

Three Roads from Sensory Awareness to Dualism

1. Introduction

Neil Mehta has written a fantastic book. *A Pluralist Theory of Perception* develops a novel theory of perception that illuminates the metaphysical structure, epistemic significance, and semantic role of perceptual consciousness. By and large, I found the core tenets of Mehta's theory to be highly plausible and successfully defended. I could quibble with some parts (e.g., his claim that our conscious awareness of sensory qualities is non-representational). But I suspect our disagreements are largely verbal, and where they are non-verbal, they are minor. Instead of focusing on disagreements, in this commentary I wish to explore the metaphysical ramifications of Mehta's theory with respect to the mind-body problem. Mehta has a great deal to say about the metaphysics of perception. Much of it seems to me to be in tension with physicalism. But throughout the book he remains officially neutral on the truth of physicalism, "in reflection of [his] genuine uncertainty" (ibid: 100). I will try to show that Mehta's commitments lead almost inexorably to dualism (or, at least, away from physicalism) by giving three arguments against physicalism that centrally rely on premises to which Mehta is committed.

Some might take these arguments to show that Mehta's theory is unacceptable. For my part, I take them to reveal an interesting consequence of some plausible parts of Mehta's theory. In my view, the arguments below are simply three sound arguments for dualism, arguments that give everyone, not just Mehta, reason to reject physicalism. They are, moreover, importantly distinct from the usual epistemic-gap arguments (Mary, zombies, inverts, explanatory gap) that have been center-stage in debates over physicalism in recent decades (excessively so, in my view). The arguments below serve to illustrate what I believe to be a general pattern, and which I take to constitute "the real case against physicalism": that when one tries to develop natural and

plausible theories of this-or-that aspect of the mind (e.g., perception, consciousness, free agency, conceptual thought, rational insight, intentionality) *without* making any strained effort to keep one's theories consistent with physicalism—as Mehta does not—one very commonly runs into conflicts with physicalism.

The first argument (§2) is the color-dilemma argument. As Mehta suggests, it strongly seems impossible for sensory qualities like color to be instantiated wholly in virtue of physical facts. This gives us reason to think that physicalism is false if colors are instantiated. But if colors *aren't* instantiated anywhere in reality, then there is no plausible physicalist account of how we could be aware of colors in experience (as Mehta claims we are). Therefore, whether colors are instantiated or not, physicalism is false.

The second argument (§3)—the history-dependence argument—sets aside difficulties with reconciling color realism and physicalism. It begins with Mehta's claim that, in illusion and hallucination, we are aware of locally uninstantiated qualities. I argue that the best physicalist account of our awareness of locally uninstantiated properties entails that this awareness constitutively depends on historical facts (e.g., facts about the selectional history of our visual systems). But given Mehta's (very plausible) views about the connection between quality awareness and phenomenal character, if awareness of sensory qualities were history-dependent, then our *current phenomenology*—what it's like to be us right now—would also be history-dependent, which is absurd.

The third argument (§4) is the revelation argument. It relies on two of Mehta's central commitments. First, our awareness of sensory qualities is “partly revelatory,” in the sense that it “reveals a substantial part [...] of the essences of its targets” (ibid: 6). Second, as Mehta himself

claims, it “seems impossible [...] for purely physical facts to make it the case that there exists partly revelatory awareness” (ibid: 97).¹

2. The Color Dilemma

The argument in outline:

C1. If colors are instantiated, physicalism is false.

C2. If colors aren't instantiated, there is no plausible physicalist account of our phenomenal awareness of colors, so physicalism is false.

C3. Therefore, physicalism is false.

C1 is supported by the commonplace observation that color (“color-as-we-see-it,” “edenic color,” “sensible color”) seems to be absent from physical scientific descriptions of the world, and seems as though it couldn't be instantiated wholly in virtue of physical facts. This gives us reason to think that, if color is out there, it must be an extra property over and above the physical properties of objects. C1 can also be motivated via epistemic-gap arguments, such as the objectual analogues of zombie arguments, inverted-spectrum arguments, knowledge arguments, explanatory gap arguments.² There are also many Leibniz's-law arguments to the effect that colors have certain features that the relevant physical properties lack, e.g., structural-mismatch arguments to the effect that no set of physical properties that might reasonably be identified with human color space has the structural profile of human color space.³ (For a catalogue of ten or so Leibniz's law arguments against reductive theories of color, see Cutter (2021: 400-1).)

