

Thomas Raleigh

## *A New Approach to 'Perfect' Hallucinations*

**Abstract:** *I consider a new, non-disjunctive strategy for 'relational' or 'naïve realist' theories to respond to arguments from 'perfect' (causally matching) hallucinations. The strategy, in a nutshell, is to treat such hypothetical cases as instances of perception rather than hallucination. After clarifying the form and dialectic of such arguments, I consider three objections to the strategy. I provide answers to the first two objections but concede that the third — based on the possibility of 'chaotic' (uncaused) perfect hallucinations — cannot obviously be dealt with by the proposed strategy. However, such 'chaotic' scenarios are also problematic for standard representational accounts of experience. Thus I conclude that perfect hallucinations pose no more of a threat to the relational theory than to its main representational rival.*

Though it surely remains the standard view, both amongst neuroscientists and amongst philosophers, that conscious experience supervenes only on the brain (or some parts/processes within the brain), in recent years there has been a very marked surge of interest in 'naïve realist' or 'relational' theories of perceptual experience.<sup>1</sup> Such theories deny that *all* conscious experiences supervene solely on factors internal to the subject's body — perceptual experiences, the so-called 'good case', are alleged to essentially or constitutively involve elements in the external environment as well as elements internal to the subject. This recent interest has been due, in large part, to the develop-

Correspondence:  
Thomas Raleigh, Norwegian University of Science & Technology.  
Email: [thomas.raleigh@ntnu.no](mailto:thomas.raleigh@ntnu.no)

---

[1] For example Martin (1997; 1998; 2002); Campbell (2002; 2007; 2009); Fish (2009), Brewer (2006; 2011); Logue (2011; 2012); Johnston (2004; 2006); Hellie (2012).

ment of new, disjunctivist strategies for responding to arguments from hallucination.<sup>2</sup> It is widely agreed that the strongest versions of such arguments that a disjunctivist faces are based on the possibility of ‘perfect’ or ‘causally matching’ hallucinations.<sup>3</sup> It is in response to this sort of possibility that M.G.F. Martin has developed his ‘negative epistemic’ account of hallucinations.<sup>4</sup> Martin’s work, I think it is fair to say, represents the current state-of-the-art in disjunctivist strategies for defending a naïve realist/relational account of experience.<sup>5</sup>

This paper is not about disjunctivism. Instead, I want to consider a different, non-disjunctive line that one sympathetic to the relational model of perceptual experience might take in response to arguments based on ‘perfect’ hallucinations. In brief: the usual thought-experimental cases of ‘perfect hallucinations’ should be treated as relational experiences of something external. I suspect that this line of response may initially seem somewhat radical; that is to say, implausible. But I hope that by the paper’s end I will have made the case that this sort of response has a much firmer dialectical footing than might first appear. Indeed, I will argue that by adopting this line of response, the relational theory ends up in the same position with respect to perfect hallucinations as its main rival, the representational theory. But to begin I will briefly discuss the general form and dialectic of arguments from hallucination. I’ll also clarify what I take the basic structure of the relational model of experience to be and how it differs from other possible accounts.

### 1. Arguments from Hallucination and from Perception

Boiling arguments from hallucination down to their simplest form, there are two premises and a conclusion:

- (1) **Natural H-premise:** Hallucinations are conscious episodes that are not essentially relational.
- (2) **Common kind assumption (CK):** Hallucinations and perceptual experiences are the same essential kind of conscious episodes.
- (3) **Revisionary P-conclusion:** Therefore, perceptual experiences are conscious episodes that are not essentially relational.

[2] See, for example, Hinton (1967; 1973); McDowell (1982; 1994); Snowden (1980).

[3] See Robinson (1985; 1994); Snowden (2005); Smith (2002); Martin (2006).

[4] Martin (2002; 2004; 2006).

[5] Though see also Fish (2009); Langsam (1997).

When discussing arguments from hallucination, most philosophical effort is typically expended in considering premise (2), the CK assumption. The two standard reasons adduced in support of CK are:

- (i) Hallucinatory and perceptual experiences might be 'subjectively indistinguishable' to the subject.
- (ii) Hallucinatory and perceptual experiences might involve exactly the same neural activity.

In each case we need a further linking conditional claim in order to support CK:

- (iii) If two experiences are subjectively indistinguishable then they are of the same essential kind.
- (iv) If two experiences involve exactly the same neural activity then they are of the same essential kind.

The disjunctivist strategy, of course, is to deny CK and so if, as has seemed plausible, we must allow the possibilities described in (i) and/or (ii) then a disjunctivist will need to deny the conditional claims made in (iii) and/or (iv). Of course, much more should be and has been said about all this, but I will say no more about the possibility or plausibility of denying CK.

In contrast to the attention lavished on CK (and on claims like (iii) and (iv)), premise (1) — the 'Natural H-premise' — in our basic argument from hallucination is rarely argued for or scrutinized. Perhaps it is taken as simply the meaning of 'hallucination' that it is a conscious episode in which the external world does not play any essential role.

