

# Some thoughts about hallucination, self-representation, and “There it is”

Jeff Speaks

Benj’s “There it is” is a characteristically original and wide-ranging exploration of the relationship between certain direct realist theories of perception and the nature of perceptual justification — with some formal semantics thrown in for good measure.<sup>1</sup> Here I’ll focus on just one of the many topics about which Benj has something to say: his remarks on the topic of the relationship that must obtain between a perceptual state and a belief in order for the former to immediately justify the latter.

Setting some of the subtleties of Benj’s story to the side, and focusing for now only on the case of veridical experience, the basic picture is as follows:<sup>2</sup> being in a veridical perceptual state involves accepting a sentence. This sentence, speaking loosely, represents the subject of the experience as having an experience of the type he is having — so, to use Benj’s example, the sentence accepted in virtue of Sam’s looking at the red color of a widget would represent Sam as having the property of looking at the red color of a widget. The relationship between veridical experiences of this sort and the corresponding sentences accepted has some interesting and unusual features: (i) whenever a sentence of this language which ascribes the property of having a veridical experience of the right sort is accepted, it is true; and (ii) everyone who has a veridical experience of the right sort accepts a sentence which self-ascribes the property of having just that sort of veridical experience. As Benj puts it, the sentences in question are infallible, and the properties they ascribe to subjects are self-intimating.

One might wonder: how could a language have these features? How could it be impossible to accept a sentence of a language without it being true? Benj’s answer is that the language in question is a “Lagadonian language” in which (at least some of) the expressions are objects and properties which refer to themselves. If Sam himself, and the property of looking at the red color of a widget are expressions in this language, and if accepting the sentence which represents Sam as looking at the red color of a widget just is a matter of the name for Sam (namely, Sam himself) instantiating a predicate which expresses the property of looking at the red color of a widget (namely, that property), then we can see how the Lagadonian language could be infallible and self-intimating.

Now, to be sure, this Lagadonian language raises some further questions. Surely Sam can instantiate some properties — like the property of gaining 2

---

<sup>1</sup> A previous draft of this paper was given in response to “There it is” at the 3rd annual Online Consciousness Conference. “There it is” is now published as Hellie (2011).

<sup>2</sup> See Hellie (2011), §3.

pounds — without representing himself as instantiating these properties. So it must be that some of Sam’s properties are predicates of the Lagadonian language, and some aren’t. But what explains this distinction between two sorts of properties? In the standard case, we explain the distinction between things which are and things which aren’t expressions of a language in terms of the use to which the expressions are put by a certain community of language users — whether use is specified in terms of Gricean intentions, Lewisian conventions of truthfulness in trust, or in less psychologistic terms. But in the case of the Lagadonian language, “use” seems to just be a matter of property instantiation — which won’t give us the wanted contrast between the property of looking at the red color of a widget and gaining 2 pounds, since Sam instantiates each.

Now, at this stage Benj is, I take it, just sketching a framework for thinking about these issues rather than giving a fully worked out theory of the Lagadonian language. The present worry is less an objection to the framework itself than a question which, it seems to me, a fuller development of Benj’s theory should be able to answer.<sup>3</sup> So let’s set worries about the Lagadonian language to the side and press on to the account of perceptual justification.

To explain how accepting a sentence of the Lagadonian language can rationalize a belief, Benj suggests that we have to understand how sentences of this language might be related to sentences of the distinct language which does underwrite beliefs. The mechanism for this is the subject’s regarding a sentence of the Lagadonian language as equivalent to a sentence of the belief language. Roughly, if  $\mathcal{R}$  is a sentence of the Lagadonian language and  $B$  is a sentence of the belief language, this requires that the subject have a certain cluster of attitudes toward the biconditional  $\lrcorner \mathcal{R} \text{ iff } B \lrcorner$  — one must regard it as trivially true and its negation as incoherent, and one must take questions about why it is true to be unintelligible. When these conditions are satisfied, this is sufficient for  $B$  to have the same content as  $\mathcal{R}$ , which in turn is sufficient for (to continue with the example above) Sam to believe that Sam is looking at the red color of a widget. Since sentences of the Lagadonian language are infallible, a belief formed in this way will always be true.<sup>4</sup>

