

## ***Rosenthal on Mental Qualities***

Alex Byrne, MIT

Rosenthal couples his higher-order thought theory of consciousness with a theory of “mental qualities”, properties of mental states. The first thesis of this paper is that there are no mental qualities as Rosenthal conceives of them. The second thesis is that Rosenthal’s residual insights are significant. They naturally lead to a simple first-order theory of consciousness.

### **1: What are mental qualities?**

‘Mental quality’ is Rosenthal’s term. Here is what he means by it:

... a visual experience of the redness of a tomato is not red in the way the tomato is. But that does not preclude its having some mental quality, distinct from the red of the tomato but characteristic of the visual sensations we have when we see red objects in standard conditions of illumination. Mental qualities are unlike the perceptible qualities of physical objects in several ways. For one thing, they are properties of states, rather than objects. Equally important, they are not perceptible; we do not come to be aware of them by perceiving them. (Rosenthal 2000: 118)

A mental quality, then, is a property of a “sensation” or “experience”; these Rosenthal thinks of as states or conditions of perceivers, “ontologically on a par with properties of objects” (Rosenthal 1999a: 106). “Sensations”, he writes, “are states of sentient creatures” (197). Mental qualities correspond to perceptible qualities. When I see a tomato in good light, I am aware of a distinctive quality of the tomato, a particular yellowish-red shade. (For simplicity, we can ignore shades in favor of determinables like red.) That quality of the tomato has nothing to do with

me—the tomato would have been red even if I hadn't glanced at it, and it will continue to be red when I leave the kitchen. It is quite different with mental qualities. When I look at the tomato, I am in a perceptual state with a distinctive mental quality, which we can label 'red†'. Indeed, when I look at any red object in suitable conditions I will be in a perceptual state which is red†. As Rosenthal puts it (using 'red' for 'red†'):

The mental quality of red occurs when one consciously sees something red under suitably standard conditions (Rosenthal 2010: 375)<sup>1</sup>

And (for the most part) conversely: whenever I am in a red† perceptual state, I am looking at red object in suitable conditions. There are exceptions: I can be in a red† perceptual state if I am suffering a color illusion and there is nothing red before me. And similarly if I have a so-called red afterimage: I am in a red† perceptual state, although I see nothing red (see Rosenthal 1999a: 107).

Do all mental states have mental qualities? One might suspect, given the explanation so far, that Rosenthal's answer is "no". And indeed it is:

[T]houghts, desires, expectations, wishes, and the like [do not exhibit] mental qualities.  
(Rosenthal 2010: 375)

Mental qualities are not representationally inert, like the color or font of letters on the page. They are not, that is, what Block calls "mental latex"—"Mental properties of the experience that don't represent anything" (Block 1996: 29). Rather, mental qualities are what Block calls "mental paint", mental properties that *do* represent:

---

<sup>1</sup> Rosenthal sometimes uses 'red\*' to stand for the mental quality of red (Rosenthal 2005b: 213).

It is these mental qualities by which perceptual experiences, unlike nonperceptual thoughts, represent things in a distinctively qualitative way. (Rosenthal 2000: 119)

Specifically, the mental quality or (as Rosenthal sometimes says) “mental color” red† represents the color red, purple† represents the color purple, and so on.

From what has been said so far, mental qualities might be theoretical entities, postulated to explain certain perceptual or psychological phenomena—like, for instance, words in a language of thought. If there is a language of thought, it is not evident from ordinary experience. Unlike the English word ‘dog’ or the Italian word ‘cane’, one cannot introspect the mentalese word DOG (which in any case is not spelled that way). But mental qualities are not like that:

Mental colors...are just the introspectible mental counterparts of...physical color properties. (Rosenthal 1999a: 103)

The colors stand in relations of similarity and difference to each other: purple and red are more similar to each other than either is to green, for instance. Whether or not introspection reveals counterparts of the colors, it certainly does *not* reveal counterparts of the colors that stand in *different* similarity relations. It is not surprising, then, that according to Rosenthal:

...the similarities and differences each mental color has to all other mental colors are homomorphic to the similarities and differences that the corresponding physical color property has to all the other physical color properties. (Rosenthal 2001: 90)<sup>2</sup>

---

<sup>2</sup> Rosenthal uses ‘homomorphism’ rather than ‘isomorphism’, which leaves open the possibility that there are similarities and differences between physical color properties that are not reflected in similarities and differences between mental colors.