Mehta would sympathize with at least some of these motivations. He echoes many others in saying that “it strongly initially seems impossible” for something to be scarlet wholly in virtue

¹ This argument is a cousin of one suggested in passing by Mark Johnston (2011: 216n): “Herein lies the deep inadequacy of reductive materialism: There is no reduction of a relation which essentially involves disclosure to any combination of relations which essentially do not.”

² Byrne (2006), Cutter (2022).

³ Hardin (1988), Pautz (2006).

of its physical properties (ibid: 76). Moreover, Mehta has extra reason to take these seemings seriously. A cornerstone of his theory of perception is that our awareness of sensory qualities reveals a substantial portion of the essences of those qualities. Although this revelation is said to be partial, according to Mehta it is nonetheless infallible. That is, “consciously perceiving any sensory quality Q does not mistakenly seem to reveal any false claims of the form <It lies in the essence of Q that . . . >” (ibid: 25), where a mental state seems to reveal p when p seems true on the basis of rational reflection on that mental state.⁴ Now, the appearance that colors can’t be grounded in physical properties presumably arises on the basis of rational reflection on our perceptions of color. Together these claims *almost* entail that colors are not instantiated purely in virtue of physical facts. The only escape hatch I can see would be to insist that, while <colors aren’t grounded in physical properties> seems true on rational reflection, <it lies in the nature of colors that colors aren’t grounded in physical properties> doesn’t seem true on rational reflection. (Maybe instead it seems to lie in the nature of physical properties (in general?) that they don’t ground colors?) But this escape hatch does not strike me as particularly inviting.

Let’s turn to C2: if colors aren’t instantiated, there is no plausible physicalist account of our awareness of colors, so physicalism is false. C2 could be supported through the usual epistemic/explanatory-gap arguments, but it can also be motivated without any controversial inferences from epistemic gaps to ontological gaps. The main motivation for C2 is just that, given physicalism, there is no relation we bear to colors that could reasonably be identified with, or could reasonably be said to constitute, our awareness of colors if color eliminativism is true. (Note that C2 presupposes that we are aware of colors. Some philosophers—not Mehta—will insist that “awareness” is not an apt word for the mental relation we bear to colors in visual

⁴ This infallibility, which I would not dispute, is the basis of Mehta’s claim that sensory-quality awareness is “non-representational,” which I would dispute. For it seems to me that we are never sensorily aware of a quality without thereby representing it as instantiated (in a way that’s not immune from error).

experience. But the arguments below apply in equal force if we rephrase C2 in terms of visually *representing* colors, or in terms of colors visually *appearing* or *seeming* to be instantiated in one's environment.)

Here it will be useful to consider Mehta's response to the challenge of explaining how we can be aware of locally uninstantiated qualities (as Mehta insists we are in cases of illusion and hallucination). Without committing to any particular account, he offers one possible dualist account and one possible physicalist account as reasonable answers to the challenge. Here is his description of the physicalist account:

[T]he physicalist might identify a proper mechanism. They might say that my ancestors encountered instances of, for example, scarlet, which were of course spatiotemporally located, spatiotemporally extended, and causally efficacious. (They would thereby incur the burden of explaining away the appearance, which I discuss in the next chapter, that colors could not possibly be instantiated in a purely physical world.) The physicalist might add that my ancestors thereby developed the capacity to be sensorily aware of the universal scarlet and could exercise that capacity even when they were not causally interacting with instances of scarlet. Because I have a capacity with this evolutionary history, I, too, can be sensorily aware of the universal scarlet even if I have never encountered an instance of scarlet. (p. 71)

Notably, he does not mention any physicalist account that wouldn't "incur the burden of explaining away the appearance [...] that colors could not possibly be instantiated in a purely physical world." I conjecture that this is because he was unable to think of a half-plausible physicalist account of color awareness that accepts color eliminativism. And I conjecture that he was unable to think of such an account, not due to any intellectual deficiency on his part, but

because there isn't one. Rather, it would seem that the only half-plausible physicalist account of our sensory awareness of properties would identify or ground sensory awareness in some broadly physical relation between the perceiver and instances of the property in question.

