Working draft v.5 — March 2025 — Comments are welcome!

What Constitutes Phenomenal Character?

Murat Aydede
Department of Philosophy
University of British Columbia
murat.aydede@ubc.ca

Abstract. Reductive strong representationalists accept the Common Kind Thesis about subjectively indistinguishable sensory hallucinations, illusions, and veridical experiences. I show that this doesn't jibe well with their declared phenomenal externalism and argue that there is no sense in which the phenomenal character of sensory experiences is constituted by the sensible properties represented by these experiences, as representationalists claim. First, I argue that, given general representationalist principles, no instances of a sensible property constitute the phenomenal character of the sensory experience that represents them. Second, I argue that, with two very plausible assumptions in place, no sensible property qua universal can constitute the phenomenal character of experiences either. At the end, I offer an alternative picture that is consistent with a naturalist psychosemantics for sensory experiences without embracing phenomenal externalism.

Suppose Sam is intently looking at a blue and round ball (call the ball, Tom) in front of her against a roughly uniform neutral background in good day light. Let's say that Sam is having a veridical visual experience VE1 as of something being blue and round.

Let *B* be the property complex, being blue and round:

$B = \lambda x$ (x is blue & x is round)

B is a type, a universal, and is instantiated by Tom. Call this particular instance of *B, b1*. Sam is seeing *b1* — the instantiation of *B* (by Tom).

Sam is intently and carefully looking at Tom for about 5 seconds. It is natural to say that she is aware of *b1*. This is a direct *de re* awareness, if anything is. Sam's visual experience seems to put Sam directly in contact with *b1* (and with Tom, of course). This experience has an immediate seemingly world-disclosing presentational character, which has a certain phenomenological profile that we may call its *phenomenal character* — there is something it's like to undergo this particular experience which seems to immediately present *b1* to Sam. What is the relation of VE1's phenomenal character to *b1*? Is this relation merely

causal, or rather is it constitutive (partly or fully[1])? If the answer is the latter, I'll say that Sam's experience is instance-involving.

Reductive strong representationalism is meant to be a view committed to a form of phenomenal externalism, according to which the phenomenal character of sensory experiences is constituted by the character of (non-conceptually or sensorially) represented sensible properties. On this view, physical duplicates being in the same state may differ in the phenomenal character of their respective experiences — if their states somehow sensorially represent different sensible properties; or physically type-distinct states of the duplicates can have the same phenomenal character if they somehow represent the same sensible property. According to representationalists,[2] the phenomenal character of a sensory experience doesn't supervene on the narrow physical constitution of the experiencing subject. Thus, the represented sensible properties are constitutive of phenomenal character. So, one would expect that a representationalist of this sort would answer the above question by saying that *b1* constitutes the phenomenal character of Sam's veridical visual experience — indeed they often say that the phenomenal character *is identical* to the represented content or feature. First, I will argue that this externalist claim about property *instances* cannot be true given what representationalists have to say about hallucinations. Second, I will show that sensible properties *qua universal* cannot constitute the phenomenal character either. My overall conclusion will be that phenomenal externalism is false, and that if representationalism entails such an externalism, it too is false. In conclusion, I'll offer an alternative picture that is (weakly) representationalist but internalist and naturalist.

1 The role of property instances

Suppose, at the end of the five seconds, God intervenes and takes over the causal route stimulating Sam's brain in such a way that Sam doesn't notice anything when God removes Tom. It's a smooth transition. Sam is now having a subjectively indistinguishable hallucinatory experience, HE1, as of a blue and round ball in front of her. We can extend the thought experiment. Another five seconds pass and God puts a qualitatively identical but numerically distinct ball (call it, Bill) back in where Tom had been when Sam was looking at it and lets Bill take over the causal operation on Sam: Sam is now having a VE2 with another instantiation of *B, b2*. Another five seconds pass and God intervenes again in the same way, smoothly removes Bill while maintaining the neural activity in Sam associated with *B*. Sam is now having another hallucinatory experience, HE2. Finally, we can suppose that after another 5 seconds, God puts Tom back where it was twenty seconds ago and lets the causal stimulation be controlled by Tom again. Sam is now having another

[1] Metaphysical constitution may be partial or whole. I'll omit this qualification from now on except when it matters — see below.

[2] From now on, when I talk about "representationalists" without qualification, I'll have in mind *reductive strong representationalists* in mind such as Dretske (1995) and Tye (2013, 2014), among others.

veridical experience, VE3, which makes her aware of *b1*. Sam has no clue about what is going on. (The following diagram may help.)

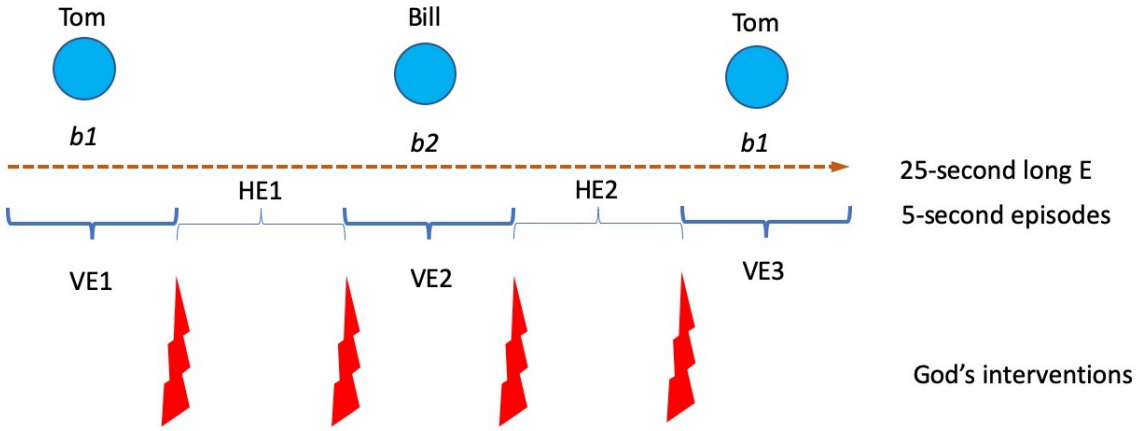

Figure 1. Imagined scenario.

By stipulation, the phenomenal characters of these smoothly connected experiential episodes (VE1, HE1, VE2, HE2, VE3) are subjectively indistinguishable. Indeed, throughout 25 seconds, Sam falsely but justifiably believed that she was looking at a blue and round ball that remained identical. The Common Kind theorists in philosophy of perception think that the subjective indistinguishability in such cases is to be explained by the presence of positive phenomenology: VE1, HE1, VE2, HE2, VE3 all have the same phenomenal character. These five episodes share a common fundamental phenomenological or experiential kind. Representationalists accept the Common Kind Thesis.

According to representationalists, the phenomenal character of sensory experiences is exhaustively a matter of what sensible properties are represented in the experience, whether or not the experience is veridical. In our example, all the five sensory episodes represent *B* as instantiated, and it is this fact that determines the identity of the phenomenal character of Sam's experience during the 25 seconds she was intently looking at the "ball." This entire experience — call it *E* — is an experience that remains phenomenally identical throughout 25 seconds, where *b1, b2* are the instances causally related to *E* during the first, third and fifth 5-second periods (therefore making VE1, VE2, VE3 accurate), while *E* has no actual objects or instances during the second and fourth periods, which makes HE1 and HE2 hallucinatory. Thus, if *E* has the same phenomenal character throughout, this phenomenal character cannot constitutively involve *b1* and *b2*. So, if the veridicality of VE1 is what partly makes for Sam's awareness of *b1*, the phenomenal character of this awareness (VE1) cannot be constituted by *b1* — similarly with VE2 and VE3. VE1 is not in this way instance-involving. The relation of VE1 to *b1* is *only causal*. So, the phenomenal identity

of *E* is not instance-involving at all. In fact, given the way the thought experiment is set up with "external" physical objects and mind-independent sensible properties, we can generalize: for Common Kind theorists who believe that subjectively indistinguishable veridical, illusory and hallucinatory experiences share a common positive phenomenal core, even in cases where the experiences are veridical, the sensible property instances the subjects are aware of never constitute the phenomenal character of these experiences.

According to representationalists, what metaphysically fixes the identity of the phenomenology of *E* is this: *E* represents *B* (as instantiated)[3]. *E* is veridical when it is sustained by an appropriate causal/informational link to phenomenologically irrelevant instances of *B*, and non-veridical otherwise.

Generalizing, the situation is the same with *all* sensory experiences: their phenomenology is never constituted by the instances of the sensory properties or property-complexes they veridically represent (when they do). Veridical sensory experiences, according to representationalists, are not only not object-involving, but are also not instance-involving.