## 2: The double-meaning theory

If the presence of the mental color red<sup>†</sup> is readily detectable by anyone with normal color vision and introspective capacities, one would expect that we have a word for it. And according to Rosenthal, we do. ‘Red’ has a “double-meaning” (Rosenthal 1999a: 100). In one sense, red is “physical red”, a property of tomatoes.<sup>3</sup> In another sense, red is “mental red”, red<sup>†</sup>, a property of sensations or experiences. The “widespread tacit assumption of univocality is unfounded”:

Terms for color and shape must therefore mean different things when applied to physical objects and to visual experiences, and attribute distinct properties in the two cases. A similar result holds for words that ascribe other perceptible properties. (Rosenthal 1999a: 96)

Let us examine Rosenthal’s case for the double-meaning theory.<sup>4</sup>

---

One complication is worth noting (although will not be pursued here). Imagine seeing one red and one green marble. On Rosenthal’s view, one’s perceptual state has a certain qualitative character, due to its possession of various mental qualities. If those include red<sup>†</sup> and green<sup>†</sup>, then one’s mental state is *both* red<sup>†</sup> and green<sup>†</sup>. It is hard to understand how that can be, if red<sup>†</sup> and green<sup>†</sup> are as dissimilar as red and green. Another issue concerns the difference between seeing two red marbles and seeing only one. If both states are red<sup>†</sup> then the difference between them with respect to color phenomenology collapses. (These problems would not arise if mental qualities were properties of sense data, rather than perceptual states.)

<sup>3</sup> Although Rosenthal himself thinks that “physical red” is a light-dispositional property (see note 15 below), the only assumption he needs is that it is a property of tomatoes that is not constitutively connected to the mental color red<sup>†</sup>.

<sup>4</sup> One point can be conceded at the outset. Rosenthal does have good reply to Jackson’s worry about the double-meaning theory. “Such a double use, [Jackson] holds, would be “a linguistic accident, a fantastic fluke”” (Rosenthal

## 2.1 “Red experiences”

One might think that the single-meaning theory is supported by our ways of describing perceptual experiences.<sup>5</sup> In particular, ‘red’ in the phrase ‘an experience *of* red’ surely has its standard “physical red” interpretation. (Likewise, ‘tomato’ in ‘an experience of a tomato’, has its ordinary “physical fruit” sense.) Rosenthal agrees that ‘red’ means *physical red* in ‘an experience of red’, but leverages this observation into an argument *against* the single-meaning theory. We can describe an experience of a tomato as an experience of red, or as an experience of a red tomato, or as an experience of a red physical object, but we can also describe it as a “red experience”. And this, Rosenthal argues, shows that ‘red’ in ‘red experience’ cannot mean *physical red*:

The predicates we apply to the experience are ‘red’ and ‘of a red physical object’; so ‘red’ in ‘red experience’ must mean the same as the entire predicate, ‘of a red physical object’. (Rosenthal 1999a: 97)

That is: ‘red experience’ and ‘experience of a red physical object’ have the same meaning, so ‘red’ in the former expression means the same as ‘of a red physical object’ in the latter. And

---

1999b: 171; quotation within quotation from Jackson 1977: 73). Rosenthal might be over-interpreting Jackson at this point—all Jackson actually says is that “Obviously, it is not a linguistic accident” and that the “simplest explanation” is the single-meaning theory. But in any event, Rosenthal notes that any plausible version of the double-meaning theory will take the ambiguity in ‘red’ to be polysemy, and so not a fluke at all.

<sup>5</sup> The “single-meaning theory” merely denies that ‘red’ has two meanings à la Rosenthal. It is not the view that ‘red’ only has one meaning (which of course it doesn’t).

since ‘of a red physical object’ does not mean the same as ‘red’ in ‘red tomato’, the double-meaning theory is true.<sup>6</sup>

The problem with this lies in the first step, which tacitly assumes that ‘red experience’ has a compositional semantics like ‘red tomato’: the meaning of ‘red experience’ is derived from the meaning of its components, the adjective ‘red’ and the noun ‘experience’, and their arrangement (adj+NP). But ‘red experience’ (in the intended sense) occurs almost exclusively in philosophical contexts, and is very plausibly an idiom, like ‘red tape’, ‘red cent’, and so on.<sup>7</sup> One can understand the expression ‘red cucumber’ the first time one sees it, by knowing the meaning of ‘red’ and ‘cucumber’, and applying compositional rules. Not so with ‘red tape’ or ‘red cent’—one has to learn the meaning of the entire expression.<sup>8</sup> And likewise with ‘red experience’. In a philosophical context one might correctly conjecture what it means, but clues need to be provided. Given that ‘red tape’ is an idiom, there is no need to seek some special meaning for ‘red’. And likewise for ‘red experience’.

