

## What Should the Naïve Realist say about Total Hallucinations?

Heather Logue

Naïve Realism is a theory of *veridical* experience—the sort of experience in which a subject perceives things, and they appear to the subject to have certain properties because the subject *perceives* those properties. For example, I'm currently having a veridical visual experience of the banana on my desk. I see the banana, and it looks to me to be yellow and crescent-shaped because I perceive the banana's yellowness and crescent-shapedness. What Naïve Realism says about veridical experience is that it *fundamentally consists in the subject perceiving things in her environment and some of their properties*. For example, it says that my veridical experience of the banana fundamentally consists in my perceiving it, and its yellowness and crescent-shapedness.

Whatever there is to be said in favor of Naïve Realism, many reject the view because it comes with a commitment to a rather counterintuitive account of *total hallucination*—the sort of experience in which the subject doesn't perceive anything in her environment at all (e.g., an experience had by a brain in a vat). Since total hallucinations *don't* involve the subject perceiving anything in her environment, they can't *fundamentally consist in* perceiving things in her environment. So the Naïve Realist has to give a totally different account of total hallucination.

The received view is that the Naïve Realist must say that total hallucination "...is nothing but a situation which could not be told apart from veridical perception" (Martin 2004: 72). All she can say is that a total hallucination is subjectively indiscriminable from a veridical experience of a certain kind; she cannot characterize it in terms of a more substantive psychological property. Many find this kind of account deeply unsatisfying. Surely we can and *should* say more than this in giving a philosophical theory of total hallucination.

I will argue that the Naïve Realist should say more than this—but also that what she can say is still rather counterintuitive. Nevertheless, I think that this reaction stems from a failure to appreciate the broader dialectical context; whereas the same cannot be said of dissatisfaction with the standard Naïve Realist account of hallucination.

Before I can make the case for that claim, more must be said about exactly what Naïve Realism is and why it is thought to lead to an account of total hallucination in terms of subjective indiscriminability. In the first section, I will elaborate Naïve Realism and sketch the space of options the Naïve Realist has in accounting for hallucination. In the second section, I will elaborate the subjective indiscriminability account, explain why I think it is unacceptable, and why it is commonly supposed that the Naïve Realist is forced to it. In the third section, I will identify a loophole in the reasoning that takes us from Naïve Realism to a subjective indiscriminability account of hallucination, and then outline (but ultimately reject) an attempt to exploit that loophole. In the fourth section, I will argue that we can avoid a subjective indiscriminability account of total hallucination if we reject (the admittedly natural) assumption that there is something it is like to hallucinate. Finally, in the fifth section, I will reply to several objections to this counterintuitive account—in a nutshell, I will argue that *if* a certain motivation for Naïve Realism holds, the objections dissolve.

## 1. Naïve Realism and disjunctivism

As I said above, Naïve Realism holds that veridical experience fundamentally consists in the subject perceiving things in her environment and some of their properties. This might not sound like a controversial or even informative theory; after all, no one but an idealist would deny that the subject of a veridical experience perceives things in her environment and some of their properties. The controversial component of the theory is the claim that veridical experience *fundamentally* consists in such a state of affairs.

What an experience fundamentally consists in are the features of it that provide the ultimate *personal-level psychological* explanations of certain phenomenal, epistemological, and behavioral facts. For example, there's "something it's like" to have the perceptual experience I'm having right now, and in virtue of having it, I'm inclined to believe that there's a banana before me (given that, e.g., I don't believe that there are lots of faux bananas about) and to move my arm in a certain direction (given that, e.g., I want to eat a banana). Of course, there are further *subpersonal* psychological facts concerning the information processing that generates experience, and further *non-psychological* facts concerning the biological and chemical underpinnings of such processing. Such facts are of course explanatorily relevant, but it's not the job of a *philosopher* to identify them. They are *causally* relevant to the features of experience mentioned above, but the aim of a philosophical theory of perceptual experience is *metaphysical* explanation. For example, a given perceptual experience is apt to cause certain beliefs and behaviors, but what a philosopher of perception wants to know is *why* this is so—what it is about that experience that makes it apt to cause *those* beliefs and behaviors as opposed to others.

It will be helpful to compare Naïve Realism to its main rival, *Intentionalism*. There are quite a few varieties of Intentionalism, but the core of the view is the claim that all perceptual experiences fundamentally consist in the subject *representing* her environment as being a certain way. For example, according to an Intentionalist, my current veridical experience fundamentally consists in my visually representing that there's a yellow, crescent shaped thing before me. Of course, an Intentionalist doesn't deny that the subject of a veridical experience perceives things in her environment. She simply denies that veridical experience *fundamentally* consists in this fact—that this fact is the *most basic* personal-level psychological explanation of the phenomenal, epistemological, and behavioral explananda mentioned above.

Discussion of Naïve Realism is usually intertwined with discussion of *disjunctivism*, which is basically the claim that veridical experiences and at least total hallucinations are fundamentally different.<sup>1</sup> The discussions are intertwined because

---

<sup>1</sup> Obviously, this statement of disjunctivism is very rough. For one thing, it doesn't cover illusion—i.e., an experience in which the subject perceives something in her environment, and it appears to have a certain property, but *not* because the subject perceives an instance of that property. Nor does it cover *partial* hallucination—i.e., an experience in which the subject perceives things in her environment, but *some* aspects of the way things perceptually appear to be aren't due to the fact that she perceives these things (but rather to, say, causally independent goings-on in her brain). Since the our focus is what the Naïve Realist should say about *total hallucinations*, we need not

Naïve Realism leads naturally to disjunctivism—as I noted above, since total hallucinations *don't* involve the subject perceiving anything in her environment, they can't *fundamentally consist in* perceiving things in her environment. So the Naïve Realist is pushed towards holding that total hallucinations have a radically different metaphysical structure than veridical experiences.<sup>2</sup>

Of course, a Naïve Realist disjunctivist cannot simply state that total hallucinations are fundamentally different from veridical experiences—she must tell us what total hallucinations fundamentally consist in. There are two broad options: *positive* or *negative* disjunctivism. Positive disjunctivism is the view that hallucination fundamentally consists in a psychological property that can be specified *independently* of veridical experience, for example, the property of *perceptually representing* one's environment as being a certain way. This would be a version of positive disjunctivism: saying that the subject of a hallucination perceptually represents her environment as being a certain way involves neither explicit nor implicit reference to veridical experience.<sup>3</sup> Negative disjunctivism is the view that hallucination fundamentally consists in a psychological property that *cannot* be specified independently of veridical experience, for example, the property of being in a state that is *subjectively indiscriminable from a veridical experience of a certain kind*. Obviously, saying that the subject of a hallucination is in a state that is subjectively indiscriminable from a veridical experience of a certain kind involves explicit reference to a veridical experience of a certain kind. For example, this account holds that a hallucination as of a yellow, crescent-shaped banana fundamentally consists in being in a state that is subjectively indiscriminable from a veridical experience of a yellow, crescent-shaped banana. As far as I know, the only version of negative disjunctivism out there is the one just outlined in terms of subjective indiscriminability. In the next section, I will elaborate this view, highlight its problems, and explain why it is nevertheless the most popular account of hallucination among Naïve Realists.

---

trouble ourselves with the related refinements of disjunctivism here. (Unless I say otherwise, in what follows when I use 'hallucination' I mean *total* hallucination.) That said, there are tricky issues regarding the implications of the account of total hallucination I will propose for what the Naïve Realist should say about illusions and partial hallucinations. Unfortunately, I must leave this issue until another time.

<sup>2</sup> Although see Johnston 2004 for a non-disjunctivist theory of perceptual experience that is compatible with Naïve Realism. However, this theory comes with commitments that most Naïve Realists aren't willing to take on (see fn. 15 below), which is why most Naïve Realists are disjunctivists.

<sup>3</sup> Throughout this paper, I will assume that the positive psychological property is a perceptual representational state, because I think that sort of state is the most plausible candidate for being what total hallucination fundamentally consists in. However, in principle, it could be a positive psychological property invoked by theories of perceptual experience *other* than Intentionalism (e.g., awareness of sense-data).

## 2. Subjective indiscriminability negative disjunctivism

In order to understand subjective indiscriminability negative disjunctivism (hereafter ‘SIND’ for short), we must clarify exactly what it means to say that a hallucination is *subjectively indiscriminable* from a veridical experience of a certain kind. For M.G.F. Martin, the originator of SIND, the relevant notion is a purely *epistemic* one: a hallucination is subjectively indiscriminable from a veridical experience of a certain kind just in case the subject is *not* in a position to *know by introspection alone* that the hallucination is *not* a veridical experience of that kind (2004: 48). (‘Introspection’ refers to the special mode of access one has to one’s own mental states, whatever that is, exactly.)