However, it is worth noticing that *even accepting the CK assumption*, a theorist who wished to hang on to a direct-relational model of perceptual experience could construct the following 'argument from perception':

- (1\*) **Natural P-premise:** Perceptual experiences are conscious episodes that are essentially relational.
- (2\*) **CK assumption:** Hallucinations and perceptual experiences are the same essential kind of conscious episodes.
- (3\*) **Revisionary H-conclusion:** Therefore, hallucinations are conscious episodes that are essentially relational.

The two arguments are entirely symmetrical — the conclusion of each is the denial of the other's first premise. Both go from an intuitively 'natural' starting point, via the CK premise, to a 'revisionary' conclusion. Given the symmetry it would seem that the only room left for argumentative manoeuvre would be to give reasons *independent of*

*either such arguments* that one or other natural premise is preferable, or equivalently that one or other revisionary conclusion is the less acceptable. (Notice: I am not suggesting that the best reasons for wanting to endorse either ‘Natural’ claim will be their (alleged) naturalness or intuitive plausibility: e.g. there are many possible reasons to want to endorse the Natural P-premise *other than* its (allegedly) fitting with the views of the naïve. For what it’s worth, I think that it’s far from clear that the non-philosophical populace do straightforwardly endorse either Natural claim, and even if they did it’s not at all clear that this should bear much dialectical weight.)

CK, in itself, does not point us towards either the Revisionary P-conclusion or the Revisionary H-conclusion. Moreover, for the conclusion of the argument from hallucination to be decisive — i.e. *preferred* to the conclusion of the symmetrical argument from perception — it is not enough just to argue that the Natural H-premise (1) is plausible and counter-intuitive to deny, it has to be shown *on grounds other than the argument from hallucination* that the Natural H-premise (1) is *more* plausible and counter-intuitive to deny than the Natural P-premise (1\*) (i.e. denying the Natural H-premise is *worse than* denying the Natural P-premise). It was supposed to be the *conclusion* of the argument from hallucination that we should deny, against our naïve intuitions, the Natural P-premise, yet now we need reasons to deny the Natural P-premise, rather than denying the Natural H-premise, *before the argument from hallucination gets started*. CK then can only get us to one revisionary conclusion rather than the other if we already have *prior*, independent reasons to prefer one revisionary conclusion to the other.

At this point you might, quite reasonably, want to protest: OK, we can all accept the elementary logical point that CK could be used in an ‘argument from perception’ as well as an argument from hallucination. One person’s *modus ponens* is another’s *modus tollens*. Nevertheless, whatever motivations in favour of (1\*) there are supposed to be, to endorse (3\*)/deny (1) is *just absurd!* *Of course* there are conscious experiences, which we normally call ‘hallucinations’, in which we do *not* have any essentially relational conscious awareness of mind-independent features (likewise with dream experiences). So the ‘argument from perception’ is a non-starter; the argument from hallucination is the only one we need consider.

I agree that it would be far-fetched to claim that *all* conscious experiences, both perceptions and hallucinations (and dreams), are essentially relational, constitutively involving mind-independent features. And yet something like this position has been endorsed by some

theorists. Both Alston (1999) and Langsam (1997) have endorsed (or at least toyed with) the view that in visual hallucinations we have (direct) conscious awareness of regions of physical space, i.e. the kind of conscious episode is still essentially a relation involving the mind-independent environment. Then there is Johnston (2004), who proposes that in hallucination we have relational conscious awareness of uninstantiated universals, or Butchvarov (1994) who claims that we have conscious awareness of Meinong-style non-existents.

Though we should at least allow for the possibility that it is our intuitive conception of hallucinations, rather than of perceptions, that we end up having to revise, I don't want to endorse any of these revisionary views concerning hallucinations *in general*. Everyone should accept that there are in fact a variety of real-life conscious episodes in which we have rich visual experiences that are not instances of relational awareness of mind-independent features — e.g. vivid dreams, sensory deprivation, drug-induced reveries, etc. But such episodes are *not* the sort of possibility described in (ii); they do not precisely replicate the neural activity or stimulation of a perceptual experience. So they cannot be used, via (iv), to support the CK assumption. And so they put little pressure on the relational theorist. Nor is it obvious that such real-life hallucinatory episodes normally satisfy condition (i) either. Whilst undergoing dreams or drug-induced hallucinations one might take the experience to be of real environmental items; one may not be able to distinguish the situation from one of perception. However, dreams and hallucinations often affect judgment as well as sensory faculties. And on later calm reflection, subjects generally *don't* want to say that such episodes are phenomenally identical to everyday perceptual experience. After all, we have adjectives such as 'dream-like' or 'hallucinatory', which are often used precisely to describe the *distinctive phenomenology* of such episodes, i.e. that an hallucinatory experience may not, at the time, be distinguishable from some normal perceptual situation does *not* show that a normal perceptual experience is, at the time *it* occurs, indistinguishable from the hallucinatory situation. So these familiar, real-life sorts of episodes, of dreaming and hallucinating, cannot be used to support CK and so they are just not relevant to the argument from hallucination.