## *2.2 The afterimage argument*

Rosenthal suggests a related argument that turns on the fact that we describe afterimages as having colors. The sentences ‘This tomato is red’ and ‘This afterimage is red’ can both be true, as uttered in appropriate contexts. But if ‘red’ is univocal:

---

<sup>6</sup> For present purposes, meanings can be taken to be intensions.

<sup>7</sup> ‘Red sensation’ sounds more acceptable outside philosophy, and might be used to describe ostensible colored spots or flashes experienced during a migraine. But not when viewing a British pillar box in daylight.

<sup>8</sup> Rosenthal himself says that expressions like ‘red experience’ are “idioms” (1999a: 97) but clearly intends this loosely.

The property of being red that physical objects have is the same as the property of being red that afterimages have. (Rosenthal 1999a: 101)

However, it seems most unlikely that afterimages, which “cannot reflect or emit light” (101), have the kinds of colors that we attribute to physical objects like tomatoes. Hence, ‘red’ is not univocal: it has a different meaning in ‘red afterimage’ than it does in ‘red tomato’.

This argument needs a separate treatment, because unlike ‘red experience’, ‘red afterimage’ is implausibly idiomatic. ‘Red experience’ is not in the typical non-philosopher’s vocabulary, but talk about the colors of afterimages is quite familiar. One can induce an afterimage experience in someone unfamiliar with afterimages, for instance by getting them to look at a white wall after staring at a blue-and-green-striped square for a minute or so. If one asks the subject “What do you see?”, a natural reply would be “An orange-and-pink striped square”, to which one might respond “That’s a complementary afterimage”. One can then go on to ask other intelligible questions about the subject’s afterimage—whether it’s moving, or makes a sound, or whatever. Semantically, ‘red afterimage’ looks very much like ‘red tomato’.

What are afterimages? In a classic paper on the “imagery debate”, Block makes a helpful distinction:

To avoid confusion, it may be useful to distinguish mental images in my sense (the internal representations of imagery) from what might be called *phenomenal images*, the “seeming objects of image experiences,” the things, if such there be, which are more or less vivid, sometimes orange-and-pink-striped, etc. (even though they reflect no light. (Block 1983: 507)

Block is actually talking about the sort of imagery studied in experiments on “mental rotation” and the like—as when one visualizes the sofa at various angles to see if it can be carried up the staircase. But the distinction applies to afterimages just as well. When one has an orange-and-pink-striped afterimage, the *phenomenal image* is the “seeming object” of the experience, the ostensibly orange-and-pink striped patch that might appear to be located on the far wall. The *mental image* is the internal representation of the phenomenal image—on some views, a neural representation in some kind of linguistic format. And there’s one more thing to distinguish: the *imagistic experience*—the state of being ostensibly aware of the phenomenal image. (Cf. Block: “even in this literature, ‘mental image’ sometimes denotes the experiences we have in imagery” (506).)

Now when people untutored in philosophical subtleties talk about their afterimages, what are they talking about? Phenomenal images, mental images, or imagistic experiences? Surely *phenomenal* images—those are the items that are, or seem to be, orange-and-pink-striped and located on the far wall. Indeed, it can seem obvious that afterimages are *seen*, even to authorities in vision science. According to the psychologist Richard Gregory, for example, an afterimage is an “image seen immediately after the intense stimulation of the eye by light has ceased” (quoted in Phillips 2013: 413<sup>9</sup>). Since it is not at all plausible that we see our experiences or neural representations the obvious interpretation is that Gregory had phenomenal images in mind.

For another example, take Smart’s classic paper “Sensations and brain processes”. With respect to afterimages, Smart carefully distinguishes between (in present terminology)

---

<sup>9</sup> See also Smart 1959: 149.



phenomenal images and imagistic experiences. The thesis of his paper is that imagistic experiences, not phenomenal images, are brain processes:

I am not arguing that the after-image is a brain-process, but that the experience of having an after-image is a brain-process. It is the *experience* which is reported in the introspective report ["I have a yellowish orange after-image"]. (Smart 1959: 150)

What are afterimages, then? According to Smart, they are nothing:

There is, in a sense, no such thing as an after-image or a sense-datum, though there is such a thing as the experience of having an image. (151)

Sensible enough.<sup>10</sup> But then the case for the double meaning theory collapses, because it relies on the premise that 'This afterimage is red' is *true*.