There are two different ways of interpreting such claims, which differ with respect to what the relata are. The first is a *de re* claim about two particular experiences: there is a particular hallucination that the subject cannot tell apart from a particular veridical experience by introspection alone. The second interpretation is a claim about a particular hallucination and the *property* of being a veridical experience of a certain kind: there is a particular hallucination such that the subject cannot tell that it *doesn’t have the property of being* a veridical experience of that kind (e.g., the property of being a veridical experience of a yellow, crescent-shaped thing). On the first interpretation, the relation is symmetric, but on the second it isn’t—the latter concerns a relationship between entities in different ontological categories (an experience, and a *property* of experiences).

Now, one might want to deny that any given veridical experience is subjectively indiscriminable from a particular hallucination (e.g., for anti-skeptical reasons). Since the *de re* interpretation is *symmetric*, this requires denying that any given hallucination is subjectively indiscriminable from a particular veridical experience. But one could still accept that a hallucination is subjectively indiscriminable from a certain kind of veridical experience on the *non-symmetric* interpretation of this claim: in principle, it’s possible that one *couldn’t* be in a position to know that a given hallucination doesn’t have the property of being a veridical experience of a yellow, crescent-shaped thing, even if one can be in a position to know that a given *veridical experience* doesn’t have the property of being a *hallucination* as of a yellow, crescent-shaped thing. Since SIND can be elaborated in terms of the uncontroversial non-symmetric interpretation, let us set the symmetric interpretation aside for simplicity’s sake.<sup>4</sup>

So, according to Martin’s version of SIND, total hallucination fundamentally consists in being a state that is subjectively indiscriminable from a veridical experience of a certain kind, where this is a matter of the subject not being able to know by introspection alone that the state lacks the property of being a veridical experience of a that kind. This view faces some difficult objections. For example, one that has loomed particularly large concerns cognitively unsophisticated creatures—creatures that don’t have the capacity to form beliefs, and hence are *never* in a position to know anything at all (Siegel 2008: 210-4) Since Martin’s version of SIND utilizes a purely epistemic notion of indiscriminability, *prima facie*, it cannot adequately account for the hallucinations of such creatures. For such a creature, any given hallucination is subjectively indiscriminable from *every* kind of veridical experience, and so we cannot even

---

<sup>4</sup> See Fish 2009: 86-8 for a more detailed discussion of these interpretations.

formulate a SIND account of its hallucinations (as there isn't a specific kind of veridical experience from which it is subjectively indiscriminable).<sup>5</sup>

In order to circumvent objections in this vicinity, William Fish proposes a version of SIND that utilizes a *broader* notion of indiscriminability. On his view, what it is for a hallucination to be subjectively indiscriminable from a veridical experience of a certain kind is for it to produce the same *cognitive effects* as the latter would in a rational subject, holding fixed all the subject's other mental states (2009: 94-5). The cognitive effects Fish has in mind include beliefs, but they also include *actions*. This is how he proposes to deal with the problem of cognitively unsophisticated hallucinators—holding other mental states fixed, for any given hallucination, there is a specific kind of veridical experience that would generate precisely the same behaviors (2009: 101-2). So, on this broader notion of subjective indiscriminability, there is a reasonably specific kind of veridical experience from which a cognitively unsophisticated creature's hallucination is subjectively indiscriminable.

Siegel then offers objections to Fish's version of SIND (see her 2008: 214-7). But rather than explore the existing dialectic further, let us take a step back to consider what is arguably the most important problem with any version of SIND. In a nutshell, the problem is this: any version of SIND takes subjective indiscriminability *for granted*, rather than explaining it in terms of personal-level psychological facts.

Martin's version of SIND cannot explain *why* a hallucination is subjectively indiscriminable from a veridical experience of a certain kind in the relevant sense—that is, it cannot explain why a hallucinating subject isn't in a position to know by introspection alone that her experience doesn't have the property of being a veridical experience of a certain kind. According to this version of SIND, a hallucination *fundamentally* consists in being subjectively indiscriminable from a veridical experience of a certain kind. So it's *built in* to the view that there is no personal-level psychological explanation of the fact that a hallucination is subjectively indiscriminable from a veridical perception of a certain kind. I suspect that much of the resistance to Martin's version of SIND is rooted in the idea that it's part of the job of a philosophical theory of perceptual experience to provide such explanations. Now, Martin would simply reject this part of the job description. But arguably, it's a crucial component of explaining the epistemological role of perceptual experience. Perceptual experience is the source of perceptual knowledge about one's environment, but it is also an *object of introspective* knowledge. So, just as we need to explain what it is about a given perceptual experience that suits it to generate and justify specific beliefs about one's environment, we also need to explain what it is about a given hallucination that renders introspection unable to tell it apart from a veridical experience of a certain kind.<sup>6</sup>

---

<sup>5</sup> Martin replies to this objection in his 2004: 51 and his 2006: 381, and Siegel rebuts this reply in her 2008: 211-3.

<sup>6</sup> In principle, we could explain this inability in our theory of *introspection* of perceptual experience, rather than in our theory of perceptual experience. The idea is that hallucination seems exactly like a veridical perception of a certain kind to introspection (even if they have no personal-level psychological commonalities whatsoever) *simply* because introspection is not sensitive to the differences. But as I will argue below in section 4, this kind of explanation is an option of last resort.

Similarly, Fish's version of SIND cannot explain *why* a hallucination is subjectively indiscriminable from a veridical experience of a certain kind in the relevant sense—that is, it cannot explain why the former produces the same cognitive effects as the latter would (in a rational subject, holding fixed all the subject's other mental states). According to this version of SIND, a hallucination fundamentally consists in being a state that generates the same beliefs and behaviors as a veridical experience of a certain kind would (with the aforementioned qualifications). So it's *built in* to the view that there is no personal-level psychological explanation of the fact that a given hallucination generates the beliefs and behaviors that it does. This clearly runs afoul of part of the job description of a philosophical theory of perceptual experience outlined above—namely, the requirement of specifying the personal-level psychological features of a hallucination that render it apt to cause certain beliefs and behaviors. Now, Fish rejects this part of the job description: basically, he says that if his theory of hallucination is correct, then this explanatory burden isn't part of the job description after all (2009: 114). However, this attempt to evade the explanatory burden is dialectically ineffective. Fish's opponent *accepts* this conditional—she's simply using it in a *modus tollens* inference to the falsity of Fish's account. She will not reject the job description without good reason for doing so (one that's *independent* of the details of any particular theory).<sup>7</sup>

In short, the basic problem with SIND is that on any version of it, the relevant sort of subjective indiscriminability is just left *unexplained*.<sup>8</sup> Nevertheless, it is commonly supposed that the Naïve Realist simply has no other option in accounting for hallucination. This is because of Martin's influential "screening-off" argument against positive disjunctivism (2004: 52-68).<sup>9</sup> The first premise of the argument is this: if we say that hallucination fundamentally consists in (say) the subject perceptually representing her environment as being a certain way, then we have to say that veridical experience does too. Martin's argument for this claim is that a hallucination can have the same proximal neural cause as a veridical experience. So if that proximal neural cause gives rise to a representational state in the case of hallucination, then it gives rise the same kind of representational state in the case of a veridical experience—it would be *ad hoc* to deny that it does.<sup>10</sup> The second premise of the argument is this: if veridical experience also fundamentally consists in being in a representational state, then this state "screens off" the Naïve Realist account from doing any significant explanatory

---

<sup>7</sup> See Logue 2010: 32-4 for a more detailed discussion of this objection to Fish's account, his response, and why this response isn't dialectically effective.

<sup>8</sup> It's not as though SIND has *no* explanation of subjective indiscriminability at its disposal. A proponent of SIND might appeal to *neural* commonalities in explaining it (Thau 2004: 249-50). But what we're after in giving a philosophical theory of perceptual experience is *personal-level psychological* explanation.

<sup>9</sup> Martin presents this argument in painstaking detail, but for our purposes here it's sufficient to summarize the key moves in it.