This result may come as a surprise to some.[4] For phenomenal externalism seems to demand that veridical sensory experiences are instance-involving. But recall the ease with which many representationalists claim that Sam is sensorially aware, *de re*, *of* a (locally uninstantiated) universal, *B*, while having HE1 and HE2.[5] This sounds mysterious and puzzling, but I'll assume that all they mean with this is that in hallucinatory experiences like HE1 and HE2 there is, in an obvious sense, still sensory representation: Sam's hallucinatory experiences still represent a (locally) uninstantiated sensible property complex, namely *B*, a universal — it just *misrepresents it as instantiated*.[6] It is this fact, according to representationalists, that determines the phenomenal character of the hallucinations — no property instances are ever involved. But given the Common Kind Thesis, this phenomenal character is also the *very same character* of the veridical episodes. Thus, even the veridical

---

[3] Sensory experiences as of a sensible property *F* are always awareness of *F*-instances when veridical. The non-existence of *F*-instances is what makes *F*-experiences illusory or hallucinatory, i.e., non-veridical. So, *pace* Tye (2014), Sainsbury (2019) and Gottlieb & Rezaei (2021), sensory representation of sensible properties (when deployed in perception rather than, say, imagination) has always assertoric force, thus accuracy conditions.

[4] For instance, Pitt (2017) and Gow (2018) seem to write with the assumption that for representationalists property instances are phenomenology-constituting in veridical experiences. Over the last few years, I have received many useful referee reports on previous versions of this work. Many referees have accused me of misrepresenting representationalists on just this point, while other referees have dismissed my claim in this section as old hat.

[5] See, for instance, Tye (2015: 485, 2013: 51–52), Dretske (1999: 107).

[6] In my view, there cannot be a sensory awareness *of* a universal uninstantiated. The function of sensory awareness is to *detect* property instances in the physical (including, bodily) environment — see next section. There is no such thing, properly speaking, as *detection* of uninstantiated properties. I simply take this talk of sensory awareness of universals as expressing (PC) — see below.

episodes don't involve instances of sensible properties as the constitutive determinants of the experiences' phenomenal character.

A different way to express the main point is this: whatever the phenomenal character of any sensory experience involves, it involves it *essentially*. But, given what representationalists say about hallucinatory experiences, property-instances are not *essentially* involved in the constitution of phenomenal character of any experiences. Hence, no sensory experiences (veridical or not) are ever instance-involving.

2 Phenomenal character as the property '*representing P*'

But then we seem to have a puzzle. If the phenomenal character of an experience of a sensible property *P* is never constituted by the instances of *P*, in what sense is it constituted by the property *P* (*qua* universal)? Indeed, what does it mean to say that the phenomenal character is constituted by *P* but not by its instances? For surely, as pointed out at the start, phenomenal externalism requires that sensible properties themselves are constitutive of the phenomenal character of sensory experiences that represent them. Representationalists keep telling us that it is the represent*ed* properties that constitute the phenomenal character. If it is not their instances, what is it for the sensible properties *qua* universals to constitute sensory phenomenology? It seems to me that the only plausible thing to say at this juncture is this:

> (PC) The phenomenal character of a token sensory experience, *s*, as of a sensible property *P* at time *t* is constituted by the fact that *s* possesses the intentional property of sensorially *representing P* (as instantiated) at *t* — whether or not *s* is veridical.

One might think that the sense in which such a view is phenomenal externalist is that *P* is a property that can be instantiated only by external physical (mind-independent) objects. But there is more to this claim as we will see in a moment. So, no instances of *P* are ever constitutive of sensory experiences as of *P*. All that is needed for a sensory state, *s*, to have the relevant phenomenal character at a time is that *s* be representing *P* (as instantiated) at that time — that is, *s* have the property at *t* of *representing P*. Veridicality, but not the phenomenal character, of *s* comes with the causation of *s* by an appropriately related *instance* of *P*. We may even say that *s* is what makes the subject sensorially *aware of* the relevant instance of *P* when *s* is veridical, i.e., appropriately caused. But the phenomenal character of *s* is not constituted by the relevant P-instance. Representationalists may point out that in veridical cases the subject is aware of the properties themselves *as well as their instances*. But again, the point is that the *phenomenal character* of such sensory awarenesses is *solely* due to the relevant states '*representing* the sensible properties *qua* universals — not due to the particular instances that are merely causally involved in the awareness. This much seems clear given what representationalists say about hallucination and their acceptance of the Common Core thesis.

Note that, if I'm right so far, the phenomenal character of a sensory experience as of *P*, *is* a property *of* the experience: it is the property of sensorially *representing P*. Having *this* property is what metaphysically constitutes the phenomenal character of a sensory experience as of *P*. In whatever sense we have introspective access to the phenomenal character of such an experience, it is to this property (*representing P*) that we have access, not just to the property, *P*.

Representationalists sometimes also say things like this:

> (PC!) The phenomenal character of a token sensory experience, *s*, as of a sensible property *P* at *t* is constituted wholly by *P* (the represented universal) — or, *is* just P!

It is completely obscure how to make sense of such claims. Suppose the sensible property in question is an *instance* of a particular shade of blue (let it be unique blue, UB) that Sam is aware of during VE1. According to representationalists, this property, UB, is a physical surface property, say, a certain set of surface spectral reflectances, SSR_{UB}. Whether or not Sam is hallucinating during *E*, representationalists claim that the phenomenal character of Sam's (colour) experience remains identical. But if SSR_{UB} *is* the phenomenal character of *E*, it would of course be completely unsurprising that this physical property (*qua* universal) has been self-identical and remaining identical — whatever that means. But of course! Nobody would take this claim to be making a philosophically controversial or even interesting point. Therefore, when representationalists make claims of this sort (PC!), we will interpret them as meaning (PC).

Our next task, then, is to understand what sorts of facts constitute a state's representing *P* (even when *P* is not locally instantiated). But it would be useful to summarize our discussion so far and draw some lessons before we do that. The most important point to keep in mind is that the phenomenal character of even normal veridical experiences of sensible properties *P* is not constituted by the instances of *P*. The instances are causally/informationally implicated in generating these experiences, and therefore, in this *causal* sense, they determine what experiences with what phenomenal character to be tokened. But the phenomenal character itself is metaphysically constituted by a property that doesn't involve the instances of *P*. The phenomenal character of an experience as of *P* is metaphysically constituted by the experience's having the property of *representing P*, according to representationalists. The instantiation of this intentional property (representing *P*) by a sensory state doesn't metaphysically require the simultaneous existence of any *P*-instances anywhere. Indeed, when Sam is aware of *b1* during VE1, the phenomenal character of her visual experience has, constitutively, nothing to do with *b1* (or, for that matter, with *B qua* being wholly present in *b1*).

Given that Sam is a *bona fide* member of human species, her internal physical/functional constitution is metaphysically sufficient to instantiate sensory states representing *B*,

whether or not there are any instances of it around. To this extent, then, Sam's internal constitution is metaphysically sufficient for her to have experiences with the phenomenal character that is here identified with *representing B*. Without any further externalist account of what the possession of this intentional property comes to, we don't yet have a phenomenal externalist position.

3 Naturalistic psychosemantics for "*representing P*"

The project now is to understand how externalism may arise out of the representationalist account of intentional facts. For reductive or naturalist representationalists, the intentional facts (i.e., the possession of the property of *representing P* by sensory states) concern some combination of facts about causal co-variation, indication, teleological function, tracking, etc. For our purposes, a simplified Dretskean version will suffice (Dretske 1995: 14ff):

> (a) The sensory state token *s* has the phenomenal character it has in virtue of the fact that it *systemically represents P*.
>
> (b) *s* systemically represents *P* in virtue of the fact that it's a token of a state type *S* whose function is to indicate (track, carry information about) instances of *P*.

Thus:

> (c) *s* has the phenomenal character it has in virtue of the fact that it's a token of a state type *S* whose function is to indicate (track, carry information about) instances of *P*.[7]

The indication function here is entirely causal/nomological with a certain historical selection condition (depending on how one understands the notion of function involved).[8] Given that the indication function isn't sufficient to make *P*-instances *constitutive* of phenomenal character of *S* tokens *now*, we can ask: is there any reason to think that tokens of *S* have had their phenomenal character constituted by *P*-instances in the evolutionary history during which the state type *S* was selected because its tokens regularly indicated instances of *P*? The answer clearly is *No*. Information transmission works, roughly, by there being a lawful causal correlation between instantiations of two properties, *P* and *S*: when the channel conditions are right, (only) *P*-instances causally determine *S*-tokens. That, then, constitutes *S*-token's indicating a *P*-instance. If any *S*-token throughout the

---

[7] More is required here for the emergence of phenomenal character such as the fact that the state type needs to have certain format constraints (e.g., non-conceptual, imagistic, analog, etc.). In particular, the sensory representation types, the *S*'s, need to belong to a sensory *system,* $\mathfrak{S}$, whose state types are systematically inter-defined according to a multi-dimensional discriminability space. And this whole system needs to be coupled to a certain kind of cognitive architecture with conceptual and conative states that extract information for further processing and behaviour. I will ignore these sorts of complication and assume that whatever else is needed is in place. I'll sometimes call these 'background' conditions for the emergence of phenomenal character.