In fact, even if phenomenal images exist, the view that phenomenal images and objects like tomatoes can share colors is not indefensible. Light might not be so intimately connected with color as some think. And if so, 'This afterimage is red' and 'This tomato is red' could *both* be true, with 'red' univocal throughout. Rosenthal makes this point himself: "Univocality has far more credibility if the terms or color and shape apply to mental objects of some sort, rather than mental states" (Rosenthal 1999a: 106). Accordingly, Rosenthal's version of the afterimage argument for the double meaning theory assumes that afterimages are experiences or states, not objects or phenomenal images.

---

<sup>10</sup> Phillips officially takes afterimages to be imagistic experiences: "Afterimages are a kind of perceptual experience, and in enjoying them we are not literally aware of any images" (2013: 421). However, for most of his paper he treats them—much more naturally—as phenomenal images (see fn. 5, 421), as ostensible objects that can appear to move, to be occluded, and so on.

Rosenthal is free to use terminology however he likes, but if his version of the afterimage argument is to be persuasive, we need a reason for supposing that ‘afterimage’ (in some ordinary sense) refers to imagistic experiences, and that ‘red afterimage’ has a compositional semantics. And, as Rosenthal notes, this is unobvious.

we describe afterimages by saying straight-out that they are red and round, we tend not to describe other visual experience that way; rather, we describe them as being *of* something red or round, or *of* a red disk. (107)

Why, then, does Rosenthal think that ‘afterimage’ refers to experiences? He gives a reason in this passage:

...we describe afterimages differently from other visual experiences not because afterimages are mental objects, but because they are obvious cases of hallucinations.

When we have an afterimage, we know full well it is not a case of our seeing a physical object. So we describe it using the terms we would have applied to the physical object, for example, by saying it is red and round. (107)

But there are two problems with this argument. First, how does Rosenthal get from “knowing full well it is not a case of our seeing a physical object” to “we describe [the experience] using the terms...”? Instead of applying ‘red’ to the experience, why don’t we just say “OK, I’m not seeing a red and round physical object, but I am having an hallucination/experience *of* something red and round”? Rosenthal thinks that we don’t say that because the ‘of’ construction indicates that there *may* be a physical object:

when we take a visual experience to be the result of a corresponding physical object, or we think that it may be, we signal our belief that some red, round physical object may be involved by saying...that our experience is *of* something red and round. (107)

Rosenthal's basis for thinking that fully informed hallucinators switch to applying predicates like 'red' to their experiences is that they have no other linguistic resources. This is dubious. It is standard to describe hallucinators as having, say, "hallucinations *of* faces", and hallucinators can describe themselves in that way.

The second problem is more serious. Although it may be evident that afterimage experiences are not cases of seeing a *physical* object, it is far from evident that they are not cases of seeing—or, more cautiously, being aware of—an object. Gregory, quoted above, claims that afterimages are seen, and there are many more examples. In "Proof of an external world" Moore gives the example of negative afterimages ("grey four-pointed stars") as things that he sees, and which are "presented in space", but which are not physical objects, because "it is not conceivable that anyone besides myself should have seen any one of them" (Moore 1939: 151-2). Moore knew that he was not seeing physical objects, but still thought he was seeing objects, to which color and shape predicates could be applied. When he describes his afterimages as grey, there is no reason to think that he was applying 'grey' to his experience.

Finally, there is some direct evidence against the double-meaning theory. One can compare the color of one's afterimages to the color of physical objects: 'The afterimage on the left is exactly the same shade of red as the tomato on the right', and so on. This kind of construction makes a univocal interpretation of 'red' hard to avoid. If 'red' had two meanings such comparisons would sound anomalous, like 'You and a low-carb diet are both healthy'.

### 3: An argument against mental qualities

The conclusion so far is that the double-meaning theory is false. In particular, ‘red’ is univocal in ‘red afterimage’ and ‘red tomato’. This does not mean that redness is a property of physical objects like tomatoes and strange objects like afterimages. Smart’s view, that there are no afterimages (even though there appear to be) remains available.

We can use the falsity of the double-meaning theory as a premise in an argument that there are no mental qualities:

P1. If there are mental qualities, the double-meaning theory is true

P2. The double-meaning theory is not true

C. There are no mental qualities

Could Rosenthal deny P1? Perhaps there are mental qualities, but—as Kind puts—they are just “*difficult* (but not impossible) to attend [to]” (Kind 2003: 230). And if mental qualities are elusive, there is no reason to expect we would have words for them.

However, Rosenthal’s higher-order thought theory of consciousness requires *easy access* to mental qualities. On the HOT theory, if one (consciously) sees a red tomato, one has a thought about the mental qualities of one’s perceptual state—that is what makes the perceptual state conscious:

if I visually perceive something red...[and] that perception is conscious, I also have a HOT to the effect that I’m in a state with a red<sup>†</sup> mental quality. (Rosenthal 2005b: 215; superscript added<sup>11</sup>)

---

<sup>11</sup> See also Rosenthal 2005a: 16.