<sup>10</sup> One might object to this case for the first premise because of skepticism about there being a "last" brain state that then causes [perceiving] (Johnston 2004: 138-9). But as long as the sort of brain state that Martin has in mind is a *constituent* of both veridical and hallucinatory experience, and it either causes or constitutes a perceptual representational state, the first premise of the screening-off argument holds.

work. The idea is that whatever the experiences have in common (e.g., their phenomenal character, and their tendency to generate certain beliefs and behaviors) will be best explained by *further* features they have in common (e.g., their representational properties). Hence, there wouldn't be much left over for *Naïve Realism* to explain; we could adequately account for the phenomenal, epistemological, and behavioral features of veridical experience entirely in terms of perceptual representation. Hence, if hallucination fundamentally consists in perceptual representation, then Naïve Realism is explanatorily redundant. So if we want to hang on to Naïve Realism, we have to endorse a version of SIND.<sup>11</sup>

We have now arrived at the present state of the dialectic concerning how the Naïve Realist should account for hallucination. Whatever Naïve Realism has going for it as an account of *veridical* experience, the screening-off argument seems to show that the view brings a commitment to an account of hallucination that simply cannot explain one of the main things it's supposed to explain (*viz.*, subjective indiscriminability). Fortunately, as I will explain in the next section, there is a loophole in the screening-off argument. Unfortunately, as I will also explain, it's not clear how the Naïve Realist can exploit it.

### 3. Positive disjunctivism, first pass

As I have argued elsewhere (Logue forthcoming-a: section 4.3) the loophole in the screening-off argument is this: even if the positive disjunctivist is forced to conclude that veridical experience involves perceptual representation, she is *not* forced to conclude that veridical experience *fundamentally consists* in perceptual representation. For example, there is logical space for the following sort of Naïve Realist positive disjunctivist package. Hallucination fundamentally consists in the subject representing her environment as being a certain way—that is, the most basic, personal-level psychological explanation of the phenomenal, epistemological and behavioral features of a hallucination is in terms of such a representational state. As for veridical experience, the Naïve Realist can accept that if a hallucination involves a certain kind of representational state, then so does any veridical experience that involves the same kinds of brain states; and she can accept that a veridical experience's phenomenal, epistemological, and behavioral features are explained in terms of the subject being in a certain representational state. However, the Naïve Realist can *deny* that a veridical experience *fundamentally consists* in the subject representing her environment as being a certain way. That is, she can insist that there is a *further* personal-level psychological fact in virtue of which the subject is in this representational state—a natural candidate would be *the subject perceiving things in her environment* (I'll call this the 'Naïve Realist state of affairs' for short). This is the Naïve Realism the positive disjunctivist aims to preserve: a view on which veridical experience fundamentally consists in the subject perceiving things in her environment.

This Naïve Realist positive disjunctivist package is short on detail. In particular, I haven't specified *what it is* about the representational state that is explained by the

---

<sup>11</sup> Note that an analogous argument applies to any other positive account of hallucination.

Naïve Realist state of affairs. (What its content is? The mere fact that the subject is in it?) And unfortunately, it's not clear how to make the proposal more precise. I suspect that this is because the view mischaracterizes the relationship between the representational state and the Naïve Realist state of affairs. It's not that the Naïve Realist state of affairs *explains* anything about the representational state. It's rather that the representational state is simply a *constituent* of the Naïve Realist state of affairs.

Note that the Naïve Realist state of affairs has *the subject* as a constituent (in addition to the objects of experience and the perceptual relation the subject bears to them). This raises the question: *which aspects* of the subject count as part of the veridical experience? Plausibly, not *all* of them do. Some of them are causal *consequences* of the experience (e.g., the belief that there's a yellow, crescent-shaped thing before her). Some of them operate in *conjunction* with the experience to produce such consequences (e.g., the desire to eat a banana operates in conjunction with the veridical experience to produce the action of reaching for the banana). And some of them are just plain *irrelevant* to the experience (e.g., the shirt she's wearing). But the Naïve Realist's account of veridical experience can and should include states of the subject that have a role to play in explaining the phenomenal, epistemological, and behavioral aspects of the experience—such as personal-level perceptual representational states (if such there be). Hence, given that veridical experience involves a perceptual representational state, it is best thought of as a constituent of the state of affairs that the Naïve Realist identifies with the experience.

So, in summary, it seems that the initial proposal started off on the wrong foot. The representational state involved in veridical experience isn't something entirely distinct from the Naïve Realist state of affairs to be explained in terms of it. Rather, it is simply a constituent of that state of affairs—one that can be appealed to in explaining various aspects of the veridical experience. And the Naïve Realist can say that the representational state involved in hallucination isn't embedded within a Naïve Realist state of affairs. The subject perceptually represents her environment as being a certain way without perceiving anything in it, and this representational state can be appealed to in explaining various aspects of the hallucination.

However, this change of heart has pushed us back into the crosshairs of the screening-off argument. The initial proposal avoided the screening-off argument by claiming that some aspect of the representational state was to be *explained in terms of* the Naïve Realist state of affairs—which would mean that the latter isn't explanatorily redundant after all. But now I've suggested that we *abandon* the idea that we can explain the representational state in terms of the Naïve Realist state of affairs, and instead hold that the former is a constituent of the latter—one which has some role to play in explaining the experiential phenomena at issue. By the reasoning of the screening-off argument, if a perceptual representational state suffices to explain the phenomenal, epistemological, and behavioral explananda in the case of hallucination, then it *also* suffices to account for those explananda when they occur in the case of veridical perception. So if we give up on the idea that the representational state is explained in terms of the Naïve Realist state of affairs, it appears that the screening-off argument goes through. There would be no reason to identify the veridical experience with the entire Naïve Realist state of affairs; we might as well just identify it with the constituent representational state (since it explains everything that needs explaining).



The only way out for the positive disjunctivist is to deny that the representational state explains *everything* that needs explaining. For positive disjunctivism to thwart the screening-off argument, there must be some feature of veridical experience that is explained by the Naïve Realist state of affairs *as a whole*, rather than just its component representational state. But it's not immediately clear how to defend such a claim. It is typically assumed that a hallucination has pretty much exactly the same phenomenal, epistemological, and behavioral features as a veridical perception of a certain kind (with a few minor exceptions to be noted later), and a perceptual representational state is sufficient to explain these common features. However, in the next section, I will present and defend a version of positive disjunctivism that *denies* that a hallucination has the same features as a veridical experience of a certain kind.

#### 4. Eliminativist positive disjunctivism

What might veridical experience have that hallucination lacks? Of course, a hallucination and the kind of veridical experience from which it's subjectively indiscriminable differ in their causal histories—veridical experience is caused in a “non-deviant” manner, whereas hallucination is “deviantly” caused. Spelling out what this notion of deviant causation of an experience amounts to is a non-trivial task. But *whatever* it amounts to, the difference isn't one the positive disjunctivist can appeal to in order to avoid the screening-off argument. For we don't need an *explanation* of why veridical experiences and hallucinations differ in their causal histories; plausibly, this is true simply as a matter of *definition*. Part of what it is to be a hallucination is to be an experience caused in a deviant way. And what we're looking for is an aspect of veridical experience that could be best *explained* by the Naïve Realist state of affairs.

As I've been insisting, the job of a philosophical theory of perceptual experience is to explain its phenomenal, epistemological, and behavioral aspects. So let us consider these aspects to see whether there's anything that veridical experience has that hallucination lacks. The prospects for appealing to doxastic or behavioral aspects of veridical experience are grim, as a hallucination will have pretty much the same doxastic and behavioral features as the kind of veridical experience from which it's subjectively indiscriminable.<sup>12</sup> Just as the subject of a veridical experience of a yellow, crescent-shaped thing is inclined to believe that there is a yellow, crescent-shaped thing before her (other things being equal), so is the subject of a hallucination as of a yellow, crescent-shaped thing. And just as the subject of the veridical experience will reach in a certain direction if she wants to eat a banana, so will the subject of the hallucination.