[8] Tye gives the following formula: "a sensory state is about a property, *P*, just in case the state is of a type that is Normally tokened if and only if *P* is tokened and because *P* is tokened." (2014, fn.20)

selection process, happened to have phenomenal character, this character wasn't constituted by the *P*-instance that caused it in the circumstance. Recall, as per (c), phenomenal character is constituted only by a state-token's belonging to a state type whose function is to indicate P-instances. This type wasn't there, to begin with, in the selection stage.

According to representationalists, then, external objects and sensible property instances *never* metaphysically constitute the phenomenal character of sensory experiences.[9] Rather the phenomenal character is constituted by what sorts of state types get to be causally tokened. What is constitutive for the token experiences to have the phenomenal character they do is that they belong to a *state type* whose tokens are under the nomic control of property instances that they track under Normal conditions. In other words, the phenomenal character of token experiences is inherited from the *type* they belong to, not from the property instances these tokens purport to indicate. This state type is a functional type whose tokens purport to indicate and are the realizers of the experiences that represent sensible properties — thus constituting their phenomenal character.

Let me clarify a point about the *causal* determination of phenomenal character. There is of course a clear sense in which the phenomenal character of Sam's experience VE1 was determined by *what* she saw, namely an instance of blue. I argued that this determination was *causal* rather than constitutive. Causal determination of this sort is more like the causal selection of a sensory state from among a system ($\mathcal{S}$) of states already possessing different phenomenal characters — as per (c) above. For instance, if the ball Sam saw were red, instead of blue, VE1 would have a different phenomenal character. Not because the particular instance of red Sam saw would metaphysically constitute the phenomenal character of Sam's visual experience, but rather by causally activating a token of a different sensory state type in Sam's $\mathcal{S}$ that has the function of indicating instances of red.

This kind of causal determination is not relevant to phenomenal externalism that representationalists usually have in mind. What they need to defend is externalism of the constitutive kind. Above we've determined that the most plausible version of the claim that phenomenal character is constituted by "external" universals is given by (PC). And (c) is one way to cash out (PC) in completely naturalistic terms — in terms of Dretskean indication functions.[10] Does it deliver what is needed? I will argue in what follows that it doesn't.

---

[9] Compare the robust phenomenal externalism of disjunctivist naïve realists. The phenomenal character of sensory experiences, in the good cases, is metaphysically constituted by physical objects and the mind-independent properties instantiated by these objects (plus perspectives, etc., perhaps). See, among others, Campbell (2002), Martin (2004), Brewer (2011).

[10] I won't bother to try out other naturalist proposals like Tye's or Millikan's. Differences in these proposals won't make a difference in my argument in what follows. Also, although I'm very sympathetic to a Dretskean psychosemantics (see Aydede & Güzeldere 2005), I won't assume here that these sorts of proposals can naturalize perceptual intentionality.

But before I do that, I would like to pause and reflect on the following conditional claim for a moment:

> (C) For any sensory experience *s* and any sensible property *P*, and any time *t*, if the phenomenal character of *s* representing *P* at *t* is never metaphysically constituted by the instances of *P*, then it is not (partly or wholly) constituted at *t* by the uninstantiated universal *P* itself either.

I have argued for the antecedent of this claim so far. In section §5 below, I will argue directly for the consequent. But it is natural to consider why this conditional looks to be very plausible at this point. The phenomenal character of any sensory experience is an episodic and categorical property of the experience that belongs to the "here-and-now" during the experience's occurrence. As such, this character is itself an instance of a certain phenomenal type. What Sam's VE1 — as a *token* instantiating a phenomenal type during the first five seconds — indicates is an instance of *B*, namely *b1*. But it turns out, *b1* has constitutionally nothing to do with the phenomenal type of which VE1 is an instance. How is this *actual phenomenal instance* then supposed to be (partly or wholly) *constituted* by an uninstantiated (un-instanced) universal? Can an instance of a property be metaphysically and literally constituted (partly or wholly) by an uninstantiated property? The mind boggles.

Properties (universals) have their causal powers in virtue of their *instances* that enact them in actual causal processes that surround us. In other words, there is no such thing as the causal powers of an uninstatiated universal — except through its instances. The nearly unintelligible insistence in a universal being constitutive of phenomenal character of an actual sensory experience like VE1, therefore, also makes a mystery out of the causal powers of this experience.

To say that the phenomenal character of Sam's VE1 is constituted by a token sensory representation of *B* but not by the instance indicated, namely *b1*, is to say something that phenomenal internalists may easily accept. This is because, for them, either (i) sensory representation of *B* is an internal affair or (ii) phenomenal character is not constituted (merely) by representation. So, an externalist representationalist must demonstrate how sensory representation can *both* be an external/relational affair *and* constitute introspectable phenomenal character *while* at the same time making sensible property instances metaphysically irrelevant to phenomenology altogether. This seems like an extremely unlikely project. Strong representationalists promote their agenda by claiming that it has the most promise of naturalizing phenomenal consciousness: it is therefore a mystery-reduction enterprise. So, they need to deny (C) without multiplying mysteries, and it is not clear how to do this given the enormous initial plausibility of (C). If (C) is true, then actually establishing its consequence independently is not needed at this point, and I can stop the paper here. But it is instructive to see how it can be established independently.

4 The alleged phenomenal externalism

Representationalists typically argue for their case in the following way. VE1 is a brain state, a certain activation of a set of neurons in the relevant circuitries implementing the quality spaces in colour and shape detection. For ease of exposition, let's just concentrate on colour and ignore the shape. Let $\mathcal{S}$ be Sam's *colour* visual system whose different state types, S_i, implement the relevant neural activations in her visual pathways and cortex — these activations corresponding to registering different specific shades of colours. In particular, let S_{UB} be the state *type* belonging to $\mathcal{S}$ that has the systemic function of indicating instances of unique blue (UB, the universal). We can say, then, the token state, s_{UB}, is the realizer of Sam's *colour* experience *e* (say, during the first 5 seconds) indicating UB, the instance of UB had by Tom.[11] A representationalist can say that even if the phenomenal character of *e* is not instance-involving (hence not constituted by UB), it does involve UB — it systemically represents UB in virtue of having the function to indicate instances of UB. So, UB is what partially but essentially individuates S_{UB} of which s_{UB} is a token.

S_{UB}, the realizer of the experience type of which *e* is a token, is a state type whose historically relevant tokens got selected because they have indicated UB-instances. Although these indication relations have all consisted of particular causal interactions between the tokens of UB and S_{UB}, the result was that S_{UB} acquired the function of indicating UB-instances, thus the power of representing UB (as instantiated) — veridically or not.[12] The individuation of S_{UB} (in fact the whole $\mathcal{S}$) thus essentially adverts to the historical and causal interactions with UB through its instances. This is what systemic representation comes to. *e*'s veridically representing UB/UB is therefore an essentially relational property of *e*. Indeed, in the original example, *E*'s systemically representing *B* (thus its having the same phenomenal character during 25 seconds) is a relational property of *E*. A representationalist would then conclude: change the relation, thus the type-identity of the token state, as per (c) above, you change the representational content of *E*, and therefore its phenomenal character: thus, you change the experience type of which *E* is a token. You've got your phenomenal externalism of the constitutive kind.

Before arguing against this, let me say a few things about Sam's $\mathcal{S}$: well, it is *Sam's* colour perception system. So, *it* doesn't have any historically relevant tokens that contributed to its own selection and passing its blueprint to Sam's descendants. Rather, $\mathcal{S}$ belongs to a type of system $\mathfrak{S}$ that is phylogenetically fixed for the human species. The only known way

[11] This is partial realization given that VE1 involves representing other properties. But ignore this for the moment. We'll concentrate on the simpler colour case. I'm using SMALL CAPS to indicate that the reference is to the token.

[12] It's extremely unlikely that it is the individual state types, independently of others, that acquired the function of indicating *specific* sensible properties. Rather, it is the system type $\mathfrak{S}$ as a whole, whose particular states are interdependent, that acquired the function of indicating a *range* of sensible properties within a certain stimulus domain. I will use this point to argue against representationalism below.

of phylogenetic development of sensory systems is at the biological level — at the level of the mechanics of biological inheritance (involving DNA replication and expression). At this level, 𝕊 has a fairly robust neurophysiological description whose "system-level analysis," as engineers call it, can be given at the neurofunctional level. So, if representing a sensible property like UB is a relation, it is a relation with two relata: UB and S_{UB}-*qua-a-neurophysiological-state-type-belonging-to-*𝕊. Therefore, the state types of Sam's 𝓢 acquire their relational character by being of the same neurophysiological system type as 𝕊 — or whatever the descriptive level required by the transmission of a phylogenetic trait may be. 𝕊 had had *millions* of more tokens after it'd acquired its function — let's idealized away all the messy variations in this phylogenetic process (we don't have any good account of when the acquisition process is considered over or why it cannot change later). In almost all these cases, the internal constitution of people with 𝕊 who are in S_{UB} metaphysically suffices for them to have a sensory experience with the attendant relevant phenomenal character — whether or not they are veridical. When the representationalists say that the sensory representation is essentially a relation, they need to specify the relata. If one relatum is the universal UB (through its instances), the other relatum cannot be S_{UB}-*qua-representing-UB*. The relata need to be specified independently of each other if they need to serve as the basis for the naturalization of phenomenal intentionality in terms of indication functions.