Given physicalism, this is likely to involve some "detection" or "tracking" relation, perhaps of a sophisticated kind. In other words, awareness of a property is (at least partly) a matter of being in a state that is causally sensitive to instances of the property, or that simply correlates with the property, or has the function of doing so, or does so in optimal conditions, or something along these lines.⁵ Clearly, if colors aren't instantiated anywhere in reality, no story even roughly along these lines can get off the ground.

I grant that a sophisticated detectionist model might allow for awareness of *some* uninstantiated properties. For example, we might be aware of uninstantiated conjunctive properties (e.g., being a golden mountain) by semantically composing simpler representations that track the individual conjuncts. But this model wouldn't apply to our awareness of colors if color is missing from the world. Even if colors are regarded as conjunctions of hue, saturation, and brightness, these "conjuncts" are surely missing from the world if color is.

Semantic composition isn't the only way to account for awareness of uninstantiated properties within a broadly detectionist framework. Here is a crude model, inspired by Karen Neander's (2017: ch. 8) sophisticated teleosemantic account of how we might represent a missing shade of blue or a weird shape that is never actually instantiated. Suppose an organism has an internal dial that tracks ambient temperatures, where certain relationships between dial states meaningfully correspond to relationships between temperatures (e.g., equal intervals in the

⁵ Cf. Dretske (1995), Tye (1995), Neander (2017), Papineau (2021). "At least partly" because the physicalist may also wish to add further functional conditions, such as that the state in question is globally accessible, or is a target of second-order detection by a suitable self-scanning mechanism within the brain.

dial reading track equal differences in temperature). If the dial extends far beyond the range where its readings track ambient temperature, the dial might have possible settings that represent temperatures that aren't (and perhaps couldn't be) instantiated anywhere, such as temperatures colder than absolute zero. The general model is to start with some device that detects some members of a family of properties (e.g., some determinates under a common determinable). We then use structural similarities between the family of properties and the state space of the representational device to represent other properties in the same family that haven't been encountered, including some that are nowhere instantiated.

But this model won't help the color eliminativist explain our awareness of colors. If just one shade of blue is missing from the world, the model might explain our awareness of it. But it won't help if colors *as a family* are missing from the world. It is essential to the model above that the internal representational device anchors on a certain family of properties via physical detection of some members of that family. Without such an anchor, nothing could make it the case that the device represents one family of properties rather than another structurally similar family (e.g., temperature rather than volume or brightness).

As far as I can see, there is no remotely plausible physicalist account of how we could be aware of colors if colors aren't instantiated in the physical world. To be clear, the worry is not that physicalism precludes our being related in *any* way to uninstantiated colors. If colors exist (despite being uninstantiated) then we bear countless relations to them that pose no threat to physicalism about the mind. For example, let f be a function that maps everything on earth to the color blue, and let R be the relation: $\lambda x \lambda y (f(x) = y)$. You bear R to the color blue. This fact should not trouble the physicalist. But it seems clear that no relation like R could constitute our phenomenal awareness of sensory qualities. This is true even for cooked-up variants of R that are

designed to be coextensive with sensory awareness. Any attempt to reduce or ground sensory awareness to such relations would be a silly gimmick.

3. The History-Dependence Argument

The next argument sets aside difficulties with reconciling physicalism and color realism. It purports to show that our awareness of colors must involve a non-physical relation, even if colors themselves can (contrary to appearances) be instantiated in a wholly physical world. (Alternatively, for those who take colors to be extra, non-physical qualities of physical objects, the argument below can be taken to show that there are additional non-physical mental facts involved in our awareness of colors.) The gist of the argument is that physicalism can't accommodate the following pair of extremely plausible claims. First, the phenomenal character of my current experience doesn't constitutively depend on events in the distant past, such as facts about which properties caused certain neural states among my ancestors. Second, there is a necessary connection between my current phenomenology and the fact that I am aware of (or that I represent, or am acquainted with, or bear some interesting mental relation to) certain sensory qualities like colors and shapes. Although it is a wild oversimplification, there is a grain of truth in the thought that physicalists divide into camps based on which of these two intuitions tugs most strongly on them. Reductive representationalists like Tye, Dretske, and Lycan hold onto the second and (perhaps reluctantly) reject the first. Internalist non-representationalists like Block, Papineau, and Neander hold onto the first, and (perhaps reluctantly) reject the second. To my mind, both are overwhelmingly plausible, and to accommodate both we must reject physicalism.