Now, there are some doxastic and behavioral differences between the subjects—arguably, only the subject of a veridical experience is in a position to form beliefs with object-dependent content (e.g., that *this thing* is yellow and crescent-shaped), and her

---

<sup>12</sup> Of course, a crucial difference in the epistemological roles of veridical experience and hallucination is that the former affords *knowledge* of one's environment. I will not pursue this avenue here, however; as I've argued elsewhere (Logue 2011) that this facet of the epistemological role of veridical experience can be adequately accounted for in terms of perceptual representation.

actions are much more likely to be successful (unless the hallucinator's experience happens to be accurate, she will fail to get her hands on a banana). Nevertheless, these differences can be easily explained in terms of the subjects' representational states. We can say that the subject of the veridical experience perceptually represents the relevant object-dependent proposition, and (arguably) we can account for the success of her action in terms of the truth of that proposition.<sup>13</sup>

What about the *phenomenal* aspects of the veridical experience? On the face of it, the prospects look even *more* grim here—surely what it's like for a subject to have a hallucination as of a yellow, crescent-shaped thing could be *exactly the same* as what it's like to veridically perceive such a thing. But what was once regarded as an obvious truth is now up for debate. In a recent book, William Fish has argued that hallucinations lack perceptual phenomenal character entirely (2009: Ch. 4). Fish is forced to this conclusion because he holds that perceptual phenomenal character is *relational* in nature—i.e., that it fundamentally consists in a relation to a mind-independent entity. In particular, he holds that that it fundamentally consists in the subject *perceiving* things in her environment and some of their properties.<sup>14</sup> *If* perceptual phenomenal character is relational in nature, then total hallucinations don't have it—since they don't involve the subject perceiving things in her environment.<sup>15</sup>

Now, Naïve Realist need not claim that hallucinations lack phenomenal character *entirely* in order for there to be something for the Naïve Realist state of affairs to

---

<sup>13</sup> There is room to criticize the proposed explanation of successful action. As Martin notes, some have argued that “the explanatory potential of...a relational explanans [e.g., perceiving an object] cannot necessarily be matched by a conjunction of a non-relational psychological fact [e.g., perceptually representing that there is something at a certain location relative to one] and some non-psychological relational facts in addition [e.g., the fact that there *is* an object at that subject-relative location]” (2004: 64). Nevertheless, Martin does not pursue this strategy for vindicating the explanatory power of the Naïve Realist state of affairs, on the grounds that it would leave it open that a Naïve Realist explanation of *phenomenal character* is redundant (2004: 64). Since I agree with Martin that such an explanation is *not* redundant (Logue forthcoming-b) I will not pursue this strategy either.

<sup>14</sup> Note that this is only a proposal about *perceptual* phenomenal character—it could still be that the phenomenal character associated with *other* sorts of mental states (e.g., imagination) is non-relational.

<sup>15</sup> Actually, the matter is not quite so straightforward. Mark Johnston holds that the phenomenal character of hallucination fundamentally consists in the subject perceiving “sensible profiles”, which are complexes of properties—in particular, the ones that would be instantiated by things in one's environment if one's experience were veridical (2004: 134). Sensible profiles aren't mind-dependent sense-data; they are complexes of mind-independent, uninstantiated properties. Hence, there is at least logical space for a relational theory of hallucinatory phenomenal character. This theory entails that we can perceive *uninstantiated* (and presumably *abstract*) properties, a claim I think is false (although explaining why would take me too far afield). Suffice it to say that I don't think we should try to avoid the conclusion that total hallucinations lack phenomenal character by embracing Johnston's theory.

explain. For she could claim that hallucinations have *non-perceptual* phenomenal character (perhaps the typically fainter but similar sort of phenomenal character associated with perceptual imagination). This option has the advantage of making a hallucinator's introspective error less egregious (see the objections from introspection in section 5). The error is that she is inclined to believe that her state has perceptual phenomenal character even though it doesn't. But if hallucinations have imagination-like phenomenal character, rather than the subject being inclined to mistake a *non-phenomenal* state for a phenomenal one, she would be inclined to mistake *imagination-like* phenomenal character for perceptual phenomenal character. One might find the latter mistake more palatable than the former. However, this option faces another objection. It entails that *veridical* experiences have too many phenomenal characters—the phenomenal character associated with sensory imagination *in addition to* perceptual phenomenal character. This claim is in tension with the deliverances of introspection of veridical experience. As I'm not entirely sure how to respond to this objection, I will defend the claim that hallucinations lack phenomenal character entirely.

Of course, the obvious question at this juncture in the dialectic is: why on earth should we think that perceptual phenomenal character is *relational* in nature? This question is too big to address within this paper. I refer the reader to other works that address it: Fish argues (briefly yet tantalizingly) that a relational account of phenomenal character can give us traction with one aspect of the "hard problem" of consciousness (2009: 75-9). And I have argued that a relational account of phenomenal character is required to explain how veridical experience puts us in a position to know what the things we perceive are like independently of experience (Logue forthcoming-b). In light of the main aim of this paper, I will assume for the sake of argument that the Naïve Realist has a sound case for the claim that perceptual phenomenal character is relational in nature—for this is the only motivation for denying that total hallucinations have perceptual phenomenal character, and this denial affords a version of positive disjunctivism that avoids the screening off argument.

Let us now spell out this combination of positive disjunctivism with eliminativism about hallucinatory phenomenal character. The basic idea is that total hallucinations fundamentally consist in the subject perceptually representing her environment as being a certain way, but they lack phenomenal character. The first part of the proposal is a positive account of hallucination, and the second part is intended to thwart the screening-off argument. According to the first plank, the most basic personal-level psychological explanation of the epistemological and behavioral effects of a hallucination is in terms of the subject perceptually representing her environment as being a certain way. For example, the subject of a hallucination as of a yellow, crescent-shaped thing is inclined to believe that there is such a thing before her (other things being equal) in virtue of the fact that she *perceptually represents* that there is such a thing before her. And given that the subject believes that bananas are yellow and crescent-shaped and that she wants to eat a banana, she will be inclined to reach in a particular direction. She has this inclination partly in virtue of the fact that she *perceptually represents* that there is a yellow, crescent-shaped thing in that direction. According to the second plank of this positive disjunctivist proposal, hallucinations lack perceptual phenomenal character. That is, there is *nothing it is like* to hallucinate. Note that it doesn't follow that hallucinators are philosophical zombies. For they may be in

other mental states that have phenomenal character, e.g., states of pain, imagination, or memory (Fish 2009: 98-9, fn. 19).<sup>16</sup>

Of course, as I noted above, the subject of a hallucination will be *inclined to believe* that she is in a state with perceptual phenomenal character. Eliminativist positive disjunctivism is committed to claiming that this belief is *false*. Hence, we need an account of why hallucinating subjects are inclined to form false beliefs of this sort. As a first pass, one might suggest that the subject of a hallucination is inclined to believe that her experience has perceptual phenomenal character because it is *subjectively indiscriminable* from a kind of state that *does* have it—namely, a veridical experience of a certain sort. A subject of a hallucination as of a yellow, crescent-shaped thing cannot know by introspection alone that her experience doesn't have the property of being a veridical experience of such a thing, and is thereby inclined to believe that her state has the same phenomenal character as such a veridical experience. But of course the story cannot end here. For we need to explain how a hallucination could be *subjectively indiscriminable* from a certain kind of veridical experience when they differ *so radically*—when there is something it's like to be in the former kind of state, but nothing it's like to be in the latter.

It will be helpful to consider the general forms an explanation of indiscriminability might take. If an X is indiscriminable from the Ys via a given mode of epistemic access, that's because the X *seems* exactly the same as a Y via that mode of epistemic access. And there are three broad kinds of explanations of why the X would seem exactly the same as a Y. The first is that the X *is* exactly the same as a Y. The second is that (a) the X differs from a Y in some but not all respects, but (b) the X is exactly the same as a Y *with respect to those features the relevant mode of access is sensitive to*. The third is that (a) the X differs from a Y in some or all respects, (b) the relevant mode of access *isn't* sensitive to any similarities there may be, and (c) the relevant mode of access is simply insensitive to differences *in any respect whatsoever*.

Obviously, when it comes to explaining the subjective indiscriminability of a hallucination from a certain kind of veridical experience, the first sort of explanation isn't available to an eliminativist positive disjunctivist (since there is a phenomenal *difference*). So the explanation must take either the second or the third form. Now, unless we have reason to believe that the mode of access at issue has the severe limitations postulated in the third form of explanation, arguably this form of explanation should be avoided if at all possible. That is, we shouldn't place all the "blame" for indiscriminability despite differences *on the mode of access*. Provided that the mode of access is supposed to be reasonably acute, we should favor an explanation of the epistemic appearance of total sameness that it generates in terms of its detecting some similarity (rather than an explanation solely in terms of a brute failure to detect

---

<sup>16</sup> I should note that I am not inclined to extend this account to illusions or partial hallucinations. Both sorts of experience involve the subject perceiving things in her environment, so there is scope for claiming that they have relational perceptual phenomenal character. However, formulating the details of such an account will be tricky—for example, what's it's like to perceive something as being yellow when it *isn't* obviously cannot be accounted for in terms of perceiving an instance of yellowness. In the case of illusion, I am inclined to adopt something along this lines of the account offered in Brewer 2008. However, this issue is beyond the scope of this paper.

any differences). And surely, introspection is *at least* reasonably acute with respect to perceptual experience.