If mental qualities were elusive, then conscious perception would be much harder to achieve than it evidently is.<sup>12</sup>

A HOT theorist need not dispute the conclusion C: Rosenthal's theory of consciousness does not require his theory of mental qualities. But once mental qualities are gone, the HOT theory becomes harder to motivate. One of Rosenthal's insights (to be discussed below) is that consciousness involves awareness of *qualities*. If red† is unavailable, then the next best candidate is the color red. But that points towards a first-order theory of consciousness.

#### **4: Three Rosenthalian insights**

We have come to praise Rosenthal, not to bury him. Consigning mental qualities to the dustbin of history enables us to see his important insights more clearly. Three are particularly relevant here.

Following Rosenthal, let *state* consciousness be “the property mental states have of being conscious” (Rosenthal 1994: 235). If we restrict attention to perceptual states, then we can elucidate state consciousness by saying that a perceptual state is conscious just in case there is “something it is like” for one to be in the state (adapting the formulation in Nagel 1974):

There is always something it's like for one to be in a conscious qualitative state; by contrast, there is nothing it's like for one to be in a...state that isn't conscious. (Rosenthal 2005b: 193)

---

<sup>12</sup> Another route to denying P1 would be to argue that although we have easy access to mental qualities, we lack words for them because they are of no interest to us. That does not sound promising, since we certainly have an interest in describing our feelings and perceptions.

Again following Rosenthal, *transitive* consciousness is *awareness*:

I'll refer to our being conscious *of* something, whether a mental state or anything else, as *transitive consciousness*. (Rosenthal 1994: 235)

Rosenthal's first insight is that state consciousness is to be explained in terms of transitive consciousness, consciousness *of*. His second insight is that the relevant items of which we are conscious, are *qualities*. And his third insight is that the bearers of these qualities "can occur without being conscious" (Rosenthal 2005a: 13). That is, these qualities can be instantiated in the absence of state consciousness (and, for that matter, transitive consciousness).

Why are these—if not a trio of insights—at least plausible hypotheses?

For simplicity, let us restrict attention to clear examples of *perceptual* states—seeing a red tomato, hearing a clap of thunder, and so on. (We are therefore ignoring other mental states that are in some intuitive sense "qualitative", like having an itch or feeling depressed, but which are not clearly perceptual.<sup>13</sup>)

When one consciously perceives, one is conscious *of* something. That is not controversial. Take a paradigmatic case of conscious perception. Could someone be conscious *of* exactly the same items—facts, objects, qualities, property-instances, sense-data, phenomenal character, whatever—and yet *not* be in a conscious perceptual state? If you think it helps, add that the perceiver is conscious of the same items *in the same way*. (Cf. the first indented quotation from Rosenthal below.) An "inner light" picture of consciousness can make an affirmative answer seem inevitable. Having a working visual system and facing a tomato in sunshine on will suffice

---

<sup>13</sup> Rosenthal thinks that pain states are not perceptual (1986: 25); for the contrary view see Byrne 2018: 147-9.

for being conscious *of* the tomato, but it won't guarantee *consciousness*. For that, the "internal light" in the mind needs to be on as well; the internal light allows one's awareness of the tomato to be *conscious*, without being yet another item one is conscious *of*. However, this makes very little sense, even as a metaphor: if one sees the tomato through one's office window, and a literal "internal light" like a desk lamp is on, one is normally aware *of* the light. Rosenthal's hypothesis, that state consciousness can be explained in terms of transitive consciousness, is an intuitive theoretical starting point.

A perceptual state is conscious, then, because the subject in the state is conscious *of* certain items. What items? As Rosenthal says, "[m]ental states are conscious in different ways; what it's like for one to have a conscious sensation of red, for example, differs from what it's like for one to have a conscious sensation of green" (Rosenthal 2005a: 4). All conscious sensations of red have a conscious aspect in common, and similarly for conscious sensations of green. The natural hypothesis, then, is that conscious sensations of red involve consciousness of a particular quality—call it 'R'—while conscious sensations of green involve consciousness of a *different* quality—call it 'G'. As Rosenthal puts it:

What it's like for one...is fixed by how one is conscious of the quality.

Of course, Rosenthal's theory is more specific:

What it's like for one to be in a state that has that mental quality...is fixed by how one is conscious of the quality, and so by the way one's HOTs represent it. (Rosenthal 2005b: 209)

*Conscious of* becomes *has a thought about*, and R is identified with the mental quality red<sup>†</sup>, a property of mental states, not tomatoes. But this is an addition: Rosenthal's first two insights can survive its deletion.