We'll formulate the first intuition with the claim that phenomenal properties—properties characterizing what it's like to be a conscious subject at a given time—are “temporally intrinsic.”

To a first approximation, this means that their instantiation at a given time is entailed by the intrinsic character of the relevant time-slice of the entire world. This is a modal gloss on the idea that what it is like to be an individual at a given time t does not *constitutively* depend on what happened before t (or what will happen well after t , for that matter), though of course it may *causally* depend on what happened before t .⁶

For the purposes of the argument, I'll officially define a temporally intrinsic property as a property that never divides "temporary indiscernibles." Temporary indiscernibles are individuals that are qualitatively indiscernible (intrinsically and extrinsically) over a given stretch of time. (Officially: a in w over interval T is a temporary indiscernible of b in w^* over interval T^* iff the T -segment of w is a duplicate of—is intrinsically, qualitatively the same as—the T^* segment of w and the qualitative role that a occupies within T in w corresponds that that which b occupies within T^* in w .) Here's the rough idea. Take a time-slice of the whole world as it is now. (We'll let the time-slice have a bit of thickness—stretching, say, a few seconds or a few minutes from the current instant, to sidestep the complications mentioned in footnote 6.) Any world with a timeslice that is a duplicate of this one will contain, at the relevant time, a counterpart of you-as-you-are-now, someone indiscernible from you as far as matters stand at the present moment. This is your temporary indiscernible.

Examples may be helpful. Some properties that intuitively are *not* temporally intrinsic are: wearing a shirt that once belonged to a lumberjack, being the first man on the moon, being 80 years old, holding a pen that will be used tomorrow to sign a peace treaty, being a (real) \$100 bill, being a widow, and being a (biological) father. On the other hand, spatial properties like

⁶ We can let t be a short-ish interval, rather than an instant, to accommodate the possibility that experience is essentially temporally extended, e.g., across a specious present.

being round, or being next to a round object, are intuitively temporally intrinsic, as are many other physical properties like weighing 50 kg or having negative charge.⁷

Let us suppose that you are currently undergoing a hallucination or illusion as of a yellow object. Let V be a phenomenal property characterizing your total visual phenomenology, as you undergo this experience. From here, we can state the history-dependence argument as follows:

H1. Necessarily, anyone who has V is aware of yellow.

H2. V is temporally intrinsic.

H3. If physicalism is true, being aware of yellow is not temporally intrinsic.

H4. Therefore, physicalism is false.

H1 just seems obvious upon introspection. It seems inconceivable that anyone should have an experience phenomenally just like (say) an ordinary experience of a lemon without thereby being aware of yellow. (Again, some will have qualms about whether “aware” is quite the right word here. But as long as one agrees that there is some significant mental relation to yellow that is entailed by our total visual phenomenology (e.g., representing yellow, being such that it seems to one that yellow is instantiated, being in a position to refer to yellow), a modified version of the argument should go through.⁸)

H2 is a consequence of the more general claim that phenomenal properties are temporally intrinsic. Here we hit intuitive bedrock, but this seems about as obvious as anything in

⁷ As the example of being next to a round object illustrates, being “temporally intrinsic” does not entail being intrinsic. Thus the claim that phenomenal properties are temporally intrinsic is importantly different from the claim that phenomenal properties are “narrow” or intrinsic (Horgan and Tienson 2002, Hawthorne 2004, Papineau 2021). While the latter might also be true, I agree with Pautz (2014) that it lacks *a priori* or intuitive support; indeed, naive reflection on our sensory phenomenal properties would seem to suggest the opposite, that phenomenal properties essentially involve acquaintance relations to qualities in our environment. As C.D. Broad observes (1952: 5): “In its purely phenomenological aspect *seeing* is ostensibly *saltatory*. It seems to leap the spatial gap between the percipient’s body and a remote region of space.”