So given that the second form of explanation is what we're after, let's make it less abstract by considering an example involving perception rather than introspection—J.L. Austin's lemon-like bar of soap (1962: 50). It is *visually* indiscriminable from lemons of a certain kind (i.e., those with a certain shape, color, and so forth) because it seems exactly the same as a lemon of that kind solely on the basis of vision. And it seems exactly the same as a lemon of that kind because, although it is different in many respects (e.g., chemical composition), it is exactly the same with respect to the properties vision is sensitive to (color, shape, and so forth). Similarly, we can explain the *subjective* indiscriminability of a hallucination from a certain kind of veridical experience as follows: a total hallucination is *subjectively* indiscriminable from a certain kind of veridical experience because it seems exactly the same as a veridical experience of that kind solely on the basis of introspection. And it seems exactly the same as a veridical experience of that kind because it *is* exactly the same with respect to the features introspection is sensitive to.

But which features do they have in common that introspection is sensitive to? At this point, the positive disjunctivist plank of the proposal comes in handy. The positive disjunctivist holds that there is a personal-level psychological commonality between a total hallucination and the kind of veridical experience it is subjectively indiscriminable from—on the view being developed here, this commonality is a perceptual representational state. And plausibly, introspection is sensitive to such states. Now, I'm not suggesting that a subject can tell by introspection that she is in a perceptual representational state. If we could, there would be no room for debate about whether experiences have representational content—a claim which some Naïve Realists dispute (see, e.g., Brewer 2006). Rather, I mean that the *content* of the representational state is introspectively accessible by the subject (e.g., that there is a yellow, crescent-shaped thing before one).

A worrying consequence of combining this explanation of subjective indiscriminability with eliminativism about phenomenal character is that the subject of a total hallucination is *not* introspectively sensitive to the presence or absence of *perceptual phenomenal character*. The proposed explanation holds that the hallucination is exactly the same as veridical experiences of a certain kind with respect to features that introspection is sensitive to. But eliminativism says that they are *not* the same with respect to phenomenal character—veridical experiences have phenomenal character, while the total hallucination doesn't. Hence, in order to reconcile the explanation with eliminativism, we must conclude that the subject of a total hallucination is not introspectively sensitive to whether or not her experience has perceptual phenomenal character; which is a counterintuitive result.

I will address this worry at great length in the next section. But first, let us compare eliminativist positive disjunctivism with the other views we've discussed so far. Unlike SIND and like Intentionalism, eliminativist positive disjunctivism offers a personal-level psychological explanation of subjective indiscriminability. But unlike Intentionalism, eliminativist disjunctivism does not appeal to sameness of phenomenal character in its explanation. Recall that SIND holds that hallucination *fundamentally* consists in being subjectively indiscriminable from a veridical experience of a certain kind, which means that there is no personal-level psychological explanation of the fact

that a given hallucination is subjectively indiscriminable from a veridical experience of a certain kind. Intentionalism, on the other hand, explains subjective indiscriminability in terms of phenomenal sameness, which it in turn explains in terms of sameness in representational state. For example, it explains the fact that a hallucination as of a yellow, crescent-shaped thing is subjectively indiscriminable from a certain kind of veridical experience by appeal to the fact that the experiences have the same phenomenal character. And it explains this sameness in phenomenal character in terms of sameness in representational state (e.g., by holding that the phenomenal character of an experience fundamentally consists in perceptual representation, and holding that both states involve visually representing that there is a yellow, crescent-shaped thing before one). By contrast, we can think of eliminativist positive disjunctivism as offering an explanation in terms of personal-level psychology, but one that cuts out the Intentionalist's middle step: it explains subjective indiscriminability *directly* in terms of sameness of perceptual representational state.

Now that eliminativist positive disjunctivism has been elaborated, let us now address some objections to it.

## 5. Objections and replies

One could object to eliminativist positive disjunctivism by taking issue with either its "eliminativist" plank (i.e., the claim that hallucinations lack perceptual phenomenal character) or its "positive" plank (i.e., the claim that hallucinations fundamentally consist in perceptually representing one's environment as being a certain way). First, I will discuss an objection directed at the "positive" plank, and then I will discuss two objections directed at its "eliminativist" plank.

The objection targeted at the "positive" plank concerns the proposed explanation of a hallucinator's inclination to believe that she's in a state with perceptual phenomenal character. In a nutshell, the worry is that this explanation applies to a *veridically perceiving* subject as well.<sup>17</sup> Recall that eliminativist positive disjunctivism explains a hallucinating subject's inclination to believe that she's having an experience with perceptual phenomenal character by appealing to the fact that her experience is subjectively indiscriminable from a veridical experience of a certain kind (which *does* have perceptual phenomenal character), and in turn explains the subjective indiscriminability in terms of the fact that both experiences involve the same kind of *non-phenomenal* perceptual representational state. Applying the same kind of reasoning used in Martin's "screening off" argument, we reach the following conclusion: if being in a certain kind of non-phenomenal perceptual representational state is sufficient to explain why the subject of a hallucination is inclined to believe that she's having an experience with perceptual phenomenal character, then being in this kind of state is sufficient to explain why the subject of a *veridical experience* is inclined to believe this as well. The fact that a veridically perceiving subject is having an experience with phenomenal character *drops out* of our explanation of why she is inclined to believe that she's having such an experience, which is an odd result. Surely the fact that she is having

---

<sup>17</sup> Thanks to David Chalmers for raising this objection.

an experience with perceptual phenomenal character should play a central role in the explanation of why she *believes* that she does.

However, being in a certain non-phenomenal perceptual representational state plausibly *isn't* sufficient to explain why the subject of a hallucination is inclined to believe that she's having an experience with perceptual phenomenal character. As I argued in section 3, in the case of veridical experience, the perceptual representational state is a constituent of that experience—it is but one part of a larger state of affairs that the Naïve Realist identifies with the experience (*viz.*, the subject perceiving things in her environment and some of their properties). Now, being in such a state presumably *wouldn't* generate the inclination to believe that one is in a state with perceptual phenomenal character if it weren't *normally* a constituent of a state with perceptual phenomenal character—namely, a certain kind of veridical experience. In the absence of a normal association between perceptual representational states and states with perceptual phenomenal character, there would be no reason to suppose that the former would generate inclinations to believe anything whatsoever about the latter.<sup>18</sup> So a hallucinating subject's being in a non-phenomenal perceptual representational state *isn't* sufficient to explain why she is inclined to believe that she's having an experience with perceptual phenomenal character—another crucial part of the explanation is that the representational state is normally a constituent of a state that *does* have perceptual phenomenal character. The explanatory power of the representational state depends upon it normally being a part of a Naïve Realist state of affairs. Hence, the fact that the subject of a *veridical experience* is in a state with perceptual phenomenal character *does* figure in the explanation of her inclination to believe that she's in such a state. Her perceptual representational state generates that inclination because it's normally a constituent of a state with perceptual phenomenal character—in particular, a veridical experience of the sort she's currently having.

Let us now turn to objections to the controversial “eliminativist” plank of eliminativist positive disjunctivism. Probably the most obvious objection to eliminativism is based on the claim that *neural stimulation is sufficient* for the generation of an experience with phenomenal character. The idea is that since such neural stimulation is also sufficient for the generation of a *hallucination*, hallucinations must have phenomenal character. This objection has been discussed at great length by (2009: Ch. 5). In my view, Fish does an excellent job of dismantling it; so my discussion will be relatively brief.

Fish points out that the *brute assertion* that neural stimulation is sufficient for phenomenal character simply begs the question against the Naïve Realist, who is claiming that phenomenal character fundamentally consists in bearing a certain *relation* to things in one's environment. So this assertion stands in need of an argument. One

---

<sup>18</sup> I mean 'normal' in the normative sense. There is something *defective* about hallucination; given that a primary function of a perceptual system is to convey information about the subject's environment to her, the system fails to perform this function in a case of hallucination. So even in a situation in which total hallucinations are the norm in the *statistical* sense (e.g., the Matrix), the perceptual representational states at issue would normally be constituents of veridical experiences (in the *normative* sense).

might seek from it the armchair (by employing thought experiments), or one might appeal to empirical considerations.