---

<sup>3</sup> Some initially plausible answers won’t work. For example, one might try to draw the distinction in terms of availability of the relevant properties for reasoning; the proposition that I am looking at the red color of a widget is immediately available to affect my beliefs and actions, whereas the proposition that I have gained two pounds might not be. But of course the proposition that I am looking at the red color of a widget might be similarly unavailable, as can be seen from cases in which I’m unsure whether I’m having a veridical or hallucinatory experience.

<sup>4</sup> See Hellie (2011), 138 and following. Strictly, what follows is just that the belief is true at the moment at which it is formed, presuming that this moment is the same as that at which the relevant experiential property is instantiated by the subject. The belief might quickly be falsified by a change in the veridicality of the subject’s experience. (Or, if we think of beliefs as having their truth-values eternally, ordinary mechanisms of ‘belief maintenance’ might quickly lead to a false belief if there’s a change in the veridicality of the subject’s experience.)

Again, one might raise some questions here about how the central terms of the theory are to be understood. It is fairly clear what it means to regard two sentences of a language like English as equivalent, in Benj's sense; but it's not quite as clear when one of the sentences is, like  $\mathcal{R}$ , a subject instantiating a certain property. As far as I can tell, I have never had any attitudes at all toward a biconditional one of whose constituent sentences is my instantiating the property of looking at something red, mainly because it never crossed my mind that my instantiating such a property could even be a sentence in a biconditional. But, if Benj's theory is to explain the rational status of my beliefs about me looking at red things, this must be something which I've done many times — and it is very puzzling how I could have adopted the cluster of attitudes described in the preceding paragraph toward the relevant biconditionals without noticing. This suggests, I think, that we need something more than the suggested interpretation of 'regarding  $\mathcal{R}$  and  $B$  as equivalent' if it is to do the work assigned to it by Benj's theory.

There are also some worries about the view of the individuation of contents implied by this story, according to which two sentences have the same content for a speaker if the speaker (in the sense sketched above) regards the two sentences as equivalent. This makes certain sorts of mistakes about equivalence impossible: if someone regards a biconditional as trivially true, its negation as incoherent, and its truth as inexplicable, it follows that that person is correct. This is in a way parallel to a familiar consequence of other coarse-grained views of content, like the view that propositions are sets of worlds, which entails that no one believes any necessary falsehoods. And, it seems to me, one might object to Benj's theory using just the same sorts of examples standardly used to argue against those coarse-grained views of content, like mathematical mistakes. Suppose that a mathematician regards a pair of formulae as equivalent, the negation of their biconditional as incoherent, and their equivalence as inexplicable (perhaps the mathematician thinks that all mathematical truths are inexplicable) — does this really entail that the formulae are synonymous out of that mathematician's mouth? Now, there are things that can be said here — roughly, the sorts of explanations that proponents of possible worlds semantics give of apparent cases of believing necessary falsehoods. But those who are unconvinced by these explanations will regard this consequence of Benj's theory as an unwelcome one — and it's not one that Benj can avoid so long as he maintains the explanation of the way in which perceptual beliefs acquire their contents from perceptions.

So far I've only been talking about Benj's approach to veridical experiences which the subject takes to be veridical; in the later sections of the paper, Benj provides an extensive taxonomy of the different ways in which experiential episodes might fall short of this norm. Here I'll just focus on what Benj has to say about the familiar case in which a subject is having a hallucinatory experience which she mistakenly takes to be veridical.

Suppose, in particular, that our subject is Sam, and that Sam is dreaming that he is looking at the red color of a widget. In this case, Sam will, by virtue of so dreaming, accept a sentence  $\mathcal{R}_\delta$  which represents Sam as instantiating the property of dreaming that he is looking at the red color of a widget — since, as above, Sam and this property are both terms which represent themselves in our Lagadonian language. What perceptual belief will Sam form in this case?