This is not an argument against the relational individuation of phenomenal character yet, but it's a corrective for a proper understanding of how to individuate 𝕊 for its service in the naturalization project. 𝕊 has a robust neurophysiological characterization and this is important in understanding how it's supposed to work as a system.

5 "Shifted" phenomenal character

Scenario 1. Now consider Kim, who is a contemporary of Sam and roughly of the same age. Both are considered to have "normal" colour vision. But Kim's colour (hue) phenomenal space, although the same as Sam's, responds to a systematically shifted colour (hue) spectrum.[13] For instance, the tokens of Kim's S_{UB} are under the nomic control of instances of PB — a slightly but noticeably reddish (purplish) shade of blue. So, Kim's tokens of S_{UB} regularly indicate instances of PB — not UB. Not only that, let's assume, almost all her hue circle is shifted slightly compared to Sam's stimuli giving rise to same colour experiences. We don't need to assume that the degree of shift is even or thoroughly systematic. There are plenty of actual cases like this (see Kuehni 2004).

What is the phenomenal character of Kim's S_{UB} states? Are they of the same kind as those of Sam's? For a representationalist, the answer depends on whether they are both *representing* the same colour property or not. For instance, it may be that Kim is

[13] In what follows I'll have in mind the hue circle rather than the full colour quality space when I talk about colours.

systematically misrepresenting the colours that she sees — she may be systematically misrepresenting an instance of PB as UB (and similarly for almost the rest of the shifted spectrum). This could be for a variety of reasons. For instance, if, due to a genetic fault, the pigments in her cones have slightly different compositions so their response curves are slightly different (preserving the ratio of their firing rates), or maybe her eye lenses are filtering some lights due to degeneration or deformation from birth, or perhaps the neural processes between her eyes and the pre-V1 areas in her visual cortex got wired differently due to a genetic mutation, etc. If this were so, even though the states of her 𝕾 have the same indication function (thus the same representational contents) as that of Sam's, they regularly would fail to perform their function successfully resulting in systematic misrepresentation. In other words, we may think of Kim's 𝕾 as having the same neurophysiological and representational profile as that of Sam but, due to some mishap, let's say, in her early (pre-V1) neural wiring, as exhibiting a slightly different mapping from stimulus classes to the particular states of her 𝕾, therefore as not fulfilling its function in the way it was selected for. The malfunctioning of Kim's 𝕾 may be an anomaly. In such a scenario, Sam's and Kim's experiences realized by S_{UB} would have the same phenomenal character despite their systematically representing different shades of colour (Kim misperceiving the ball as PB and Sam seeing an instance of UB on the surface of the ball). Of course, it is very likely that, given how widespread the shifted spectrum cases actually are among the normally colour sighted people, both Sam and Kim may be systematically misrepresenting colours all the time. No serious problem so far for the representationalists.

Scenario 2. But let's modify the example slightly. Let's introduce Eve who, by all standard tests, has a normal colour vision. Just like Kim's, the particular states of her 𝕾, which is neurophysiologically type-identical to Sam's and Kim's 𝕾, respond to a slightly shifted hue spectrum. In fact, let's assume that Kim's and Eve's 𝕾 have the same neurofunctional (informational) profile: Kim and Eve reportedly agree systematically on the binary structure of the hues they see. A surface that reportedly looks UB to Kim also looks UB to Eve, and they reportedly agree more or less on the character of what they see for the rest of the hue circle. But there is a crucial difference between Kim and Eve: although Kim's colour experiences are systematically shifted and *mistaken*, Eve's experiences, although similarly shifted, are *accurate*. In other words, Eve's colour experiences represent the colours correctly. How is this possible? Well, following the naturalistic psychosemantics under discussion, we'll have to assume that it is the *function* of Eve's S_{UB} states to indicate instances of PB, and similarly it is the *function* of her visual system 𝕾 to respond to the light spectrum in this "shifted" way. So, we are assuming that Eve's visual system came to be where it is now due to an evolutionary process that selected for it. Eve is a member of a group of human species who has evolved slightly differently.[14]

[14] Perhaps Eve's ancestry traces to the oldest South American indigenous people who migrated to the continent tens of thousands of years ago where vegetation colours were remarkably different. But it's not crucial to the argument that Eve can be actual or belongs to the empirically possible nearby worlds — see below.

Now there is a question about whether to count Eve's visual system as of the 𝕊 kind. The issue here concerns how narrowly or broadly we should individuate 𝕊. Visual colour processing starts with photons hitting the cones and its later stages involve whatever neural circuitry (perhaps including opponent processes running through LGN and various parts of the visual cortex) implements the final discrimination behaviour that underlies the colour quality space — sometimes known as the three-dimensional colour solid. The processing in the cones as well as the retinal and early post-retinal processing may be manipulated without massive differences resulting in the implementation mechanisms of the colour quality space. There is no reason to think that among the normally sighted but shifted colour spectrum cases people have different colour (hue) quality spaces.[15] I will just stipulate that 𝕊 be individuated without including these very early processes. It is an empirically plausible assumption that most people with "shifted colour qualia" share the same colour quality (in particular, hue) space implemented in more or less neurophysiologically type-identical neural structures. If we individuate 𝕊 this way, then our assumption about Eve amounts to the assumption that she belongs to a group of community whose 𝕊-state types have evolved to acquire a different indication function due to some differences in their environment and/or in their early (pre-V1) neural processing.[16] So, for instance, while Eve's 𝕊(S_{UB}) has the function to indicate instances of PB, Sam's 𝕊(S_{UB}) has the function to indicate instances of UB. The result is that the two tokens of the same neurophysiologically identified state type 𝕊(S_{UB}) sensorially represent different colours for Eve and Sam.

Now we have reached the kind of phenomenal externalism that representationalists have in mind — the constitutive kind. In this last scenario, Sam and Eve share a neurophysiologically identified 𝕊 whose type-identical states, S_i, represent different colours. When Sam looks at the ball, being in 𝕊(S_{UB}) she is veridically seeing an instance

[15] The evidence for the claim that among the normally sighted the hue quality space is roughly the same interpersonally and implemented by the same or similar (more centrally located) neural structures has been accumulating in the last two decades — although we are very far from having the fine details of the implementing neural assemblies. It seems that specific areas in the ventral V4 and VO1 are crucially involved and the contribution of earlier V1 and V2 areas to these is now being more readily acknowledged and mapped out. See Conway (2014), Bohon *et al.* (2016), Kim *et al.* (2020), Siuda-Krzywicka *et al.* (2020), Chatterjee *et al.* (2021), Li *et al.* (2022). Although we don't have the details that would empirically settle this, one point to note is that this claim is the underlying assumption in these brain studies — not to mention the fact that many earlier studies were using macaque monkeys that have colour vision brain areas homomorphic to humans'. Note also that the findings of these studies are mostly orthogonal to whether the opponent process models of the binary/unique structure of hue circle are empirically correct. There seems to be some evidence that the cone opponency that dominates earlier layers is dropped starting V2 — see Brouwer & Heegar (2009, 2013) and Du *et al*. (2022).

[16] In order for this to work, we'll probably need to assume that the communities Sam/Kim and Eve belong to have been relatively isolated from each other throughout the relevant part of the evolutionary process — or at least they haven't mixed their lineages much. In fact, for the thought experiment to work, all that is needed is the *metaphysical* possibility that the states of 𝕊 may have acquired distinct indication functions (indicating slightly different spectra) in two different phylogenetic lineages.

of UB. When Eve looks at *another* ball that is slightly reddish (purplish) blue, being also in $\mathfrak{S}(S_{UB})$, she is veridically seeing an instance of that colour (PB). Representationalism delivers the result that the phenomenal character of Sam's and Eve's experiences is of different kinds because they represent different colour properties (hues *qua* universals). Thus, despite their neurophysiological type-identity, the phenomenal character of Sam's and Eve's experiences is constituted by different colour universals — different sensible properties that are instantiated only by "external" mind-independent entities. We can even think of Sam and Eve sharing all their narrow ($\mathfrak{S}$ and post-$\mathfrak{S}$) internal constitution relevant for conscious colour processing. Representationalists claim that Sam's and Eve's experiences despite being realized by the same internal physical state — $\mathfrak{S}(S_{UB})$ — have different colour phenomenology because they represent different colours. This claim, we have already seen, doesn't entail that the *instances* of these colours, PB and UB, are metaphysically relevant to the constitution of the phenomenal character of their respective experiences. The role the colour instances play is merely the causal generation of their respective $\mathfrak{S}(S_{UB})$-tokens that nevertheless differ in their phenomenal character. Similarly, when Sam and Eve look at the same UB ball they have experiences with the same phenomenal character despite the fact that they occupy different states of $\mathfrak{S}$. Sam occupies S_{UB}, and Eve occupies, let's say, S_{PB}, when they both look at the ball in good light, and they correctly represent the colour of the same ball. If the description of this scenario so far is coherent, representationalism entails that Sam and Eve are having type-identical colour phenomenology when they look at the ball, despite their occupying neurophysiologically different states of their respective $\mathfrak{S}$'s that represent the ball's colour correctly. Note that the description of this scenario so far contains nothing that a representationalist should in principle object about. In fact, it is a detailed description of one plausible way to fill in what phenomenal externalist representationalism quite straightforwardly entails.