As for the first option, one might suggest that it is conceivable that neural stimulation is sufficient for perceptual phenomenal character—e.g., that we can conceive of a brain in a vat being in states with perceptual phenomenal character. From there, one could say that if it's conceivable, then it's possible; and if it's possible, then perceptual phenomenal character isn't relational after all. However, the first step in this conceivability argument is inadmissible in this dialectical context. Given that there is a sound argument for the claim that perceptual phenomenal character is relational (which I'm assuming for the sake of argument here), then it isn't *ideally* conceivable that neural stimulation is sufficient for perceptual phenomenal character. It was *prima facie* conceivable before we were convinced by the argument, but that sort of conceivability is not a reliable guide to possibility (Chalmers 2002). Hence, this conceivability argument gets off the ground only if the Naïve Realist cannot defend her argument for the claim that perceptual phenomenal character is relational—in which case the conceivability argument becomes *unnecessary* (since there would no longer be any *motivation* for claiming that hallucinations lack perceptual phenomenal character in the first place).<sup>19</sup>

The other strategy is to appeal to empirical considerations in support of the claim that neural stimulation is sufficient for perceptual phenomenal character. For example, one might suggest that since there is a strong correlation between certain kinds of neural goings-on and experiences with a certain kind of perceptual phenomenal character, it is plausible that bringing about such neural goings-on is sufficient for bringing about experiences with that kind of perceptual phenomenal character (Fish 2009: 134-6). However, as Fish points out, an equally plausible explanation of the correlation is that the neural goings-on are *necessary* but not *sufficient* for perceptual phenomenal character. In particular, it could be that the visual processing carried out by neurons is a necessary condition of getting into a state with *relational* perceptual phenomenal character—it is but one component in the process that reveals some qualitative aspects of a subject's environment to her, rather than *the* component that generates an experience with (non-relational) perceptual phenomenal character all on its own (Fish 2009: 137-8). What we would need in order to support the sufficiency claim is a uncontroversial case of neural stimulation *without perception* accompanied by perceptual phenomenal character, and Fish argues that there is no such case (Fish 2009: 123-34).

Let us now turn our attention to a family of objections to the eliminativist plank that Fish does not address. The basic worry is that eliminativist disjunctivism

---

<sup>19</sup> I should note that this reply to the conceivability argument isn't Fish's. He concedes that it is conceivable (in a possibility-entailing sense) that neural stimulation is sufficient for perceptual phenomenal character, but says that this shows only that non-relational perceptual phenomenal character is logically possible, not that it's *physically* possible (2009: 122). I don't think the Naïve Realist should be too quick to concede the ideal conceivability of the scenario at issue. For if the argument that perceptual phenomenal character is relational *isn't* based on contingent facts about the physical constitution of our perceptual apparatuses, then it may well be incompatible with even the logical possibility of non-relational perceptual phenomenal character.



characterizes introspection of perceptual experience as being considerably less powerful than it really is. One way of developing this idea is to claim that the belief that one is in a state with perceptual phenomenal character is *infallible*—if a subject believes it, then it's true. But eliminativist disjunctivism entails that such beliefs *aren't* infallible, since it holds that such a belief would be false if it were formed by hallucinating subject.

Why might one think that such beliefs are infallible? One potential reason is the idea that beliefs about “feelings” in general are infallible.<sup>20</sup> A way of getting a grip on the notion of perceptual phenomenal character is by characterizing it as the distinctive kind of *feeling* associated with perceptual experiences (or at least some of them, if eliminativism about hallucinatory phenomenal character is correct). Now, when it comes to other feelings, such as bodily sensations like pains and tickles, it might seem absurd to suppose that one could be wrong about whether one has them. If one believes that one is in pain, then one *is* in pain. So why shouldn't we say the same about feelings in general, and perceptual phenomenal character in particular?

However, this line of thought is too quick. Note that I'm *not* denying the analogous claim about perceptual experience, namely: if one believes that one is having a perceptual experience, then one is having a perceptual experience. Infallibility about whether one is having a *perceptual experience* is in principle compatible with fallibility about whether one's experiences have *perceptual phenomenal character*. I'm denying only that we are infallible about whether our perceptual states have perceptual phenomenal character. So the analogous claim in the case of pain is really this: if one believes that one is in a state that has the sort of phenomenal character associated with pain states (let's call it 'painful phenomenal character'), then one is in such a state.<sup>21</sup>

Even so, one might find the strictly analogous infallibility claim just as plausible—that if one believes that one is in a state with painful phenomenal character, then one must be in such a state. However, matters are not quite so straightforward in this dialectical context. Arguably, if a certain kind of phenomenal character is *relational* in nature, then one is *not* infallible about whether one is in a state that has it. In particular, if a certain kind of phenomenal character consists in a certain kind of *awareness of* an entity of some sort, and the subject can be *mistaken* about whether she is aware of such an entity, then she can be mistaken about whether or not she is in a state with that kind of phenomenal character.

To make this rather abstract point more concrete, let us illustrate it with an example. There is at least logical space for a relational theory of painful phenomenal character, and on such a theory, it would be possible to falsely believe that one is in a state with painful phenomenal character. Suppose (for the sake of argument) that

---

<sup>20</sup> Thanks to Anandi Hattiangadi for raising something along the lines of this objection.

<sup>21</sup> One might suggest that there is no difference between painful phenomenal character and pain itself. Now, the word 'pain' is sometimes used to refer to painful phenomenal character. But it is also used to refer to a mental state that *has* that phenomenal character, along with other features (e.g., a certain functional role). This is the sense in which I am using the word here, because it is analogous to the sense of 'perceptual experience' in play in this context—the latter term is intended to refer to a mental state that might have perceptual phenomenal character, along with other features (e.g., a certain functional role). (The point of the 'might' is to allow for eliminativism about hallucinatory phenomenal character.)

painful phenomenal character fundamentally consists in interoceptive awareness of bodily damage. Now, in principle, one can be mistaken about whether one is interoceptively aware of bodily damage. An unwitting amputee waking from surgery might believe that she is interoceptively aware of bodily damage in her left calf, but be wrong (because her whole left leg has just been amputated, and so she has no interoceptive access to it). This possibility, in conjunction with the toy theory of painful phenomenal character just sketched, entails that one can be mistaken about whether one is in a state with painful phenomenal character. If one wrongly believes that one is interoceptively aware of bodily damage, one is nevertheless inclined to believe that one is in a state with painful phenomenal character. But, given the theory, this belief will be *false*: according to the theory, there is no painful phenomenal character without interoceptive awareness of bodily damage. Now, I'm not suggesting that this theory of painful phenomenal character is correct. The moral is simply that there is a connection between infallibility about a mental phenomenon and whether or not it is relational: in particular, *if* a mental phenomenon is relational in nature, *and* if one can be mistaken about whether that relation obtains, then one is not infallible about that mental phenomenon.

This general point has implications for the dialectic concerning eliminativism and infallibility about perceptual phenomenal character. Recall that I am assuming for the sake of argument that perceptual phenomenal character is relational in nature. (That's the whole motivation for claiming that hallucinations lack perceptual phenomenal character in the first place, so there's really no point in taking this claim seriously unless we spot the Naïve Realist the motivation behind it.) So given that we are assuming that perceptual phenomenal character is relational, and the fact that one can be mistaken about whether one perceives things in one's environment, we must also assume in this context that one is *not* infallible about whether one is in a state with perceptual phenomenal character. Hence, the claim that we *are* infallible about this is not admissible in this dialectical context—at least, not as a brute assertion. If the objector can offer an *argument* for this infallibility claim, that argument (in conjunction with the reasoning above) would count as an argument against the claim that perceptual phenomenal character is relational. But, at least as far as I can tell, there is no such argument. Many regard the claim that one is infallible about whether one is in a state with perceptual phenomenal character as simply obvious and intuitive, but such considerations are relatively flimsy compared to a sound argument for the claim that perceptual phenomenal character is relational (which, again, I'm assuming that there is for the purposes of this paper). So, in short, the claim that we are infallible about perceptual phenomenal character is not admissible in this dialectical context as a brute assertion of an allegedly obvious intuition.