Benj’s idea is that the way to answer this question is by looking at Sam’s conditional evidential policies, which we can suppose to include the following two:

(A) Regard  $\mathcal{R}$  and “I am looking at the red color of a widget” as equivalent if I am looking at the red color of a widget.

(B) Regard  $\mathcal{R}_\delta$  and “I am dreaming that I am looking at the red color of a widget” as equivalent if I am dreaming that I am looking at the red color of a widget.

The question is then how these policies are related to the content of the belief actually formed. It can’t be that adopting policies (A) and (B) is sufficient for one to form, in a particular situation, whichever belief (A) and (B) dictate — for, if this were sufficient, one would always believe that one is veridically perceiving when one is, and always believe that one is dreaming when one is. And of course we don’t do this, since we can be mistaken about whether we are dreaming or veridically perceiving.

The fact that we don’t do this seems to me to be a problem for Benj’s theory. Recall that, in the veridical case, the content of a belief is determined by the proposition associated with the perceptual experience via the subject’s regarding the experience as equivalent to the sentence in her “belief language.” As noted above, one might worry about what, exactly, it means to regard a sentence of the Lagadonian perceptual language as equivalent to a sentence of the belief language. But, *whatever* it takes for a subject to regard these sentences as equivalent, it seems that a subject might take exactly the same attitude toward an episode of dreaming and a sentence of the belief language. And if the subject can do this, it is hard to see, on Benj’s picture, why this should not be sufficient for the subject to believe that she is dreaming. But one simply can’t, in this way, form true beliefs about whether we are dreaming or having a veridical experience — it isn’t that easy! This, I think, casts some doubt on Benj’s explanation of how belief formation works in the case of veridical experience.

To press this point for just a moment: consider the veridical case, in which I instantiate the property of looking at the red color of a widget, and the dreaming case of type D/M, in which I instantiate the property of dreaming that I am so looking but don’t know that I do, and let’s stipulate that in each case I’m equally convinced that my experience is veridical. (My dispositions to act are the same,

in each case I assert that my experience is veridical, I am disposed to take just the same bets about the veridicality of the experience, etc. — add in whatever seems required.) What I think needs some explanation is why in the veridical case I manage to regard my instantiation of the relevant experiential property — in that case, the property of looking at the red color of a widget — as equivalent to some belief sentence, and that in the dreaming case I don't manage to do that with my instantiation of the relevant dreaming property. This looks mysterious to me because it seems that in the two cases my attitude toward the experiential episode I'm undergoing is exactly the same.

It's natural to try to answer this challenge by appealing to the conditional evidential policies (A) and (B); but it's not obvious to me that this helps. Even if it is my policy to regard  $S$  and  $S^*$  as equivalent only when  $p$  is the case, it does not follow that I *will* regard them as equivalent when  $p$  is the case, or that I won't when it isn't. But if this is granted, then it should be possible, in the dreaming case just described, that I regard  $\mathcal{R}_\delta$  and “I am looking at the red color of a widget” as equivalent. (I am, after all, as certain as I ever am that this is just what I am doing.) But then, given Benj's claims about regarding as equivalent, it follows that  $\mathcal{R}_\delta$  and “I am looking at the red color of a widget” are synonymous for me. Since  $\mathcal{R}_\delta$  is a sentence of a Lagadonian language, I take it that it cannot change its meaning; which implies that, for me, “I am looking at the red color of a widget” means that I am dreaming that I am looking at the red color of a widget. And this, in turn, means that I believe that I am looking at the red color of a widget. But I plainly don't believe this in the above case.