But, how, do you think, would Sam and Eve describe their own colour experiences — their hue phenomenology, when they are looking at the same ball? Above, I said that Kim and Eve reportedly agree with the character of their hue experience when *they* report the ball looking purplish blue to them. If we believe they are correctly reporting their phenomenology, we seem to have an internal inconsistency in the scenarios. Before addressing this question, for convenience and ease of following the argument, let's combine the scenarios by bringing all three of them (Sam, Kim, Eve) together and changing the example so that all three of them simultaneously look at the same Munsell color chip called '5B' under the same material conditions.[17]

Scenario 3. Sam, Kim, and Eve are looking at 5B (the right-hand side in Figure 2). This chip looks exactly like the colour of the ball to Sam — let's stipulate that the chip's and the ball's surfaces have the same exact surface spectral reflectance. We've stipulated that Kim's looking at this chip would cause Kim's $\mathfrak{S}$ to enter a state different than the state Sam's $\mathfrak{S}$

[17] See Figure 3 below for the naming of Munsell chips. Many normally sighted people (like Sam) describe this chip as looking unique blue.

enters. Indeed, we've stipulated that Kim reports the chip as looking PB to her whereas Sam reports it looking UB to her. Representationalists would have us believe that Sam's and Eve's hue experiences share the same phenomenal character when they look at 5B if, as described, they represent the colour of B5 correctly. So, we have the following situation as depicted in Figure 2 (right hand occasion).[18]

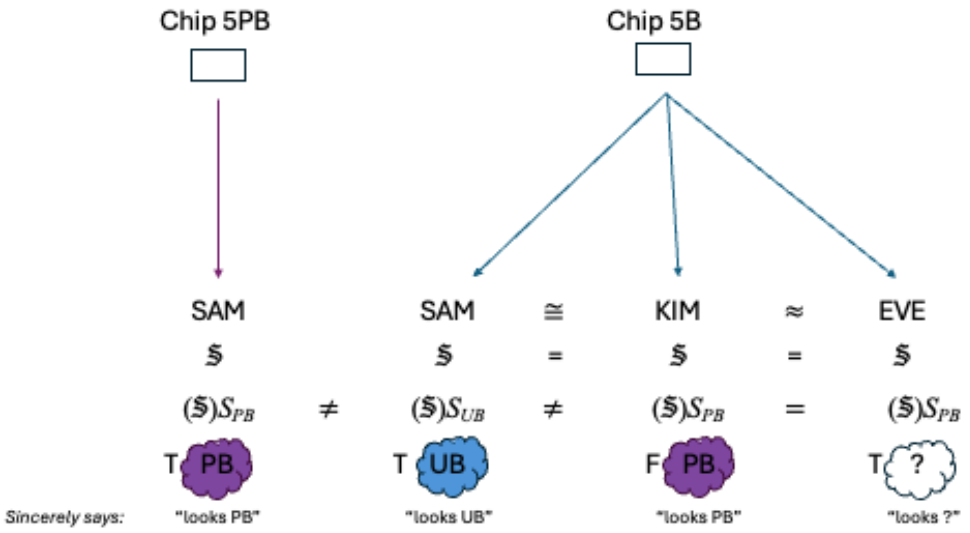

Figure 2. Viewing of the left and right chips by Sam are two separate occasions. '=' expresses near neurophysiological type-identity, '≠' type-distinctness. '≈' expresses "nearly narrow neurofunctionally type-identical" and '≅' expresses "nearly neurofunctionally type-identical post-§." 'T' and 'F' express veridicality and non-veridicality respectively. Clouds are meant to indicate the phenomenological character of hue experiences.

Following the scenarios, we are assuming that, unlike Kim's, Sam's and Eve's hue experiences represent the colour of 5B correctly. On the other hand, unlike Sam, Kim and Eve are narrow functionally type-identical and they are in the same §-state. They all speak their native tongue, English, competently. We also stipulate that they are as reliable as anyone can be in correctly reporting how the colours of surfaces *look* to them. We now ask all three how the colour of 5B looks to them.

Sam: "it looks unique blue"
Kim: "it looks slightly purplish blue"
Eve: "it looks…?"

[18] For dramatic effects, we may assume that Kim and Eve more or less share their internal (narrow) neurofunctional constitution (indicated with '≈' below) — in fact, make their entire physical *individual* histories as similar as they can empirically can be. All three share the same narrow neurofunctional constitution and organization post-§ (indicated with '≅' in Figure 2).

There is every reason to believe that Eve's report will be "it looks slightly purplish blue" just like Kim's, correctly reporting the phenomenal character of her colour experience. So she report her colour experience to have a different phenomenal character than Sam's. If this is correct, then contrary to the prediction of representationalism, same representational content doesn't guarantee same phenomenal character. But then phenomenal externalism of the kind representationalists have had in mind is refuted. The kind of relational individuation of 𝕊's states does not assign the correct phenomenal character to them. It looks like the neurofunctional type-identity of $\mathbb{S}(S_i)$ will trump the representational individuation any time the two schemes come apart.[19]

Is it plausible to insist that Eve, despite her narrow neurofunctional type-identity to Kim, would correctly report her own experience as "[5B] looking unique blue"? This would generate a profound mystery about how to explain Eve's behaviour, including her linguistic behaviour — especially given the fact that on those occasions when Sam's 𝕊 enters S_{PB}, Sam reports the surfaces she sees as "looking purplish blue", and similarly, on those occasions when Eve's 𝕊 enters S_{UB}, Eve utters the words "… looks unique blue". The evidence for Eve's undergoing a phenomenology different than Sam's is overwhelming — almost decisive under any plausible metrics.

The structure of the above argument is simple. Two actual "normal" perceivers, Sam and Eve, one having slightly shifted colour spectrum relative to the other, are looking at the same chip that instantiates a colour property, *C*. Their colour experiences represent the colour of the chip *accurately*. If the phenomenal character of their respective colour experiences were constituted by the colour properties (qua universals) they represent, then the phenomenal character of their experiences would be constituted by *C*. So representationalism implies that their experiences have identical phenomenal character. But this implication is falsified by the fact that the phenomenal character of their experiences differs as revealed by Eve's report (and her other discrimination behaviour). We arrive at this conclusion by making two very plausible assumptions. First, in many actual shifted spectrum cases, like in Sam and Kim, the hue quality space is roughly the same and implemented by the same or similar (more centrally located) neural structures

[19] There is a version of a very similar argument given by Adam Pautz (2006) that conceives of Eve (his twin-Maxwell), in a certain sense, as identical to Kim in a counterfactual world — where we assume the physical type-identity of Kim and twin-Kim including their *individual* histories. Depending on how we conceive of their *evolutionary* histories (selection history of their ancestors), therefore, we get *different* phenomenal characters under this scenario despite their type-identical internal physical constitutions. Pautz doesn't cast his argument in terms of *actual* shifted spectrum cases, and heavily relies on the hypothesized Opponent-Process mechanism of colour vision against which there is an increasing experimental literature (see Arstila 2017 for discussion and references). I don't rely on this hypothesis, although I assume that the resulting colour hue circle has a unary-binary structure that can be phenomenologically appreciated. My assumptions for the actual "shifted-spectrum" cases seem less expensive: all I need are two hard-to-deny assumptions — see below. Nevertheless, the arguments are, no doubt, very similar. Thanks to Jonathan Cohen for alerting me to this similarity.

(whatever they are).[20] Second, it is *metaphysically* possible for people with shifted spectrum to have slightly different evolutionary histories so that whatever the naturalistic psychosemantics is needed for errorless representation, it is in place in Sam and Eve.

As far as I can tell, there are two broad venues for reductive strong representationalists to resist this conclusion.

5.1 Not more than one evolutionary lineage?

One venue is to insist on the impossibility of the second scenario involving Eve, saying that there can at most be one evolutionary development of human colour system that sets the correctness conditions of colour experiences. There are plenty of actual individuals living among us with shifted colour spectrum.[21] The representationalist ought to claim that none of their kinds (except one, perhaps) could have acquired their shifted colour vision through a relatively independent evolutionary process that selected for it. But the modal strength of this claim seems empirical, not metaphysical. Representationalists need to establish this claim as a metaphysical necessity. I don't see how this can be done. In fact, for all we know, it wouldn't be too surprising if it turns out that this claim is in fact *empirically* false. The claim that some people could have acquired their shifted colour spectrum through a relatively independent evolutionary process that selected for it is clearly *nomologically possible*, although for contingent factors it may be empirically extremely unlikely. We just don't know. I will return to this point later.