In so far as perfect hallucinations pose the real threat, a relational theorist might hold that *for the restricted class of 'perfect hallucinations' only*, we *can* endorse a more limited version of the argument from perception:

- (1#) **Natural P-premise:** Perceptual experiences are conscious episodes that are essentially relational.
- (2#) **Restricted CK assumption:** Perfect hallucinations and perceptual experiences are the same essential kind of conscious episodes.
- (3#) **Revisionary Perfect H-conclusion:** Therefore, so-called perfect hallucinations are conscious episodes that are essentially relational. (And so such episodes should not really be called ‘hallucinations’ after all.)

If this is a viable strategy, then the relational theorist might simply accept (i) and (ii) whilst holding onto her preferred model of perceptual experience. The class of experiences that is supposed to cause the most problems for the relational model will have been accommodated without recourse to the subtleties of disjunctivism. Now, to anticipate, there remains an even more restricted, hypothetical class of perfect hallucinations that cannot be treated in this way — what I’ll call Perfect Chaotic Hallucinations, in which there *is just nothing* external to the brain for it to be related to. But, as I will argue in Section 6, these highly fanciful possible cases are equally problematic for standard versions of the representational view of experience. And so the dialectical position we are left in is that perfect hallucinations pose no more of a problem for the relational theory than they do for the main rival representational position.

## 2. What Exactly is the Relational Model of Experience?

It is worth briefly clarifying what might be meant by a ‘relational model of experience’.

Here are two very schematic diagrams which, taken together, illustrate the standard, common factor model for both perceptual and hallucinatory experience:

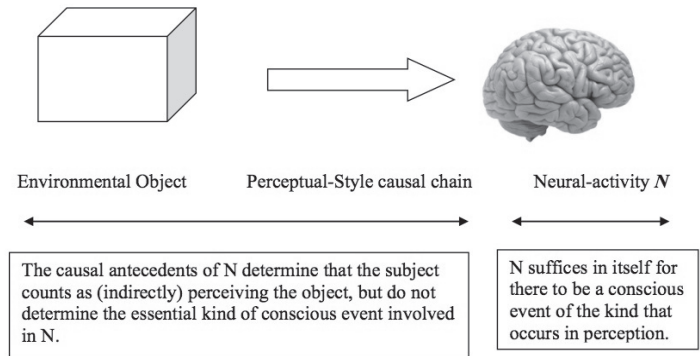


Figure 1. Phenomenal internalist model of perception.

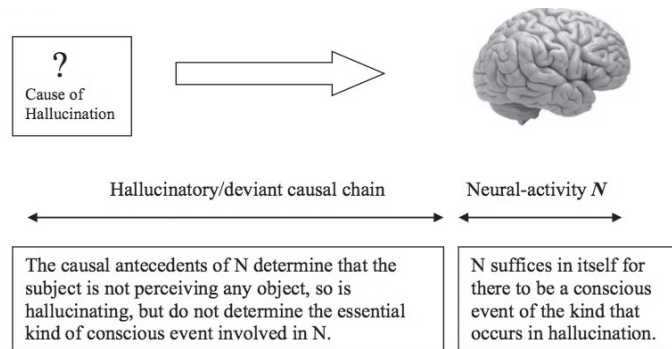


Figure 2. Phenomenal internalist model of hallucination.

On the common factor conception of perceptual experiences — which combines phenomenal internalism about perception and hallucination — the kind of conscious event that occurs in perceptual and hallucinatory cases can be the same essential kind of consciousness. The causal antecedents in each case determine whether the subject counts as perceiving or hallucinating but they have no bearing on the essential kind of conscious awareness that occurs in each case, which supervenes solely on factors internal to the subject’s brain (and nervous system).

Now consider these diagrams of two different conceptions of how an external environmental object or feature might be essentially involved in a conscious perceptual episode:

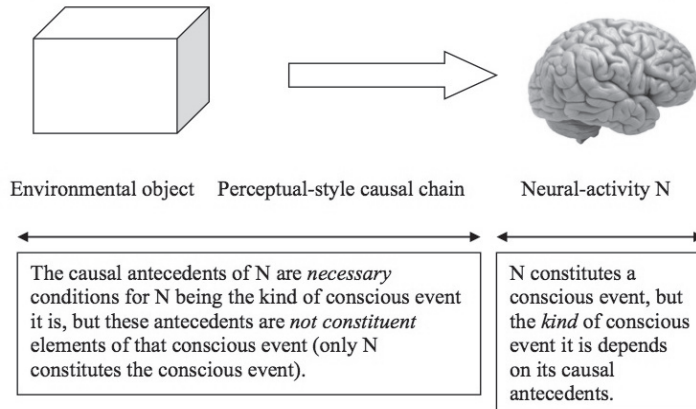


Figure 3. Non-proximate necessary causal preconditions (perceptual).