Nevertheless, even if we retreat from the infallibility claim, there still seems to be something wrong with the idea that introspection could be susceptible to such a *massive* error. Of course, introspection can get the finer details wrong, especially when we're being hasty or inattentive—for example, I might hastily judge that something looks scarlet to me when it actually looks vermilion. In special cases, introspection can even get it wrong when it comes to something as coarse-grained as *which mental states* one is in—for example, one might deceive oneself into thinking that one is happy even though one isn't. But eliminativism seems to imply that even an attentive, careful subject who isn't at all prone to self-deception is inclined to believe that her experience

has perceptual phenomenal character even though it doesn't. It's hard to believe that introspection can systematically lead one astray with respect to the very coarse-grained question of whether one is in a state with perceptual phenomenal character. As I myself have previously argued:

Introspection might be unreliable when it comes to certain fine-grained mental matters, and even about coarse-grained mental matters when it is employed in unfavorable contexts (e.g., in the presence of motives conducive to self-deception). But surely introspection isn't *so* bad as to systematically generate the belief that one has an experience with phenomenal character when in fact *the state lacks it entirely*. It is plausible that introspection is extremely reliable when it comes to the very coarse-grained question of whether one is in a state with phenomenal character, at least when the context is favorable for its operation (e.g., one isn't suffering the delirium typical of dreams). (Logue 2010: 35)

And although the result that introspection is an unreliable guide to perceptual phenomenal character is bad enough in itself, it leads to an even worse conclusion: for one could use it to argue that eliminativist disjunctivism leads to *scepticism* about perceptual phenomenal character. Of course, we want to be able to say that a subject of a *veridical experience* can know that she is in a state with perceptual phenomenal character by introspection. Now, plausibly, for introspection to afford knowledge that one is in a state with perceptual phenomenal character, it is at least a necessary condition that introspection is a *reliable* guide to whether one is in such a state. But, as we've seen, eliminativism about hallucinatory phenomenal character entails that introspection is *not* a reliable guide to the instantiation of perceptual phenomenal character. So it appears that eliminativism leaves us in a sceptical predicament: given that it entails that introspection is an unreliable guide to the instantiation of perceptual phenomenal character, it also entails that we cannot know by introspection that it *is* instantiated.

I think the eliminativist disjunctivist should respond to these objections by arguing that any situation in which a subject is having a total hallucination is an *unfavorable* context for introspection of perceptual experience. If this is right, then we would be able to soften the blow of the result that introspection isn't a generally reliable guide to the instantiation of perceptual phenomenal character. For we would be able to maintain that it is at least reliable in *favorable* contexts, which in turn affords a response to the skeptical worry outlined above—the unreliability of introspection in *unfavorable* contexts does not impugn its ability to generate knowledge about experience in *favorable* contexts. Compare: the unreliability of *perceptual experience* in unfavorable contexts (e.g., in unusual lighting conditions, from great distances, and so forth) does not impugn its ability to generate knowledge about *things in one's environment* in favorable contexts.<sup>22</sup>

---

<sup>22</sup> Of course, there are *other* well-known arguments for the claim that perceptual experience cannot afford knowledge about one's surroundings (e.g., arguments from the inability to rule out skeptical scenarios). But the point here is that the *unreliability* of perceptual experience in certain restricted contexts is not a basis for such an argument.

One might think that the suggestion that total hallucination constitutes an unfavorable context for introspection is rather *ad hoc*. Indeed, I used to think so: “Surely it is in principle possible for a hallucinating subject to be exceedingly attentive to her experience, to suffer no distractions from such a careful introspective endeavor, to have no motives conducive to self-deception regarding the phenomenal aspects of her mental situation, etc.” (Logue 2010: 36). I think my past self was too hasty. In particular, I was assuming that we could identify all of the features of a situation that make it an unfavorable context for introspection of perceptual experience *independently* of which theory of introspection we adopt. Now, an ability to be exceedingly attentive to one’s experience, and free of distractions and motives conducive to self-deception, are plausibly *necessary* conditions on a context’s being favorable. But arguably, at least on some theories of introspection of perceptual experience, they are not *sufficient*. In what follows, I will outline a theory of introspection of perceptual experience on which total hallucination constitutes an unfavorable context for introspection of perceptual experience (including its phenomenal character). I will not argue that this theory of introspection is *correct*—such a task is too large to accomplish in this paper. Rather, my aim will be the less ambitious project of arguing that the objections just outlined are not decisive unless this theory can be ruled out.

The broad kind of theory of introspection I have in mind is one on which it involves attending to what one’s mental state is *about*. This sort of theory has been mostly discussed in relation to introspection of *belief*. An inspiration for this sort of theory is Evans’ famous suggestion that “...in making a self-ascription of belief, one’s eyes are, so to speak, or occasionally literally, directed outward—upon the world. If someone asks me ‘Do you think that there’s going to be a third world war?’ I must attend, in answering him, to precisely the same outward phenomena as I would attend to if I were answering the question ‘Will there be a third world war?’” (1982: 225) The idea is basically that introspection of belief doesn’t involve “looking within”, whatever that might amount to. Rather, we acquire knowledge about our beliefs by attending to *extra-mental* facts pertaining to what our beliefs are about (e.g., facts that could precipitate or prevent a third world war). (Of course, on this model, ‘*introspection*’ is not an apt label for the mode of access one has to one’s own beliefs. But I will continue to use it nonetheless.)<sup>23</sup>

Although Evans thought that there are important differences between introspection of belief and introspection of perceptual experience, he thought that they are broadly similar in that the latter also involves “looking without” (1982: 227-8). Plausibly, in making a self-ascription of *perceptual experience*, one’s eyes are *literally* directed outward upon the world. To continue to adapt Evans’ language to the analogous point concerning experience: if someone asks me ‘Do you have a visual experience of something yellow?’, I must attend, in answering her, to precisely the same outward phenomena as I would attend to if I were answering the question ‘Is there something yellow before you?’ The basic idea is that introspection of perceptual experience is mainly a matter of attending to the *objects* of one’s experience. In the phrase ‘objects of experience’, I’m using ‘objects’ broadly to refer to *the entities one*

---

<sup>23</sup> This kind of theory is controversial. But for our purposes, it only matters whether an analogous theory concerning introspection of *perceptual experience* is true; we need not take a stand on the theory concerning introspection of belief.

*perceives*—including not just objects in a narrower sense of the term, but also regions of space, events, property instances, and so forth. So, for example, I know that I'm having a visual experience of a yellow, crescent-shaped thing by attending to my immediate environment: in particular, the thing I see, and its yellowness and crescent-shapedness. This proposal captures how we actually go about introspecting our perceptual experiences. If someone were to ask me about my current visual experience, or instruct me to introspect it, I would have no idea what to do other than attend to what my experience is *of*.

More recently, Alex Byrne (2012) has offered a theory in this spirit. Unlike the theory just sketched, his theory is restricted to *seeing*. That is, it's only a theory of knowledge that one is *visually* perceiving something, as opposed to perceiving something in any of the other sense modalities (although, at least *prima facie*, one could easily extend the theory to the other modalities). Moreover, as stated, Byrne's theory isn't a theory of how one knows about one's *visual experiences* (which don't always involve seeing things, as in the case of a visual total hallucination). Regardless of these differences, there are some initial obstacles that any theory of this sort must overcome, and Byrne elaborates his theory in order to deal with them. I won't get into these obstacles or the details of Byrne's view, as they aren't of immediate importance for our purposes. Here, I just want to argue that if we generalize *something* along the lines of Byrne's theory to introspection of perceptual experience in general, it turns out that total hallucination constitutes an *unfavorable* context for introspection of one's experience.

Of course, this kind of theory cannot be straightforwardly extended to introspection of total hallucination. It holds that one gets knowledge about one's experience by attending to the things in one's environment one perceives, but in the case a total hallucination, one doesn't perceive any such thing. Nevertheless, this kind of theory still seems to capture how the subject of a total hallucination would go about introspecting her experience: plausibly, the subject would have no idea of what to do other than to attend to her surroundings. A natural way of extending this sort of theory to accommodate introspection of total hallucination is as follows: introspection of non-hallucinatory perceptual experience involves attending to things in one's environment (the objects of the experience), whereas introspection of total hallucination involves *trying and failing* to attend to such things.<sup>24</sup>

This proposal allows us to capture how the subject of a total hallucination would actually go about introspecting her experience. And, with a natural addition, it gives us a principled story to tell about what makes total hallucination an *unfavorable* context for introspection of one's experience. Once we've accepted that introspection of perceptual experience involves *trying* to attend to things in one's environment, it is natural to suppose that *ideal* introspection involves *succeeding* in this attempt—whereas trying and *failing* amounts to a *defective* form of introspection. Now, if ideal introspection requires successfully attending to the objects of one's experience, then any introspective endeavor carried out by the subject of a total hallucination is *guaranteed* to be defective (since there are no objects of experience to attend to). Hence, if

---

<sup>24</sup> Of course, the subject of a total hallucination wouldn't be in a position to *know* that she's failed to attend to things in her environment, given that her experience is subjectively indistinguishable from a veridical experience of a certain kind.

something along the lines of this theory is right, then total hallucination constitutes an unfavorable context for introspection of perceptual experience.

