pages 215–271. Cambridge University Press.
- Russell, B. (1948). *Human Knowledge: Its Scope and Limits*, chapter "Mind and Matter", pages 224–231. Simon and Schuster, New York.
- Shoemaker, S. (1994a). Phenomenal Character. *Noûs*, 28:21–38.
- Shoemaker, S. (1994b). Self-Knowledge and 'Inner Sense.' Lecture III: The Phenomenal Character of Experience. *Philosophy and Phenomenological Research*, 54:291–314.
- Shoemaker, S. (2000). Introspection and Phenomenal Character. *Philosophical Topics*, 28:247–273.
- Shoemaker, S. (2003). Content, Character and Color. *Philosophical Issues*, 13:253–278.
- Strawson, G. (2003). Real Materialism. In Antony, L. M. and Hornstein, N., editors, *Chomsky and His Critics*, pages 49–88. Blackwell Publishing.
- Tye, M. (1992). Visual Qualia and Visual Content. In Crane, T., editor, *The Contents of Experience: Essays on Perception*, pages 158–176. Cambridge University Press.
- Tye, M. (1995). *Ten Problems of Consciousness: A Representational Theory of the Phenomenal Mind*. MIT Press.
- Tye, M. (2000). *Consciousness, Color, and Content*. MIT Press.
- Tye, M. (2002a). Representationalism and the Transparency of Experience. *Noûs*, 36:137–151.
- Tye, M. (2002b). Visual Qualia and Visual Content Revisited. In Chalmers, D. J., editor, *Philosophy of Mind: Classical and Contemporary Readings*, pages 447–456. Oxford University Press.

Van Gulick, R. (2004). Outing the Mind—A Teleopragmatic Perspective. In Schantz, R., editor, *The Externalist Challenge*, Current Issues in Theoretical Philosophy, pages 255–284. De Gruyter, Berlin.

DRAFT