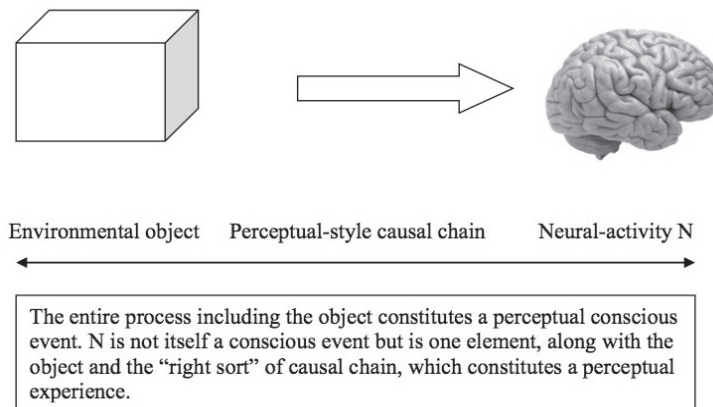


Figure 4. Constitutive environmental element (perceptual).

According to the necessary causal preconditions conception (illustrated in Figure 3), the environmental object has some kind of essential involvement in determining the fundamental *kind* of conscious state that occurs — a key role that is familiar from externalist accounts of mental content. But this kind of object-dependence is quite different to the sort of object-dependence conceived by the constitutive model of experience (illustrated in Figure 4). According to this latter view, the environmental object is literally a *part* of the experience itself, not just an essential precondition for the experience to occur. These two different conceptions of the perceptual case suggest two parallel possible conceptions of hallucinatory experience:



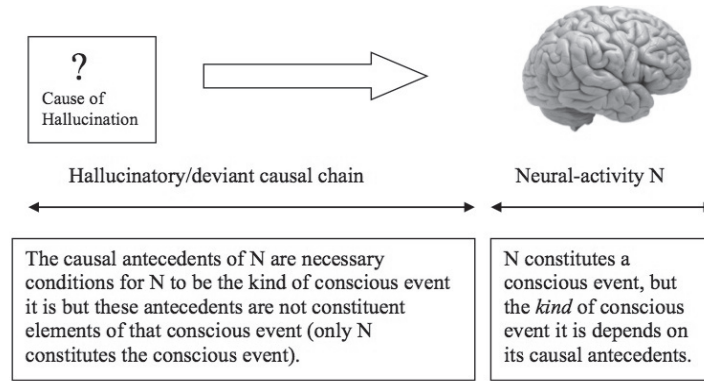


Figure 5. Non-proximate necessary causal preconditions (hallucinatory).

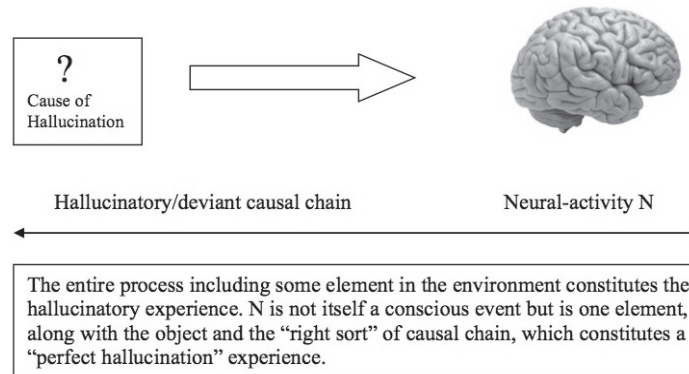


Figure 6. Constitutive environmental element (hallucinatory).

Whilst the orthodox common factor view is a combination of a phenomenal internalist view of perception with the phenomenal internalist view of hallucination (Figures 1 and 2), the various models above suggest a range of other possible combinations. If I understand him right, the combination that Martin (2006) argues for is to treat perceptual experiences as having necessary causal preconditions (Figure 3) whilst accepting the phenomenal internalist model for hallucinations (Figure 2) — that is, continuing to treat the hallucinatory case the same way as a common factor theorist. This approach then has to concede that when N occurs in the perceptual case it is *both* a conscious event of the distinctively perceptual kind *and* a conscious event of the kind that occurs in hallucinations also. To concede that there is a

common kind of conscious event that occurs in both the perceptual and hallucinatory cases might seem pretty damaging to the relational view, as this common factor threatens to ‘screen off’ the distinctively perceptual kind of event from doing explanatory work. But Martin claims to avoid the ‘screening off’ threat with his negative epistemic account of hallucinations, on which there is *nothing more* to being an hallucination than to be indistinguishable by reflection from the corresponding perception. Given this ‘parasitic’ account of hallucination, the distinctively perceptual event is, so Martin argues, no longer ‘screened off’ from being explanatory.