5.2 Multiple simultaneous colours?[22]

The second venue of resistance may come from a suggestion made by Byrne and Hilbert (1997). They claim that the chip is simultaneously and objectively both unique blue and purplish blue. According to Byrne and Hilbert, representable maximally specific shades of colours are sets of surface spectral reflectances (SSRs) and these sets intersect. So, a *particular* SSR can perhaps be a member of two or more different such sets. In our case, it can be that the chip that both Sam and Kim are looking at (under the same viewing conditions) happens to have a particular SSR that belongs to both sets to be identified with unique blue and purplish blue (specific hue shades of certain brightness and saturation — call these sets SSR_{UB} and SSR_{PB} respectively). What determines the set membership for each shade? The intended answer is: normal perceivers in a species where we know there are variations (such as visual shifted colour spectra). The rough and a bit idealized procedure would look something like this: let each normal perceiver pick the chip they

[20] See fn. 15 above for references. Note that this assumption is not a universal claim and can tolerate there being *some* normally sighted but "qualia shifted" individuals whose colour quality spaces are slightly different or implemented by slightly different brain mechanisms.

[21] Hardin (1993: 79–80), Kuehni (2004).

[22] If you are not at all moved by so-called "selectionist" proposals that multiply simultaneous surface colours of objects, this subsection may be skipped without any disruption to the flow of the content in what follows.

consider to be unique blue under the same viewing conditions (let's say, out of 40 visually equidistant Munsell hue chips with same brightness and saturation level — see Figure below). This would amount to about seven consecutive chips (7.5PB to 5B inclusive). So, take the set for SSR_{UB} (unique blue) to consist of all and only those particular SSRs that are the metamers of the SSRs of these seven chips. Next, let each perceiver pick the chip they consider slightly purplish blue (by asking them, let's say, to pick what they consider to contain 25% red). It is a reasonable guess that this would give us around seven chips (let's stipulate for the sake of the example that the range is from 10B to 5P). Then the set for SSR_{PB} would consist of all and only those particular SSRs that are metameric matches of these chips. We can follow this procedure for more or less each discriminable shade around the hue circle. This would create sets that significantly intersect with each other. Most any surfaces would then literally have more than one colour simultaneously — in most cases, quite a few, depending on the empirical details and the initial stipulation about who counts as normal perceivers.

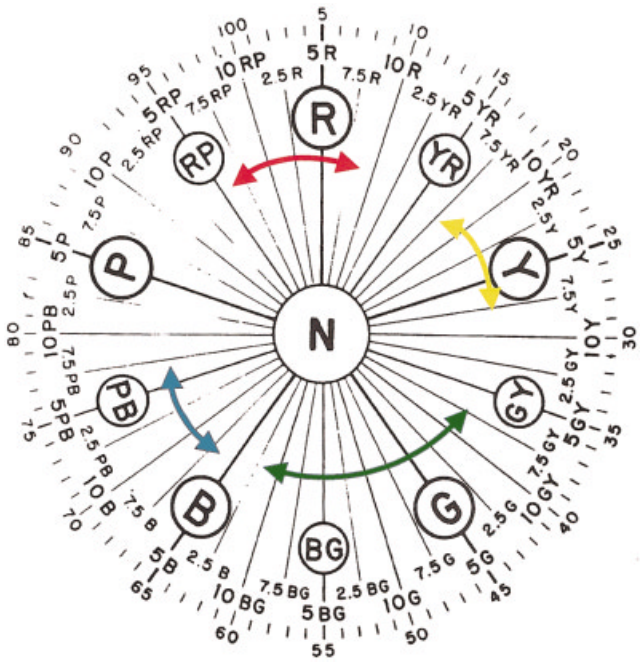

Figure 3. Munsell system perceptual chromatic diagram with inner circle segments (coloured double-arrows) indicating the approximate unique hue ranges among normal perceivers based on viewing colour chip ranges. (Kuehni 2004)

Intuitively, the idea is to let the specific phenomenal character of a shade of colour for each normal perceiver *select* the range of SSRs that will then constitute the set to be identified with that colour shade, so that the phenomenology doesn't come apart from the represented colour (= SSR set). When the selection is found out empirically on the basis of "normal perceivers" of a species (say, by letting them pick Munsell chips), the SSR sets to be identified with specific colours heavily overlap. How much overlap there is is contingent

on how much variation exists among normal perceivers. This process can be narrowed down by putting restrictions on the range of "normal perceivers" to allow for misrepresentations. But it is highly doubtful that there is any principled way of doing that.

Representationalists who take this road are now committed to a claim that they would otherwise be very happy not to take on board: objects turn out to have more than one colour simultaneously and most of these colours are not visible to many normal perceivers. One might be tempted to tolerate the proliferation of a few simultaneous colours when they are *very* close to each other. But note that 5B is both unique blue and almost unique green, this is because some normal perceivers pick it as almost unique green and some pick it as unique blue — so 5B's SSR belongs to multiple sets. There is a big phenomenal difference between seeing unique blue and unique green.[23] If I turn out to see 5B accurately as unique green, I'm invited to agree by representationalists that 5B *is* also unique blue simultaneously — it's just that I can't see the other colour (in fact many other colours it also has). This seems to me to be an intolerable consequence of this sort of representationalist response to the criticism that colour universals can't enter into the constitution of colour phenomenology. This "selectionist" response is radically at odds with some of our most fundamental beliefs about colours: that, for instance, an object's surface cannot be both unique blue and unique green at the same time under the same conditions. Saying that only one colour can be seen by a single perceiver doesn't remove the paradoxical nature of the claim. If phenomenal character is constituted by the objective colours we see, then our grasp of the nature of colours must be fairly direct. On the basis of the grasp we have, here is another fact we seem to know: that there cannot be colours that I can sensorially represent but can't see them on many occasions even under optimal conditions. Denying these intuitively evident truths is a hard bullet to bite.

But more importantly, this particular selectionist proposal is completely *ad hoc* and not workable. To see this, let us ask why I can't see the other colours that 5B has. According to representationalists, what colours I can see is a matter of what colours I can sensorially represent. Being a normal perceiver, I can actually represent all the discriminable colours — so I can experience them all, and so can you. When I look at 5B, the phenomenal character of my seeing unique green is constituted by my experience's representing this colour property. What is this colour property according to the representationalist? Under the current proposal, it is the set of all and only those SSRs that would strike (most) normal observers (including me) as unique green (UG). Because there are huge variations among normal perceivers, this set will include a huge number of SSRs that will not strike *me* as unique green — some of these may strike me as almost unique blue, let's say. Similarly, when you look at 5B, you'll accurately see it as, say, unique blue. This is because 5B is

[23] In fact, Kuehni (2004) reports that when spectral lights (instead of Munsell chips) are used, the range of unique blue covers some lights that are seen by others as unique green. In what follows, I'll just set up the example for convenience as a case where 5B can be seen as UB as well as UG by different normal perceivers.

also unique blue, that is, its particular SSR is a member of the set of SSRs identified as unique blue (UB) by (most) normal perceivers.

Consider now the chip titled 5PB. This chip looks unique blue to me, because let's say, it indeed is (following the suggestion made by Byrne and Hilbert). The particular SSR of this chip (SSR_{5PB})[24] belongs to the set identified with *this* shade (UB). But this very chip SSR_{5PB} will look to you accurately, let's say, as purplish blue, because it also belongs to the SSR set identified with *this* slightly reddish (purplish) blue (PB).

The phenomenal character of my visual experience is constituted by UG when I look at 5B, because the particular neural state type S_i that my visual system $\boldsymbol{S}_{me}$ is in, as a result of causal interaction with SSR_{5B}, represents UG. The phenomenal character of your visual experience, on the other hand, is constituted by UB when you look at the very same chip 5B, because the neural state type *SJ* that your visual system $\boldsymbol{S}_{you}$ is in, as a result of causal interaction with SSR_{5B}, represents UB. 5B is indeed both UG and UB (and many other nearby colours). If you and I belong to the same species with the same evolutionary history, the system of representation ($\boldsymbol{S}_{you}$) realizing the hue quality space in you and the one realizing mine ($\boldsymbol{S}_{me}$) are tokens of the same system type $\mathfrak{S}$ that has been selected for its function. $\mathfrak{S}$ is what characterizes the relata of the representational relations (between the particular state S_j you are in and UB on the one hand, and between the particular state S_i I am in and UG on the other). As before, the consequence of assuming a phylogenetically sustained evolutionary selection history is that $\boldsymbol{S}_{you}$ and $\boldsymbol{S}_{me}$ are neurofunctionally type identical (more or less) and so are their particular states. So, we may as well talk of particular neural state type $\mathfrak{S}(S_{UB})$ representing UB and neural state type $\mathfrak{S}(S_{UG})$ representing UG among normal perceivers.

If SSR_{5B} is an instance of both UG and UB, we may ask: in virtue of what does your visual system select UB whereas mine select UG when we are causally interacting with the same SSR_{5B}? In other words, what makes it the case that SSR_{5B} causes *you* to enter the neural state $\mathfrak{S}(S_{UB})$ that has the function to indicate instances of UB, but causes *me* to enter a distinct state $\mathfrak{S}(S_{UG})$ with the function to indicate instances of UG? More perspicuously, what makes it the case that my situation is correctly described as accurately seeing the colour of 5B (its being UG) rather than as failing to see other colours it has, say, its being UB — in other words, as a case where my particular $\mathfrak{S}(S_{UB})$ fails to detect the colour of this chip (i.e., its being UB)? Is there a principled answer to this question that is also consistent with representationalists' preferred psychosemantics? I fail to think of any.