⁸ Pautz (2010) defends such a modified version of H1 under the heading of “external-directedness,” which plays a central role in his arguments against identifying phenomenal properties with intrinsic neural properties.

philosophy. It is no less plausible than the claim that spatial properties like being spherical are temporally intrinsic.

The case for H3 is just that, as mentioned earlier, the best physicalist accounts of awareness will be broadly detectionist, and the most promising detectionist accounts will identify awareness with a temporally extrinsic property (e.g., being in a state with the function to indicate such-and-such, where functions are cashed out in terms of selectional history (Dretske 1995, Tye 1995, Neander 2017)). Pautz (2014: 174) notes that it would be very difficult to devise a reductive theory of phenomenal representation that validates the intuition that phenomenal properties are temporally intrinsic, “since all standard theories of representation appeal to historical facts or forward-looking facts to help settle what external features our inner states have the ‘biological function’ of tracking or track under ‘optimal conditions’ in the present.” Parallel remarks apply to phenomenal awareness.

4. The Revelation Argument

The third argument is the Revelation Argument. Its central premise is Mehta’s thesis that experience essentially reveals a substantial portion of the essences of certain universals: the sensory qualities. Let “Q-awareness” name the relation of conscious awareness we bear to sensory qualities.⁹ Mehta claims that Q-awareness is revelatory. Roughly, this means that when a subject is Q-aware of a sensory quality, he is disposed to have many true seemings, and no false seemings, concerning the essence of that sensory quality. (There are further qualifications, e.g., the subject needs to be conceptually sophisticated and engaged in rational reflection, which I’ll ignore below for simplicity.)

The argument:

⁹ Mehta calls it “deep awareness,” presumably because he thinks it gives us deep insight into the natures of sensory qualities. But this label would seem tendentious in the present context.

R1. If physicalism is true, Q-awareness is not a revelatory relation.

R2. Q-awareness is a revelatory relation.

R3. Therefore, physicalism is false.

R1 is hard to resist. First, some (including Mehta (ibid: 97)) may find it directly intuitively plausible. Second, physicalism goes naturally with a reductive account of color, but every reductive account of color conflicts with claims about the essence of color that seem true on the basis of rational reflection of color experience (or so I've argued elsewhere (Cutter 2018, 2021)). Third, and most importantly, we should accept (1) because, among physicalism-compatible relations, the most plausible candidates for being, or grounding, Q-awareness are not revelatory. That is, these relations do not entail that one has any tendency to have many true seemings and no false seemings about the target's essence. For example, any kind of straightforward tracking relation, such as being in a globally accessible state with the function of indicating Fs, would not entail any such tendency.

To be clear, the claim is not that revelatory relations cannot be instantiated in a physicalist world. The claim is that no revelatory relation whose instantiation is compatible with physicalism can plausibly be said to constitute Q-awareness. As before, it is easy to cook up a physicalist-compatible relation to properties that is revelatory. Indeed, given any relation R, we can define a stronger relation R* that is revelatory:

$xR^*y =_{\text{def}} xRy$ and x has many true seemings and no false seemings about the essence of y .

For R, we can plug in some physicalistically acceptable tracking relation between organisms and sensory qualities. R* would then be the relation: *x tracks y and x is disposed to have many true*

seemings, and no false seemings, about the nature of y. This relation is revelatory. And, since tracking is physicalism-compatible, R* is physicalism-compatible as well, provided that having seemings about the natures of sensory qualities is physicalism-compatible.¹⁰

But it would be deeply implausible to identify Q-awareness with anything like R*. Apart from being gimmicky, the identification would raise a serious explanatory challenge. Why is it that when we track a property, we tend to track* the property as well? More precisely, why is it that, whenever we track a property in a way that directly leads us to have some seemings about its essence, those seemings are always true? This calls for explanation, and I don't see how the physicalist could explain it. It would be as though, whenever we stood next to an object, we have many true seemings and no false seemings about that object's essence. This would be a bizarre circumstance that calls for explanation. What mysterious force prevents us from merely standing next to a thing without an accompanying set of reliable seemings? The physicalist view under consideration posits a similarly bizarre and seemingly inexplicable circumstance.