We're thus forced to the conclusion that it is impossible for a subject who does not have the correct beliefs about which experiential property she is instantiating to regard her instantiating that property as equivalent with any sentence — since, otherwise, she would, contra our supposition, have the true beliefs about, e.g., whether she is dreaming or veridically perceiving. My problem is that I don't quite see what “regard as equivalent” could mean which would secure this result.<sup>5</sup>

Returning to policies (A) and (B), it's clear that neither policy has anything to say about the case in which I am dreaming that I am looking at the red color of a widget, but believe that I am looking at the red color of a widget. So what should we say about this sort of case? Benj says:

---

<sup>5</sup> One might say: this is impossible because ‘regard as equivalent’ is sufficient for synonymy, which makes it impossible that a subject should ever regard as synonymous  $\mathcal{R}_\delta$  and “I am looking at the red color of a widget.” But I think that this gets Benj's preferred order of explanation backwards: “I am looking at the red color of a widget” and other sentences of the language of belief are supposed to get their contents *from* being regarded as equivalent to the relevant Lagadonian sentences; they don't have meanings independently which are available to constrain the objects which the subject is able to regard as equivalent. Otherwise, I think, we'd lose Benj's explanation of the truth of the beliefs formed in the ordinary veridical case.

“The question is not easily posed from the first-person perspective. If one is under the impression one is looking, then from the first-person perspective things are this way: I am looking. Fixing this, the question of what to do if one tokens  $\mathcal{R}_0$  is then a question of what to do in an incoherent situation. Rationalizing policies and rules provide answers about what do if things are this way or that way; given a way things can’t be, such policies are silent. ....

At this point, we see what I take to be the root of philosophical perplexity about perception. In a delusive case, *one’s perspective is incoherent*: the perceptual aspects of one’s perspective affirm a certain hypothesis; the doxastic aspects affirm a certain incompatible hypothesis. In such circumstances, all bets are off from the point of view of intentional psychology....”<sup>6</sup>

The idea is that, just in virtue of dreaming that he is looking at the red color of a widget, given our remarks about the Lagadonian language above, Sam represents himself as dreaming that he is looking at the red color of a widget. But he believes himself to be looking at the red color of a widget; since it is impossible to be both looking at the red color of a widget and dreaming that one is looking at the red color of a widget, the proposition which is the content of Sam’s belief is inconsistent with the proposition associated with his perceptual state.

This leads Benj to say two surprising things about Sam. The first is that there is “no coherent answer” to the question of what it is like for Sam. The second is that, for the reasons just given, Sam’s perspective is incoherent, and that for this reason, in Sam’s case, “considerations of rationality do not apply.” I find both of these conclusions hard to accept; I’ll discuss them in turn.

About what it’s like to be Sam, Benj says

“So what then *is* it like for Dreaming Sam? The question admits of no coherent answer, because the condition the world would have to meet in order for it to be faithful to how the world is ‘for Dreaming Sam’ is unsatisfiable.”<sup>7</sup>

But this seems to me to be a *non sequitur*. Even if (and here I agree with Benj) there is a certain kind of equivalence between what it’s like for Dreaming Sam and the condition which the world would have to meet to be faithful to Sam’s experience, we can’t infer from the fact that the world could not satisfy this condition that there is no such condition — any more than we can infer from the

---

<sup>6</sup> Hellie (2011), 153.

<sup>7</sup> Hellie (2012), 8.

necessary falsehood of a mathematician's belief that there is nothing that that mathematician believes. If we can coherently describe incoherent beliefs, why not say the same about Sam?

Further, it seems to me that there must be at least some coherent things that we can say about what it is like for Sam. Consider Sam', who is like Sam but for the fact that he is dreaming that he is looking at the green color of a widget. Surely what it's like to be Sam' is different than what it's like to be Sam; if we deny this, then it seems that we've lost track of the notion of 'what it's like' which made it seem interesting in the first place. But if we accept this, then, contra what Benj says, it seems to me that there must be facts about what it is like for Sam and Sam'.<sup>8</sup>