A combination that Snowdon (2005) argues is ‘dubiously coherent’ would be to treat perceptual experience according to the constitutively environmental model (Figure 4) whilst treating hallucination according to the phenomenal internalist model (Figure 2). Such a combination of models would entail that there are two experiences in the perceptual case, one that is constituted by N alone and one in which N is just a component.

...[I]t seems to me that this view is dubiously coherent. It is hard to understand the possibility of, say, simultaneously undergoing an experience that is a hallucination as of a pink elephant against a black background, and also another perceptual experience of a pink elephant against a black background. As total experiences they seem to *compete*. (Snowdon, 2005, p. 303)

Instead, Snowdon tentatively suggests a disjunctivist might think of perception as requiring a constitutive environmental element (Figure 4) whilst thinking of hallucinations according to the necessary causal preconditions model (Figure 5). Those already committed to a common factor theory might find it implausible that an hallucination, a paradigmatically ‘inner’ conscious event, should have externalist necessary conditions for that kind of conscious event to occur. But anyone sympathetic to a relational model, or to externalism about the mind more generally, might well feel that this is a pretty minor revision, a small price to pay, if it allows one to hold onto the relational model of perception.

This leaves further possible options. Most obviously a theorist might think of both the perceptual and the hallucinatory case according to the necessary causal precondition models (Figures 3 and 5). This would seem a natural position for a theorist to take in so far as it posits one set of non-proximal necessary conditions that make N a conscious event of one kind and other non-proximal necessary conditions that make N a conscious event of another kind. This would seem

to avoid the issue that faced Martin's combination — externalist preconditions for perception with phenomenal internalism for hallucination — where the perceptual episode involves a conscious event that is essentially of both kinds. So long as in principle all the possible non-proximal conditions divide into those that make N an instance of perceptual awareness and those that make N an instance of hallucinatory awareness, then N will never be a conscious event of both kinds. Now there are, of course, notorious difficulties in providing neat, non-circular criteria for deciding whether a causal path from object to subject is of the right kind for the subject to count as perceiving (as opposed to, say, veridically hallucinating). But just these same difficulties will eventually have to be faced by common factor theorists as well. Everyone wants to be able to mark *some* kind of distinction between perceptions and hallucinations even if one thinks that the essential kind of conscious state/event that occurs is the same in both cases, and so one still faces the question of how to distinguish the 'right kind' of causal chains from 'deviant' chains even if one is a phenomenal internalist about both perception and hallucination.

Regarding the causal preconditions model of perception (Figure 3), one might well wonder whether it really captures the core 'naïve realist' notion of perceptual experience being 'openness to the world'. The non-proximate necessary conditions on N counting as an event of *perceptual* consciousness are analogous to externalist conditions on the content of a mental state. But the idea of perceptual experience being a genuine relation, or *encounter*, involving both subject and object, is, you might think, quite different from the notion that perceptual experience possesses object-dependent *individuation conditions*. The worry then is that appealing to external causal preconditions is more suitable to an intentionalist theory of experience, on which the content of perceptual experience is distinctively object-dependent. Could we really have *genuinely relational awareness* when N *by itself* is admitted to constitute the conscious experience and the object merely provides a necessary precondition for that kind of conscious event to occur? This is not, of course, any kind of *argument* that appealing to necessary external conditions on certain kinds of experience cannot capture the idea of 'direct' or 'unmediated' awareness of a mind-independent feature; I am merely voicing a worry or suspicion.

In any case, for the remainder of this paper I will simply assume that a relational or 'naïve realist' theory of perceptual experience is best captured by the constitutive environmental element model (Figure 4). The possibility I now want to explore is whether one might also

coherently think of a *perfect* hallucination along the same constitutively environmental lines (Figure 6).

### 3. (Most) So-Called ‘Perfect Hallucinations’ are Perceptions

The usual way in which the thought-experimental scenario of a ‘perfect hallucination’ is set up goes something like this:<sup>6</sup> take some normal perceptual experience — in which one sees, say, a lemon — and somehow monitor exactly the pattern of stimulation that occurs in the subject’s optic nerves resulting from the cells ‘upstream’ firing in the subject’s retina. Now detach the subject’s retina from the optic nerve, but recreate exactly the pattern of stimulation that the optic nerves received by attaching them to some sort of hypothetical machine. Alternatively, one might prefer not to use the same subject at two different times — for the overall neural states of the brain are bound to be slightly different given the passage of time, the new memories formed, etc. So perhaps instead we should imagine a type-identical Twin subject, or type-identical brain-in-a-vat being administered the type-identical pattern of stimulation. Or yet again, we could frame things in terms of the same individual, in exactly the same neural state, getting the same pattern of stimulation in two different possible worlds, in one due to a lemon, in the other due to some neuroscientific trickery. However, with Twin subjects, or with the same subject in two different possible worlds, it becomes harder to make sense of the two experiences being ‘subjectively indiscriminable’ — there is no opportunity for a single subject to discriminate between the two. So perhaps then it is simplest to appeal to an evil demon, whose precise means of recreating exactly the same neural state and exactly the same pattern of stimulation as in the normal scenario need not be further specified.