Turning to R2: Mehta's argument for revelatory awareness begins with examples. An experience of scarlet, mango orange, and lime green positions you to know a number of truths about the essence of these qualities, e.g., (i) these qualities exclude one another, (ii) the first is more similar to the second than to the third, (iii) any instance of scarlet is spatially extended, and so on. Mehta suggests that even if one has never experienced color before and then miraculously gains normal sight, one could come to know these truths simply by having the relevant color experiences and rationally reflecting on them. This all seems right. Like Mehta, I find it hard to doubt that claims like (i)-(iii) are true, that they capture something "essential" about the relevant colors (leaving open how best to understand "essence" talk), that we can know such truths, and

¹⁰ I myself doubt that the latter condition holds, but we can grant it for the sake of argument, since the worries here are more general concerns about physicalism's compatibility with any form of mentality, which have little to do with revelation specifically.

that consciously perceiving them has a role in explaining this knowledge. So I think the argument from examples makes a solid case for something in the ballpark of R2. More work would need to be done, of course, to generalize the point to other sensory qualities and other sense modalities, and to justify Mehta's view that the revelation associated with Q-awareness is infallible. For this, I'll refer readers to the book, which offers further arguments for his preferred revelation thesis and responses to objections. But I'll note that, for the purposes of arguing against physicalism, we likely won't need all the trappings of Mehta's preferred revelation thesis. For example, we probably won't need infallibility. As long as being Q-aware of a sensory quality generally puts a (conceptually sophisticated, rational) subject in a position to know, or have many true seemings, about the quality's essence, the argument should go through, since it's plausible that the relevant physicalism-compatible relations lack this feature.

If Mehta remains agnostic about physicalism after reflecting on the arguments above, I would be interested to hear which premises he is unwilling to accept and why.

References

- Broad, C. D. (1952). Some elementary reflexions on sense-perception. *Philosophy* 27 (January):3-17.
- Byrne, Alex (2006). Color and the Mind-Body Problem. *Dialectica* 60 (3):223-244.
- Cutter, Brian (2018). Paradise Regained: A Non-Reductive Realist Account of the Sensible Qualities. *Australasian Journal of Philosophy* 96 (1):38-52.
- Cutter, Brian (2021). Perceptual illusionism. *Analytic Philosophy* 62 (4):396-417.
- Dretske, Fred (1995). *Naturalizing the Mind*. MIT Press.
- Hardin, Clyde L. (1988). *Color for Philosophers: Unweaving the Rainbow*. Hackett.

- Hawthorne, John (2004). Why Humeans are out of their minds. *Noûs* 38 (2):351-58.
- Horgan, Terence & Tienson, John (2002). The Intentionality of Phenomenology and the Phenomenology of Intentionality. In David John Chalmers (ed.), *Philosophy of Mind: Classical and Contemporary Readings*. New York: Oxford University Press USA. pp. 520--533.
- Johnston, Mark (2011). On a neglected epistemic virtue. *Philosophical Issues* 21 (1):165-218.
- Mehta, Neil (2024). *A Pluralist Theory of Perception*. Cambridge, Massachusetts: The MIT Press.
- Neander, Karen (2017). *A Mark of the Mental: A Defence of Informational Teleosemantics*. Cambridge, USA: MIT Press.
- Papineau, David (2021). *The Metaphysics of Sensory Experience*. Oxford: Oxford University Press.
- Pautz, Adam (2006). Can the physicalist explain colour structure in terms of colour experience?. *Australasian Journal of Philosophy* 84 (4):535 – 564.
- Pautz, Adam (2010). Do theories of consciousness rest on a mistake? *Philosophical Issues* 20 (1):333-367.
- Pautz, Adam (2014). The Real Trouble for Armchair Arguments Against Phenomenal Externalism. In Mark Sprevak & Jesper Kallestrup (eds.), *New Waves in Philosophy of Mind*. London: Palgrave-Macmillan. pp. 153-181.
- Tye, Michael (1995). *Ten Problems of Consciousness: A Representational Theory of the Phenomenal Mind*. MIT Press.