Note that when we both look at 5PB, I enter the state $\mathfrak{S}(S_{UB})$ that correctly represents UB, and you enter $\mathfrak{S}(S_{PB})$ that correctly represents purplish blue (PB). We both can represent UB, and when we do, we represent them with (more or less) the same neural state $\mathfrak{S}(S_{UB})$. But you and I cannot see the same colour when we both look at the same chip (token

[24] And its metamers — I'll omit this qualification in what follows. We can take SSR_{5BG} to name the metameric equivalence class. Similarly with the particular SSRs of other chips.

object). The other colours of this chip magically become invisible to us — we fail to detect them. (Of course, I know the phenomenal character of your colour experience when you look at 5B, which is different than mine when I look at 5B — assuming I've been in state $\mathfrak{S}(S_{UB})$ before.) So, our respective systems (tokens of $\mathfrak{S}$) fail to deliver full information by detecting at most one colour among the many colours that an arbitrary object usually has. On anybody's story, this would be a failure rather than a success story about the function of our colour systems. How could such a system manage to have evolved in our species?

The move to proliferate colours that an object can have simultaneously was an attempt to make sure that phenomenal character and representational content (colours) don't come apart where it is implicitly acknowledged that phenomenal character goes along with neurofunctionally type-identified states that belong to $\mathfrak{S}$. The question then becomes what makes the case that each particular state $\mathfrak{S}(S_i)$ represents what it does. The answer is that each has the function to detect a set of SSRs. But these sets heavily intersect because not all the members of a set cause a normal perceiver to enter the same state of $\mathfrak{S}$ (≈ see the same colour). Thus, this suggestion makes magic out of how each of us manages to see the colour we end up seeing rather than failing to see other colours. The relations like 'sensorially *representing UB*' become very difficult to understand on any naturalistic psychosemantics.[25]

I conclude that a Dretskean psychosemantics doesn't deliver the kind of phenomenal externalism for which the view has been advertised. As I said, I will generalize this conclusion, without argument, to all extant naturalistic proposals (versions of informational and/or teleological psychosemantics) about what it is a for a sensory state to *represent* a sensible property. This is because, it seems to me, whatever naturalistic conditions are required for sensory states to represent colour properties, they can be met in such a way that not only metaphysically but also nomologically allow for errorless representations of a single shade of colour with demonstrably different phenomenal characters, or for there being the same phenomenal character correctly representing different "shifted" colours.

6 How does being in $\mathfrak{S}(S_{UB})$ make a perceiver commune with the universal UB?

So far, I have argued against the antecedent and the consequent of (C) separately. To repeat:

> (C) For any sensory experience *s* and any sensible property *P*, and any time *t*, if the phenomenal character of *s* representing *P* at *t* is never metaphysically constituted by the instances of *P*, then it is not (partly or wholly) constituted at *t* by the uninstantiated universal *P* itself either.

[25] Also note that if a representationalist opts for this response by proliferating simultaneous object colours, the response in **Scenario 1** becomes unavailable: it will be very difficult to argue that among normal perceivers there are systematic *misperceivers* of colour.

Although I've briefly touched upon some of the reasons for thinking whether (C) itself is plausible, I've not elaborated a more direct argument for it. It's time to reflect on (C) more deeply given that its antecedent seems fairly well established.

Representationalists think that sensory phenomenology is constituted by what sensory properties are represented in a sensory experience so long as they are represented appropriately with the background conditions in place. It turns out that no *instances* of these properties play any essential role in the constitution of phenomenal character according to representationalism. It turns out that a sensible property like UB constitutes the phenomenal character of consciously sensing UB only in the sense that sensing UB reduces to sensorially representing the *universal* UB (as instantiated) which in turn requires some causal relations between the instances of UB and the historical tokens of a sensory state type, $\mathfrak{S}(S_{UB})$, in the selection of this type in the evolutionary history of our species. The state type S_{UB} when tokened in Sam appropriately counts as *constituting* the phenomenal character of Sam's colour experience in virtue of the fact that it makes Sam somehow "commune" with the *universal* UB that S_{UB}-token represents (as instantiated) in Sam's environment. The naturalistic psychosemantics that makes *this* allegedly possible (makes us allegedly understand how *this* feat is possible) has in its disposal only the *causal* connections that relate instances of UB and instances of S_{UB} in the selection history of the S_{UB}-type. In other words, whatever *relation* a representationalist envisions in explaining how the nature of phenomenal character is *constituted* by the universal UB, it is no stronger than a *causal* one. This is because the nature of sensory representation proposed as constituting the phenomenal character requires a relation no stronger than a *causal* relation. But causal relations are too feeble to make us understand how Sam can commune with the universal UB — how the phenomenology of Sam's experience is constituted by UB — when she is seeing an instance of UB. According to available naturalistic psychosemantic models, remember, sensorially representing UB requires relations no stronger than causal ones. But we have already established that causal determination of phenomenal character (even during the selection history) is not to be confused with constitutive determination. Put succinctly: If sensorial representation of UB reduces to some sort of indication function understood naturalistically, then sensorial representation of UB cannot explain how UB (*qua* universal) can *constitute* Sam's phenomenology when Sam is in S_{UB} — how Sam communes with UB. Phenomenal externalist representationalism just multiplies the mysteries. Why not drop externalism and say that being in S_{UB} (as opposed to, say, being in S_{PB}) *qua belonging to a neurofunctional type* constitutively determines the *particular* phenomenal character of experiencing UB? Saying this would *not* require denying that sensory representation is necessary for phenomenal character *in general* (see below). So the proper conclusion to draw is that once property instances are made irrelevant to the constitution of sensory phenomenal character, there is no naturalistic framework that can make phenomenal externalism work: a naturalist representationalism inevitably needs to learn how to live with some form of phenomenal internalism. In other words, denying (C) is a hopeless move once its antecedent is granted.

But, in fact, the situation is worse for an externalist representationalist. I've briefly mentioned above that the evolutionary selection of the hue system $\mathfrak{S}$ has almost certainly happened at the system level. In other words, it is empirically highly unlikely that the selection worked at the level of particular states types (S_i's) that $\mathfrak{S}$ can be in. In fact, this is acknowledged even by representationalists:

> Colour vision is very likely to a large extent a package deal: the representation of colour, fine- or coarse-grained, is systematic and plausibly selection did not even have the option of favoring those with the ability to identify true blue and not those with the ability to identify [purplish]-blue. Having a colour vision system, with the consequent ability to identify a variety of fine- and coarse-grained colours, confers a selective advantage. Given that the colour vision system comes as a more-or-less complete package, natural selection might have produced the ability to represent shades like true blue, even if that colour had never played any significant role in the ancestral environment. (Byrne & Hilbert, 2006, longer version)

So, this opens up the serious empirical possibility that among our ancestors (perhaps even currently) there were people who could consciously see surfaces in ways they would describe them as "looking purplish-blue" even though the particular state type (S_{PB}) they occupied had not been selected through any causal interaction with any instances of any colour property "represented" by S_{PB}. So, I invite you to imagine someone in good standing as a member of our species encountering a surface that looks purplish-blue to her for the first time in our history. According to externalist representationalism, the state she would be in, (S_{PB}), would make her commune with a universal PB (without, of course, the PB-instance she would be seeing for the first time constituting the phenomenal character of her experience). That is to say, the phenomenal character of her visual experience would be metaphysically constituted by the (mind-independent) universal PB. Phenomenal externalists would no doubt insist that there is nothing mysterious about any of this. I envy their assurance and self-confidence. On my part, the right conclusion to draw is that the *particular* state-type in question (*qua* neurofunctional type belonging to $\mathfrak{S}$) contributes to the *particular* phenomenal character in a way that goes beyond the fact that this type sensorially represents PB, and say, not UB. Representationalists, in other words, don't have to be phenomenal externalists, which is all for the best, because phenomenal externalism is not born out by the facts and regularly yields incredibly counter-intuitive results.