The machine/demon/neuroscientist is supposed to provide a stimulus input to the brain, but it is not supposed to meddle with the internal processing of the brain. And for good reason — if the demon/device is intervening in the neural processing then we do not have neural similarity with the good case and so we lose one main motivation to treat the two cases as being the same essential kind of conscious experience. (As we’ll see below, to plausibly match a temporally extended perceptual experience, the device would also have to monitor and respond to outgoing afferent impulses to saccade, change head position, focus attention, etc.)

---

[6] See, for example, Robinson (1994); Smith (2002); Johnston (2004).

In any case, proponents of these scenarios want us to conceive of it as one in which the machine has *caused* one to have an 'inner', *non-object-involving* visual experience — an experience where things look just as if there were a lemon before one. A blow to the back of the head might cause one to visually experience 'spots before one's eyes', but such an experience is not a visual experience *of* the punch (nor *of* the back of one's own head). Likewise, we are supposed to imagine that the machine caused the lemon-looking experience, but that the experience is not *of* the machine — we do not have (direct) visual awareness of the machine. I suggest that a relational theorist should resist this conception and contend instead that in such scenarios we *would indeed* be having a visual experience *of* some aspect or feature of the mind-independent machine. I will not be trying to show that such a relational account of a perfect hallucination is *superior* to the normal common factor approach. I aim only to show that a relational account of perfect hallucinations is a coherent and defensible view that anyone *already* sympathetic to the relational account of perception should seriously consider.

There is an analogous line of thought, much discussed in the philosophical literature on scepticism, which holds that a brain-in-a-vat (BIV) would have largely true beliefs about its vat-world, and so would presumably also count as *perceiving* its vat-environment. Putnam (1981), Davidson (1986), and Chalmers (2005) are perhaps the most prominent philosophers to have urged that a (long-term) BIV would have true beliefs about its vat-environment — though the idea goes back at least to Bouwsma (1949).<sup>7</sup> Putnam argues for this on the basis of a causal theory of reference, whilst Davidson appeals to his preferred interpretational account of reference. Chalmers argues that were a 'Matrix' hypothesis true, this would not show that the world we experience is unreal, just that there is a further computational level to reality we had not suspected before and that the universe was bigger than we had previously thought — which would allow most of a Matrix-dweller's everyday beliefs to still be true. Putnam, Davidson, and Chalmers focus on the BIV's beliefs, but on all of these accounts we are also presumably meant to think of the BIV as *perceiving* its vat-world in so far as its visual experiences are allowing the formation of true beliefs about features of the virtual vat-world or Matrix. There is some distinguished support, then, for the idea that a BIV could be thought of as *perceiving* mind-independent features — not features of

---

[7] See Button (2013) for an excellent recent defence of a broadly Putnamian position on BIVs.

the environment that the vat is physically situated in, but rather mind-independent features in a virtual, computer-generated environment. Putting this in terms of representational content: the *content* of the (life-long) BIV's visual experience is *about* the vat-generated virtual environment, not about the physical environment that the vat and brain are both physically situated in. And so such experience should, in general, count as veridical perception rather than hallucination.

The position I am considering, on behalf of the relational theorist, agrees then with Putnam, Davidson, and Chalmers to the extent that a BIV should count as perceiving a mind-independent feature (albeit an unorthodox sort of feature in a very unorthodox sort of environment). However, there are also very significant differences between the view I am advancing and the Putnamian position on BIVs. Putnam, Davidson, and Chalmers are all making an *anti-sceptical* point. They are primarily concerned with the reference and truth of the subject's *beliefs*, not with the metaphysical nature of her experiences — i.e. theirs is an epistemological thesis concerning BIVs. In contrast, I am proposing a thesis about the metaphysics of experience; specifically the metaphysics of a perfect hallucinatory experience.

A common factor theorist about the metaphysics of experience can happily accept the Putnam/Davidson/Chalmers conclusion — that a BIV's beliefs would be largely true and would refer to elements in the virtual vat-environment — whilst insisting that a neurally identical normal subject and BIV will be in the same fundamental kind of conscious state. Likewise, it's possible for a common factor theorist to accept that a BIV's experiences veridically represent the vat-environment, whilst a neurally identical normal subject's brain veridically represents the normal physical environment. They just need to hold that experiences are fundamentally individuated on internal/phenomenal grounds and not according to their (wide) contents. After all, for a common factor theorist the question of whether the BIV is perceiving or hallucinating has no bearing on the essential kind of experience the BIV enjoys, which is held to be fundamentally the same in either case.