Rosenthal's third insight is that the relevant qualities—R, G, and so on—can be instantiated without consciousness.<sup>14</sup> *x* can have R, even though *x* is not itself conscious, and no one is currently conscious of *x* and thereby conscious of R. Granted the first two insights, the third is—if not inevitable—at least very plausible. If R is not the mental quality red<sup>†</sup>, the obvious alternative is that R is the color red. And the hypothesis that tomatoes *are* red, and moreover remain red when in closed refrigerators, has held up pretty well to centuries of philosophical skepticism.<sup>15</sup>

### **5: Three insights - red<sup>†</sup> = Dretske**

Here's where we are: mental qualities have been swept away, while Rosenthal's three insights have been retained. The only credible candidate for R is the color red, the familiar property of tomatoes. This gives us a theory of consciousness held—at one time—by Dretske:

Conscious mental states—experiences, in particular—are states that we are conscious *with*, not states we are conscious *of*... When I speak of experiences making a creature aware of something, I mean to be speaking of the properties, not the objects (if any) the

---

<sup>14</sup> As he sometimes puts it: they are not “intrinsically conscious” (Rosenthal 1999b: 161), or: “the property of being conscious is not intrinsic to such states” (Rosenthal 2005b: 179).

<sup>15</sup> In fact, the more specific view—endorsed by Rosenthal (2005a: 11)—that identifies the colors with light-dispositional properties is also in decent shape (Byrne and Hilbert 2020). For a defense of the view that colors are *sui generis* (and instantiated), see Allen 2016.



experience makes the creature aware of...pinkness, rat-shapes...[when creatures] “see” pink rats. (Dretske 1995: 101-2)

And:

My visual experience of a barn is conscious, not because I am introspectively aware of it (or introspectively aware that I am having it), but because it (when brought about in the right way) makes me aware of the barn. It enables me to perceive the barn. (Dretske 1993: 280)

To this theory, Rosenthal has (at least) two objections. First, he complains that:

... it's arguable that this account implies that all perceiving is conscious, since perceiving something automatically makes one conscious of it. (Rosenthal 2005b: 180)

This is more than arguable. If I perceive a barn, I am conscious of some (apparent) quality of the barn—if the barn is red and conditions are good, the color red will be one of those qualities. On Dretske's view, it follows that my experience is conscious, with its distinctive qualitative character due (in part) to the fact that I am conscious of the color red.

If I had seen a green barn, I would have been conscious of a different color, and so would have had a conscious experience with a different qualitative character.

Rosenthal thinks that the implication is a problem because there is:

...compelling evidence that perceptual sensations also occur without being conscious.

Masked priming experiments provide situations in which detailed qualitative information occurs of which subjects are wholly unconscious, results that fit with everyday cases of

peripheral vision and subliminal perceiving. And striking work with blindsight patients suggests the same conclusion. (Rosenthal 2000: 108-9)<sup>16</sup>

However, it is by no means clear that this evidence *is* “compelling”.<sup>17</sup> Admittedly orthodoxy is on Rosenthal’s side. Block, for instance, argues that “[s]eeing is a single fundamental natural kind, of which conscious and unconscious seeing are sub-kinds” (Block and Phillips 2016: 169). But Phillips, in particular, has mounted an extensive defense of the view that perception is always conscious (Block and Phillips 2016, Phillips 2018). Representations of masked primes may have perceptual effects merely by facilitating subsequent perceptual processing, without those representations characterizing qualities perceived.<sup>18</sup> Blindsight patients deny (at least in some experimental conditions) that they see anything at all, but that might be the result of a conservative response bias. Blindsight could simply be severely degraded (conscious) vision.<sup>19</sup> Indeed, even if we concede that these explanations are not the whole story, another possibility is that subjects who purportedly perceive unconsciously sometimes suffer from an introspective or memorial failure rather than the absence of consciousness.<sup>20</sup> At a minimum, the issue is not settled in Rosenthal’s favor.

---

<sup>16</sup> See also Rosenthal 2011: 432-3, Rosenthal 2019, Lau and Rosenthal 2011.

<sup>17</sup> Dretske later wrestled with these kinds of cases (Dretske 2006); see Rosenthal 2005b: 181.

<sup>18</sup> That is, the representations might be at the “sub-personal” level, in one sense of that ambiguous term (Drayson 2014).

<sup>19</sup> On blindsight, see also Phillips 2020.

<sup>20</sup> See Block 2001: 203.