One might worry that this theory entails that the subject of a total hallucination cannot know *anything at all* about her experience via introspection. For if only *ideal* introspection of one's experience can afford knowledge about it, then the subject of a total hallucination cannot have any introspective knowledge about her experience. One option is to embrace the conclusion that the subject of a total hallucination doesn't know anything about her experience. Of course, the subject will have *beliefs* about her experience, but there is room to deny that these beliefs amount to *knowledge*. Another, less radical, option is to deny that ideal introspection is the only route to any knowledge about one's experience whatsoever.

All we need in order to counter the objection to eliminativism is that introspection in an unfavorable context deprives the subject of knowledge about *perceptual phenomenal character*. It could still be the case that defective introspection can put one in a position to know about *other* aspects of one's experience—e.g., that one's experience is as of a yellow, crescent-shaped thing (in other words, what its representational content is). Of course, we would need a principled reason for thinking that defective introspection deprives one of knowledge about perceptual phenomenal character, but doesn't deprive one of knowledge about other features of one's experience. I think that the account of hallucination I'm defending can provide such a reason. Roughly, the idea is that successful attention to the objects of experience is required for knowledge of *relational* features of the experience (e.g., its perceptual phenomenal character), but it is not required for knowledge of *non-relational* features of the experience (e.g., its representational content).<sup>25</sup> Developing the details of this story, however, would take us too far afield.

Let's take stock. I've argued that the objections from introspection can be avoided if total hallucination constitutes an unfavorable context for introspection of experience, and that this is the case if ideal introspection involves successfully attending to the objects of experience. At this point, one might wonder: why should we think that ideal introspection of experience involves attending to its objects? The response to the objections depends upon a controversial claim about how introspection of experience works, and if it's false, the whole house of cards collapses.

However, the situation might not be quite so precarious. What introspection of experience involves might depend on what experience *is*. That is, in principle, one could argue from a theory of the *metaphysical structure* of perceptual experience to a theory of *introspection* of perceptual experience. Recall that, according to Naïve Realism, veridical experiences are *relational* states of affairs that encompass things outside of one's head. If veridical experiences really do encompass things outside one's head, then one might reasonably expect that introspection is at least partially directed *outward* towards those very things (after all, they're part of the experience too). Of course, there

---

<sup>25</sup> Given externalism about mental content, there is a weaker sense in which the content of experience is relational—being in a state with that content requires that (say) one's evolutionary ancestors bore the right sorts of relations to certain things. However, the content of an experience is non-relational in the stronger sense of the term being used in this context: one can perceptually represent that there is something yellow before one without *perceiving* anything yellow.

are many gaps in this line of thought that need to be filled in. In this context, I simply want to suggest that the theory of introspection that the response depends upon *might* be considerably less controversial if Naïve Realism is true. Alas, defending this suggestion will have to wait until another time.<sup>26</sup>

## 6. Conclusion

The standard Naïve Realist account of total hallucinations, subjective indiscriminability negative disjunctivism (SIND), is typically regarded as counterintuitive. I have argued that this reaction is likely rooted in the fact that SIND cannot discharge a key explanatory burden of a philosophical theory of perceptual experience, namely, the task of giving an account of *the personal-level psychological facts in virtue of which* a total hallucination is subjectively indiscriminable from a certain kind of veridical experience. I argued that the Naïve Realist can avoid SIND and explain subjective indiscriminability in terms of perceptual representational states. However, this comes at the cost of embracing what some might take to be an even more counterintuitive claim—the claim that total hallucinations lack perceptual phenomenal character.

Nevertheless, several of the objections underlying this understandable reaction to eliminativism about hallucinatory phenomenal character lose their force once we appreciate the bigger dialectical picture. We must ask: is there any *good reason* to think that total hallucinations lack perceptual phenomenal character? A natural answer for the Naïve Realist to give is that perceptual phenomenal character is *relational* in nature, fundamentally consisting in the subject perceiving things in her environment. This answer undermines several of the objections. Some of the objections beg the question because they are based on claims that presuppose that perceptual phenomenal character is *non-relational* (e.g., the claim that neural stimulation of a certain sort is sufficient for perceptual phenomenal character, and the claim that we are infallible about whether a state has perceptual phenomenal character). And, if perceptual phenomenal character is relational, there is a way to avoid the unwelcome result that introspection is generally unreliable regarding the instantiation of perceptual phenomenal character. If introspection of perceptual experience involves attending to its objects, then it can be argued that total hallucination is an *unfavorable* context for introspection of perceptual experience—which would mean that its unreliability in the case of total hallucination doesn't impugn its reliability in the case of veridical experience. Moreover, if perceptual phenomenal character is relational, that would provide at least some support for a theory of introspection of this sort.

So the structure of the dialectic is this: if perceptual phenomenal character is relational, then the Naïve Realist can dispense with some of the most troubling objections to eliminativist positive disjunctivism. As I said above, the question of

---

<sup>26</sup> Another objection to eliminativism is based on the claim that one can learn what it's like to experience a property by hallucinating it (Johnston 2004: 130-1). If hallucinations lack perceptual phenomenal character, it's not obvious how they could afford such phenomenal knowledge. I don't know what to say in response to this objection at present, other than to bite the bullet and deny that hallucinations are a source of such knowledge.

whether the antecedent is true is beyond the scope of this paper. My point here is simply that the viability of these objections depends upon whether perceptual phenomenal character is relational. Hence, we cannot simply dismiss eliminativist positive disjunctivism on the basis of the former without engaging with the arguments for the latter.<sup>27</sup>

## References

- Austin, J. 1962. *Sense and Sensibilia*. Oxford: Oxford University Press.
- Brewer, B. 2006. Perception and content. *European Journal of Philosophy* 14: 165-81.
- . 2008. How to account for illusion. *Disjunctivism: Perception, Action, Knowledge*, ed. A. Haddock and F. Macpherson. Oxford: Oxford University Press.
- Byrne, A. 2012. Knowing what I see. *Introspection and Consciousness*, ed. D. Smithies and D. Stoljar. Oxford: Oxford University Press.
- Chalmers, D. 2002. Does conceivability entail possibility? *Conceivability and Possibility*, ed. T. S. Gendler and J. Hawthorne. Oxford: Oxford University Press.
- Evans, G. 1982. *The Varieties of Reference*. Oxford: Oxford University Press.
- Fish, W. 2009. *Perception, Hallucination, and Illusion*. Oxford: Oxford University Press.
- Johnston, M. 2004. The obscure object of hallucination. *Philosophical Studies* 120: 113-83.
- Logue, H. 2010. Getting acquainted with Naive Realism: critical notice of Perception, Hallucination, and Illusion. *Philosophical Books* 51: 22-38.
- . 2011. The skeptic and the Naive Realist. *Philosophical Issues* 21: 268-88.
- . forthcoming-a. Good news for the disjunctivist about (one of) the bad cases. *Philosophy and Phenomenological Research*.
- . forthcoming-b. Why Naive Realism? *Proceedings of the Aristotelian Society*.
- Martin, M. G. F. 2004. The limits of self-awareness. *Philosophical Studies* 120: 37-89.
- . 2006. On being alienated. *Perceptual Experience*, ed. T. Gendler and J. Hawthorne. Oxford: Oxford University Press.
- Siegel, S. 2008. The epistemic conception of hallucination. *Disjunctivism: Perception, Action, Knowledge*, ed. A. Haddock and F. Macpherson. Oxford: Oxford University Press.
- Thau, M. 2004. What is disjunctivism? *Philosophical Studies* 120: 193-253.

---

<sup>27</sup> Various versions of this paper were presented at the Centre for Metaphysics and Mind at the University of Leeds, the Philosophical Society at the University of Oxford, and a conference on Phenomenality and Intentionality hosted by the University of Crete. Many thanks to the members of the audiences for their helpful comments and questions.