Conversely, the position I am proposing — that in a perfect hallucination the subject enjoys a relational, constitutively environmental experience — does *not* commit us to the Putnamian view that a (long-term) BIV, or a subject who has enjoyed perfect hallucinations throughout her life, would have largely true beliefs. The thesis is solely about the metaphysics of a perfect hallucinatory experience, *not* about the truth or the reference of any beliefs that get formed downstream from the experience.

Another significant difference is that the Putnam/Davidson/Chalmers thesis is only supposed to apply to *long-term* BIVs. If the brain in the vat has only recently been envatted, having previously been housed in a normal subject's cranium, then its thoughts and beliefs will still (to begin with at least) refer to items in the standard physical environment rather than to items within the virtual vat-environment. And likewise, presumably, given a standard sort of story about mental content, for the BIVs perceptual experiences: if I am suddenly transported into my vat after a previously unvatted life, then perhaps my visual experience as of a lemon will be counted as an hallucination of a real lemon. Whereas if I have been a life-long BIV (or a species of BIV that is adapted to the vat environment) then perhaps my visual experience will be counted as a perception of a vat-lemon, or a lemon\*. In contrast, according to the relational account of perfect hallucinations that I am proposing, even an otherwise normal subject who has just a brief, one-off 'perfect hallucination' would be in relational experiential contact with some feature external to their own minds — *viz.* some element or structure within the hallucination producing apparatus.

To be clear, then, the present proposal is not that subjects have exactly the same type of conscious experience when seeing a normal lemon and when having a vat-machine-involving experience indistinguishable from seeing a lemon. In one case the subject's experience essentially involves an actual lemon, in the other it essentially involves some aspect/feature of the stimulating vat-machine. But in both cases the experience has the structure of a genuine relation to a mind-independent feature. So the hypothesized perfect 'hallucination' would effectively be a rather novel and unorthodox kind of *perceptual* episode, no different in principle to seeing a plastic lemon, or a hologram of a lemon, or a trompe l'oeil painting of a lemon, etc. Note: some might balk at using the term 'seeing' when there are no eyes, nor photons, involved. Likewise some might want to resist calling this 'perception'. But these labels are unimportant — so long as it can be maintained that an experience with a genuinely relational structure occurs, where the mind-independent feature is a genuine constituent, 'shaping' the subjective phenomenology.

In short: whatever external factor (evil neuroscientist, vat-machine, demon, etc.) is causing the 'perfect' hallucination, the relational theorist can simply insist that this item (or some part/aspect/feature of it) is the object in a *relational* experience. Within the literature on the argument from hallucination and naïve realism, this position has not, so far

as I know, been explored at all. For the rest of the paper I will try to flesh it out in enough detail so as to make it seem at least a *live option*.

There are three obvious objections/issues this strategy faces:

- (1) You admit that there are cases of experiences — normal, everyday, imperfect hallucinations, dreams, seeing stars after being hit on the head, etc. — that are not essentially relations to external objects. In these cases, whatever external cause there may be for the experience — e.g. the object that strikes your head, the mes-caline consumed, etc. — is admitted NOT to be the *object* of experience. So why should these cases of perfect hallucination be treated any differently?
- (2) What is more, you admit that in these perfect hallucinatory cases the experience would *seem* just like the normal case of, say, seeing a lemon. So you admit then that the experience does *not remotely seem* to be a perception of a complicated neuroscientific machine (nor an evil demon, etc.). Why then should we think that this experience, apparently as of a lemon, is in fact a ‘perception’ of a neuroscientific machine (or an evil demon, etc.), when what is subjectively presented appears nothing like such a machine (nor a demon)?
- (3) Your strategy requires that there *is* some or other external cause of the experience in the perfect hallucination case. But we can imagine a possible situation in which there is exactly the same pattern of neural activity, but *due to no external causal factor whatsoever*. E.g. a lone brain floating through the void might recreate exactly the same pattern of activation as in a perception through sheer random chance — quantum fluctuations inside the neurons, or some such. If your strategy is unable to deal with this ‘chaotic’ sort of perfect hallucination, why bother with it in (non-chaotic) cases where there is an external cause?

I will respond to these objections in turn over the next three sections.

#### **4. Perception vs. Sensation**

A relational theorist should be committed to the idea that a perceptual experience is not instantaneous — it always has some non-negligible duration. It also requires the right kind of structured stimulus and the right kind of ‘exploratory’ activity/impulses on the subject’s part. When these conditions are not met, the subject does not achieve perceptual contact with the external world, and instead enjoys *mere* visual sensation. (On the relational theory seeing is *not* built out of



visual sensations; rather, to see is to *cease* to have mere visual sensations and instead to have visual awareness of discriminated environmental entities.)