7 Conclusion and an alternative internalist picture

So, we still don't have phenomenal externalism. This conclusion shouldn't be all that surprising. It is difficult to fathom a philosophical account of sensory perception that accepts the Common Kind Thesis and offers a truly phenomenal externalist position. Non-reductive representationalism has been uniformly phenomenal internalist. It would have been somewhat perplexing if reductive representationalism of the Dretskean sort had turned out to be phenomenal externalist. If you have sympathies for phenomenal externalism you

should look at the naïve realist or disjunctivist camp — although I would not hold my breath for their ability to successfully deal with shifted spectrum cases either. For my money, the overall conclusion to draw is that phenomenal externalism is just false. If reductive strong representationalism entails phenomenal externalism, then, it too is false. In fact, once it is realized that, for representationalists, instances of sensible properties we are sensorially aware of play no constitutive role (as opposed to a causal role) in determining the phenomenal character of our sensory awareness, the job of finding a constitutive role for a sensible property (*qua* universal, in terms of sensorially representing it) becomes somewhat obscure, and as they say, "academic." But a naturalistic story about how this intentional property (sensorially 'representing *P*') is acquired doesn't deliver a constitutive role for the universal either: Sam and Eve are related to the same shade (universal) when they look at the same chip and accurately represent its colour but their experiences have different phenomenal character — ignoring the second, "selectionist," response. Once the role of property instances is reduced to causal but not constitutive determination of phenomenal character, all the intuitions start crying out for an internal contribution to the metaphysical determination of colour phenomenology. A naturalistic psychosemantics, as we have seen, doesn't change this at all. Phenomenology follows internal structure rather than external representation. (This is in fact the source of the temptation to proliferate simultaneous colours of objects.)

Note that the argument so far hasn't been against some form of intentionalism *per se* about sensory experience, or even against some naturalistic psychosemantics for such intentionalism. Rather it has been against the claim that (broad) representational content constitutes phenomenal character; more accurately, against the claim that the phenomenal character of a sensory experience *s* as of *P* is constituted by *s*'s 'representing *P*'. It is left open that *s* sensorially represents *P* while its *particular* phenomenology is constituted by internal structures functioning in the service of delivering information about *P*.

As an alternative, I offer the following picture, which is naturalist, intentionalist, but phenomenal internalist. Let's treat $\mathfrak{S}$ as before having internally interdependent state types, S_i, purporting to indicate instances of most determinate colour shades along the axes of the colour quality space. $\mathfrak{S}$ is a genetically transmitted and neurofunctionally specifiable system with an informational function (with an overall functional/computational profile). I will just say that the particular states of $\mathfrak{S}$ — $\mathfrak{S}(S_i)$ — all *purport to indicate* instances of colours. This *general* fact (if it's a fact) may be necessary for *any* particular state to have *some* phenomenal character (with the background conditions in place). But what *particular* character they each will have may be (at least partly) internally determined at the level of engineering. And what instances of particular colour shades each will purport to indicate may vary in different people having tokens of the same $\mathfrak{S}$ — whatever empirical accommodations are required to explain the widespread phenomenon of "shifted colour qualia." We can think of each of the S_i as a sensory predicate belonging to a *system* of analog representations ($\mathfrak{S}$) — each attributing a specific shade of colour to what is seen. These states would be the colour predicates of a colour representational system. In other

words, the particular states of $\mathfrak{S}$ may be taken as parts of syntactically structured representational vehicles whose semantic values are assigned according to local laws and whatever naturalistic psychosemantics is in place.[26] They would still have the job of indicating/representing colours, yes, but without this fact metaphysically determining the *particular* phenomenal characters each may have. But one can still maintain that a sensory system's having an indication function for a range of magnitudes for sensible properties is a necessary condition for sensory phenomenology to arise in general. In the older jargon, in other words, one may allow for the possibility of inverted or shifted qualia without *ipso facto* allowing for the possibility of absent qualia. Such a view needs to elaborate what it is about informational functions and the way they are imbedded in a larger, richer, and more complex information processing architecture that allow them to reductively explain phenomenal character.[27,28]

References

Arstila, Valtteri. (2017). "What Makes Unique Hues Unique?" *Synthese* 195 (5): 1849–72.

Aydede, Murat, and Güven Güzeldere (2005). "Cognitive Architecture, Concepts, and Introspection: An Information-Theoretic Solution to the Problem of Phenomenal Consciousness." *Noûs* 39 (2): 197–255.

Aydede, Murat (2019). "Is the Experience of Pain Transparent? Introspecting Phenomenal Qualities." *Synthese* 196 (2): 677–708.

Aydede, Murat (2020). "What Is a Pain in a Body Part?" *Canadian Journal of Philosophy* 50 (2): 143–58.

Bohon, K. S., Hermann, K. L., Hansen, T. & Conway, B. R. (2016). "Representation of Perceptual Colour Space in Macaque Posterior Inferior Temporal Cortex (the V4 Complex)." *eNeuro* **3**, ENEURO.0039-16.

Brewer, B. (2011). *Perception and Its Objects*. Oxford University Press.

[26] See Aydede (2019, §6) for more of this sort of line. Note that this framework could easily accommodate some kind of realist/physicalist ontology for colours in the form of SSRs or productances, or whatever. It's just that phenomenal character of colour experiences would be determined internally in the form of *how* physical colours *look* to observers.

[27] See Aydede & Güzeldere (2005) for a comprehensive attempt. Representationalists also typically accuse phenomenal internalists of violating perceptual transparency and of having implausible views about introspection. But these are relatively separate worries and have, in my view, satisfactory resolutions: see Aydede (2019, 2020) for internalist explanations of transparency and introspection.

[28] This paper grew out of a small discussion group discussing Pitt (2017) and Gottlieb & Rezaei (2021). I'd like to thank the authors for providing such a stimulating material as well as for their valuable comments on an earlier version. I am also grateful to Dom Alford-Duguid, Jonathan Cohen, Matt Fulkerson, and Laura Gow for their comments and questions.

Brouwer, G. J. & Heeger, D. J. (2009). "Decoding and Reconstructing Colour from Responses in Human Visual Cortex." *The Journal of Neuroscience* **29**, 13992–14003.

Brouwer, G. J. & Heeger, D. J. (2013). "Categorical Clustering of the Neural Representation of Colour." *J. Neurosci.* **33**, 15454–15465.

Byrne, A., and Hilbert, D. R. (1997). "Colours and Reflectances." In A. Byrne & D. R. Hilbert (Eds.), *Readings on colour*, (Vol. 1, Chap. 14, pp. 263–288). Cambridge, MA: MIT Press.

Byrne, A. & Hilbert, D. (2006). "Truest blue." (A longer version of an article published in *Analysis* **67**, 87–92, 2007).

Byrne, A. & Tye, M. (2006). "Qualia ain't in the head." *Noûs* **40**, 241–255.

Campbell, J. (2002). *Reference and Consciousness*. Oxford University Press.

Chatterjee, S., Ohki, K. & Reid, R. C. (2021). "Chromatic micromaps in primary visual cortex." *Nat. Commun.* 12, 2315.

Conway, B. R. (2014). "Colour signals through dorsal and ventral visual pathways." *Vis. Neurosci.* 31, 197–209.

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

Dretske, F. (1999). "The Mind's Awareness of Itself." *Philosophical Studies* 95 (1): 103–24.

Du, X. *et al.* (2022). "Representation of Cone-Opponent Color Space in Macaque Early Visual Cortices." *Front. Neurosci.* **16**, 891247.

Hardin, C. L. (1993). *Colour for Philosophers* (Expanded Edition). Indianapolis: Hackett.

Gottlieb, Joseph, and Ali Rezaei (2021). "When Nothing Looks Blue." *Synthese*, May. Springer Netherlands, 1–9.

Gow, Laura (2018). "Why Externalist Representationalism Is a Form of Disjunctivism." *Ratio* 31: 35–49.

Kuehni, R. G. (2004). "Variability in Unique Hue Selection: A Surprising Phenomenon." *COLOUR Research and Application* 29 (2): 158–62.

Kim, I., Hong, S. W., Shevell, S. K. & Shim, W. M. (2020). "Neural representations of perceptual colour experience in the human ventral visual pathway." *Proc. Natl. Acad. Sci.* **117**, 13145–13150.

Li, M. *et al.* (2022). "Perceptual hue, lightness, and chroma are represented in a multidimensional functional anatomical map in macaque V1." *Prog. Neurobiol.* **212**, 102251.

Martin, M. G. F. (2004). "The limits of self-awareness." *Philosophical Studies* 120: 37–89.

Pautz, A. (2006). “Sensory Awareness Is Not a Wide Physical Relation.” *Noûs* 40 (2): 205–240.

Pautz, A. (2012). "Tracking Intentionalism and Optimal Conditions: A Reply to Byrne and Tye." *On-Line Conference on Consciousness organized by R Brown*, 2012.

Pitt, David (2017). “The Paraphenomenal Hypothesis.” *Analysis* 77 (4): 735–41.

Sainsbury, Mark (2019). “Loar on Lemons.” In *Sensations, Thoughts, Language Essays in Honour of Brian Loar*, edited by Arthur Sullivan, Oxford UP.

Siuda-Krzywicka, K. & Bartolomeo, P. (2020). "What Cognitive Neurology Teaches Us about Our Experience of Colour." *Neurosci.* **26**, 252–265.

Tye, Michael (2013). “Transparency, Qualia Realism and Representationalism.” *Philosophical Studies* 170: 39–57.

Tye, Michael (2014). “What Is the Content of a Hallucinatory Experience?” In *Does Perception Have Content?* Edited by Berit Brogaard, 1–22. Oxford UP.

Tye, Michael (2015). “Yes, Phenomenal Character Really Is Out There in the World.” *Philosophy and Phenomenological Research* 91 (2): 483–488.