To introduce the second objection, consider the following passage, where Rosenthal seems to assume that some sort of higher-order theory of consciousness is motivated by the initial explanation of the subject-matter:

A state's being conscious is a matter of mental appearance—of how one's mental life appears to one. If somebody is in a mental state but doesn't seem subjectively to be in that state, the state is not conscious. (Rosenthal 2011: 431)<sup>21</sup>

The obvious reply is that there is another less-loaded and more standard way of explaining state consciousness, which was deployed in the previous section. Namely, a state is conscious just in case there's "something it's like" to be in it. Philosophers of mind typically find this at least mildly illuminating, and Rosenthal is no exception.<sup>22</sup> This explanation seems to leave it entirely open whether consciousness is "a matter...of how one's mental life appears to one".

Not quite, according to Rosenthal. When what is implicit in 'there's something it's like' is made explicit, namely the phrase 'for so-and-so', we see that the higher-order theory is implied:

As Nagel stressed in the article that launched that phrase, what it's like to have an experience is what it's like *for* the individual that has the experience. When a person enjoys the taste of wine, thereby enjoying gustatory phenomenality, there is something it's like *for that person* to experience the taste of the wine.

---

<sup>21</sup> Another example: "The basic motivation [for higher-order theories] is that no mental state is conscious if one is wholly unaware of being in such a mental state" (Lau and Rosenthal 2011: 365).

<sup>22</sup> He does think "what it's like" talk can lead to confusion, though (Rosenthal 2005a: 1). For the view that this locution is best avoided altogether, see Lycan 1996: 77 and Snowdon 2010.

Not so in cases of visual extinction; there is nothing it's like *for* an extinction subject to have a qualitative experience of the extinguished stimuli. That's why seeing visual extinction as the having of phenomenality without one's knowing it does not fit comfortably with the explanation of phenomenality in terms of what it's like to have an experience. (Rosenthal 2002: 656).

Visual extinction is caused by parietal lobe damage, and patients fail to report stimuli on the opposite side to the lesion, when stimuli are presented to both sides. Rosenthal is suggesting that the extinction patient is enjoying a "qualitative experience" of the contralesional stimuli which is not conscious because there is nothing it's like *for* the patient. And, properly understood, 'there's something it's like *for* so-and-so' has the fingerprints of the higher-order theory all over it:

As many, myself included, use that phrase, there being something it's like for one to be in a state is simply its seeming subjectively that one is in that state. (Rosenthal 2011: 433)

Weisberg puts the idea particularly clearly:

On one reading [of 'there's something it's like *for* the organism'], which we can...term the moderate reading, the 'for' stressed by Nagel is crucial: the notion indicates a subjective awareness of an organism's mental states by the organism itself. (Weisberg 2011a: 439)<sup>23</sup>

As against this, consider this (updated) example from Snowdon: 'What will it be like for the UK to leave the European Union?' (2010: 24). The question is not about whether the UK is aware of anything. A plausible paraphrase is: 'How will (the UK's) leaving the European Union affect the

---

<sup>23</sup> See also Weisberg 2011b: 411, quoted in Stoljar 2016: 1193.

UK?’—the ‘for’-phrase just indicates the thing affected. Similarly, ‘What will it be like for Josh to leave the party?’ can be paraphrased as ‘How will (Josh’s) leaving the party affect Josh?’ In one context, we might be interested in the financial effects on Josh (maybe he’ll miss out on the grand prize by leaving). In more typical contexts we will be interested in the psychological effects—will leaving make him sad or relieved? And in Nagelian contexts ‘What is it like for David to see red?’ asks for the *experiential* effects seeing red has on David. Those effects may or may not include David’s mental life appearing to him. But—as Stoljar puts it—“there is nothing in Nagel’s definition of consciousness to suggest that a reflexive [higher-order] account is correct” (Stoljar 2016: 1195).<sup>24</sup>

\* \* \*

The theoretical edifices erected by philosophers usually crumble to the ground on closer examination. Even those erected by the best philosophers almost always have superfluous pinnacles and gargoyles. But underneath these distracting adornments is something enduring and valuable. Such is the case with Rosenthal’s theory of mental qualities.<sup>25</sup>

---

<sup>24</sup> The points in this paragraph are distilled from Stoljar 2016. For an earlier and less persuasive attempt at disconnecting ‘for so-and-so’ and higher-order theories, see Byrne 2004: 213-6.

<sup>25</sup> Many thanks to an audience at Taipei Medical University, E.J. Green, and especially to David himself.