Let's turn now to Benj's claim that "considerations of rationality do not apply" to Sam. One way to bring out just how surprising this claim is is to imagine Sam and Sam\*, each of whom are dreaming that they are looking at the red color of a widget and each of whom mistakenly takes themselves to be looking at the red color of a widget. On the basis of this experience, Sam comes to believe that there is a red widget before him, and Sam\* instead forms the belief that there is a blue widget before him. Surely there is a straightforward sense in which Sam's response to his dream is more rational than Sam\*'s — even if we want there to be a sense in which Sam's response is less fully rational than the response of a subject who forms this belief on the basis of a veridical experience of the red color of a widget. Benj tries to capture this intuition by saying that it is indeed more natural to form Sam's belief than Sam\*'s — but while this is no doubt true, I don't think that this succeeds in capturing the intuition, which I find quite compelling, that Sam was rational to form his belief, and Sam\* (bizarrely) irrational to form his.

This might just boil down to a battle of intuitions, and Benj might fairly point out that this bullet might well be worth biting to preserve the sort of direct realist picture to which he is drawn. But I wonder whether one could preserve much of that direct realist picture without having to say these surprising things about Sam and Sam\*.

Even if we grant that in Sam's case the proposition associated with his dream state is inconsistent with a proposition he believes to be true, this fact doesn't by itself show that Sam is now wholly outside the realm of rationality. Even proponents of coarse-grained views of contents think that we have to say something about the rationality of subjects with inconsistent commitments, if only because inconsistency is so common. Indeed, it is especially common when the subject's inconsistent commitments are such that the subject herself fails to recognize their incompatibility. And the sort of inconsistency which arises in the dreaming case seems — given the aspects of Benj's framework sketched above —

---

<sup>8</sup> I think that Benj makes these claim about Dreaming Sam in order to deny PI. But one could deny PI, and admit the existence of indiscriminable but genuinely distinct "what it's likes", without denying that there is anything that it's like to be Dreaming Sam.

to be a case of compartmentalization of just this sort, since the subject who believes that she is perceiving veridically is apparently in no position to know that she is correctly representing herself (in the Lagadonian language) as dreaming that she is looking at the red color of a widget. Given that we should have something to say about subjects whose commitments are globally inconsistent the claim that “considerations of rationality do not apply” to subjects like Sam and Sam\* seems like an overreaction.

It’s also worth noting that the idea that Sam’s dream and his belief are inconsistent is not an essential part of the direct realist picture; the alleged contradiction is generated by (i) the self-representational aspect of Benj’s theory, on which the proposition which a subject accepts is not just about the red color of the widget apparently before him, but also about the subject’s relation to that widget and (ii) the claim that dreaming, just as much as veridically experiencing, has this self-representational aspect. But one might wonder — especially from the perspective of a direct realist who is unafraid to think of veridical experiences and matching hallucinations as belonging to very different categories — about the motivation for (ii). Remember that, so long as we want to avoid the conclusion that every subject represents himself as having every property which he has, that we have to find some way of distinguishing between those properties which are expressions of the Lagadonian language and those which are not. So why not think that the property of looking at the red color of a widget is one of the properties in the former category, and the property of dreaming that one is looking at the red color of a widget is not? Why not say that the mechanism by which we form true beliefs in the veridical case is radically different than the mechanism by which we form true beliefs about our hallucinatory and illusory experiences? This would avoid the conclusion that subjects who are dreaming represent themselves as such, and hence would avoid the conclusion that subjects who are “taken in” by a hallucination are thereby inconsistent as well as simply mistaken about the scene before them.

The availability of this option is important even if the theory which results from taking it ends up not being attractive. It is important because it shows that the claim which Benj takes to be the “root of philosophical perplexity about perception” — namely, that, “[i]n a delusive case, *one’s perspective is incoherent*” — is not generated by Benj’s direct realism. Quite the opposite: it is generated by Benj’s commitment to there being a certain kind of *commonality* between veridical and hallucinatory experience: namely, that both involve accurate self-representation, in the Lagadonian language, of one’s current experiential state.

#### REFERENCES

- Benj Hellie (2011), “There it is,” *Philosophical Issues* 21: 110-164.  
Benj Hellie (2012), “There it was,” this volume.