## References

- Allen, K. 2016. *A Naive Realist Theory of Colour*. Oxford: Oxford University Press.
- Block, N. 1983. Mental pictures and cognitive science. *Philosophical Review* 93: 499-542.
- — — . 1996. Mental paint and mental latex. *Philosophical Issues* 7: 19-49.
- — — . 2001. Paradox and cross purposes in recent work on consciousness. *Cognition* 79: 197-219.
- Block, N., and I. Phillips. 2016. Debate on unconscious perception. *Current Controversies in Philosophy of Perception*, ed. B. Nanay. New York: Routledge.
- Byrne, A. 2004. What phenomenal consciousness is like. *Higher-Order Theories of Consciousness: An Anthology*, ed. R. J. Gennaro. John Benjamins.
- — — . 2018. *Transparency and Self-Knowledge*. Oxford: Oxford University Press.
- Byrne, A., and D. R. Hilbert. 2020. Objectivist reductionism. *The Routledge Handbook of Philosophy of Colour*, ed. F. Macpherson and D. Brown. Routledge.
- Drayson, Z. 2014. The personal/subpersonal distinction. *Philosophy Compass* 9: 338-46.
- Dretske, F. 1993. Conscious experience. *Mind* 102: 263-83.
- — — . 1995. *Naturalizing the Mind*. Cambridge, MA: MIT Press.
- — — . 2006. Perception without awareness. *Perceptual Experience*, ed. T. S. Gendler and J. Hawthorne. Oxford: Oxford University Press.
- Jackson, F. 1977. *Perception: a representative theory*. Cambridge: Cambridge University Press.
- Kind, A. 2003. What's so transparent about transparency. *Philosophical Studies* 115: 225-44.
- Lau, H., and D. M. Rosenthal. 2011. Empirical support for higher-order theories of conscious awareness. *Trends in Cognitive Sciences* 15: 365-73.
- Lycan, W. G. 1996. *Consciousness and Experience*. Cambridge, MA: MIT Press.
- Moore, G. E. 1939. Proof of an external world. *Proceedings of the British Academy* 25: 273-300. Page reference to the reprint in Moore 1993.
- — — . 1993. *G. E. Moore: Selected Writings*: Routledge.
- Nagel, T. 1974. What is it like to be a bat? *Philosophical Review* 83: 435-50.
- Phillips, I. 2013. Afterimages and sensation. *Philosophy and Phenomenological Research* 87: 417-53.
- — — . 2018. Unconscious perception reconsidered. *Analytic Philosophy* 59: 471-514.
- — — . 2020. Blindsight is qualitatively degraded conscious vision. *Psychological Review*.
- Rosenthal, D. M. 1986. Two concepts of consciousness. *Philosophical Studies* 49: 329-59. Page reference to the reprint in Rosenthal 2005a.
- — — . 1994. First-person operationalism and mental taxonomy. *Philosophical Topics* 22: 319-49. Page reference to the reprint in Rosenthal 2005a.
- — — . 1999a. The colors and shapes of visual experiences. *Consciousness and Intentionality: Models and Modalities of Attribution*, ed. D. Fissette. Kluwer Academic Publishers.
- — — . 1999b. Sensory quality and the relocation story. *Philosophical Topics* 26: 321-50. Page reference to the reprint in Rosenthal 2005a.
- — — . 2000. Introspection and self-interpretation. *Philosophical Topics* 28: 201-33. Page reference to the reprint in Rosenthal 2005a.
- — — . 2001. Color, mental location, and the visual field. *Consciousness and Cognition* 10: 85-93.
- — — . 2002. How many kinds of consciousness? *Consciousness and Cognition* 11: 653-65.

- — —. 2005a. *Consciousness and Mind*. Oxford: Oxford University Press.
- — —. 2005b. Sensory qualities, consciousness, and perception. *Consciousness and Mind*. Oxford: Oxford University Press.
- — —. 2010. How to think about mental qualities. *Philosophical Issues* 20: 368-93.
- — —. 2011. Exaggerated reports: reply to Block. *Analysis* 71: 431-7.
- — —. 2019. Consciousness and confidence. *Neuropsychologia* 128: 255-65.
- Smart, J. J. C. 1959. Sensations and brain processes. *Philosophical Review* 68: 141-56.
- Snowdon, P. 2010. On the what-it-is-like-ness of experience. *Southern Journal of Philosophy* 48: 8-27.
- Stoljar, D. 2016. The semantics of ‘what it’s like’ and the nature of consciousness. *Mind* 125: 1161-98.
- Weisberg, J. 2011a. Abusing the notion of what-it’s-like-ness: A response to Block. *Analysis* 71: 438-43.
- — —. 2011b. Misrepresenting consciousness. *Philosophical Studies* 154: 409 - 33.