An example familiar from 'ecological psychology': a uniform level of light stimulation from all directions results in an experience of uniform whiteness or '*Ganzfeld*'. There is a visual sensation — described as 'white' — in response to optical stimulation but there is no perceptual awareness of one's surroundings. Likewise, exposure to a sudden flash of light after adapting to darkness is experienced as no more than a sensation — there is no discrimination of any environmental entity or feature.

The point here — made familiar to philosophers from the work of Noë, Thompson, O'Regan, Hurley<sup>8</sup> — is that we do not gain perceptual awareness of the entire scene before us all at once, the moment we open our eyes. Noë calls this mistaken idea the 'snapshot conception' of visual experience (Noë, 2004) — that at any instant we have visual awareness *of the entire scene before us* that is 'sharply focused, uniformly detailed and high-resolution' as well as 'uniformly colourful'. A range of empirical results — most famously those on 'change blindness' — show that this simply is not the case. Our visual awareness of a rich stable environment occurs over time and requires a process of exploring the scene with our eyes; through the saccades our eyes make three times a second and through movements of head and body. Indeed, if our eyes are prevented from saccading and the head from moving (and the scene before the subject is unchanging), after a few seconds the subject will lose visual awareness of the scene.<sup>9</sup> Simon Ings, commenting on (and perhaps slightly dramatizing) this result, writes:

The eye exists to detect movement. Any image, perfectly stabilised on the retina, vanishes. Our eyes cannot see stationary objects, and must tremble constantly to bring them into view. (Ings, 2007, p. 45)

For our purposes the key lesson here is that gaining *perceptual visual awareness* of environmental features is a process that takes some non-negligible span of time to occur, it requires that the eyes actively (though mostly involuntarily) explore and 'probe' the objects in question and it requires some degree of stability and order in those objects during this period of scanning exploration. When these conditions are not met — if the stimulus reaching the eyes is too brief, if the eyes

[8] See, for example, Noë (2004); Thompson and Varela (2001); Hurley (1998), O'Regan (2011).

[9] See Hubel (1988, p. 81).

cannot move around, if the stimulus is varying too wildly over time or is completely constant — then the subject will be afforded only visual sensation, conscious awareness that is *not* a genuine relation with environmental features as constituents. John Campbell provides a nice analogy for how brain processing should be conceived of on the relational view:

If you are caught by the idea that the existence of brain processing means that a Representationalist view must be correct, it may be helpful to consider another analogy. Suppose we have a medium which, like glass, can be transparent. But suppose that, unlike glass, it is highly volatile and needs constant adjustment and recalibration if it is to remain transparent in different contexts. Suppose, in fact, that the adjustment required is always sensitive to the finest details of the scene being viewed... You might think of visual processing as a bit like that. It is not that the brain is constructing a conscious inner representation whose intrinsic character is independent of the environment. It is, rather, that there is a kind of complex adjustment that the brain has to undergo, in each context, in order that you can be visually related to the things around you; so that you can see them in other words. (Campbell, 2002, p. 119)

Continuing with Campbell's metaphor, in order for the mechanisms of our visual system to 'turn transparent' and so 'reveal the world' there is a certain minimum time and amount of activity required — they need, as it were, to warm up and become attuned to the world. Likewise, the environment providing the stimulus cannot be too 'un-patterned' or wildly varying if the visual system is to have the opportunity to actively explore it a bit. Given favourable circumstances of operation, these visual mechanisms afford the conscious subject direct, relational awareness of mind-independent features, they do not generate arrays of inner sensations. When these conditions are lacking the subject experiences a mere sensation. Or so says the relational theorist.

How does all this bear on the hypothetical perfect hallucinations and on the first objection mentioned above?

Well, given the foregoing constraints on a perceptual experience, the hypothetical hallucination machine is *not* to be imagined as simply providing a *momentary stimulus* to some region of the visual cortex, so as to produce, say, a momentary sensation of red. If we are to imagine that the hypothetical machine exactly recreates the pattern of the stimulus and neural activity that would occur when the subject has a *perceptual experience* of a lemon, say, the machine must be producing a *rich pattern* of the stimulus that lasts some non-negligible span of time.

Moreover, this pattern of the stimulus must take account of and be responsive to the active saccading that the subject would be performing. It would *not* suffice for the machine to provide the kind of stimulus pattern that might occur as a normal perceiving subject saccades around a visual field containing a lemon. The machine must take account of the subject's outbound saccadic impulses and volitions — otherwise the stimulus that the subject receives will not vary in accord with how she takes herself to be varying her point of focus. If it did not take into account these actual saccadic impulses then the subject (or the subject's visual system) might be expecting to shift focus to the left but receive visual stimulus as-of a shift of focus to the right of a lemon. In such circumstances it seems clear that condition (ii) is not met: the visual system + brain as a whole is not functioning identically to the perceptual case. Nor, presumably, would things seem the same phenomenally, as the subject would be aware of a confusing succession of sensations that bear no relation to their attentional activity, rather than a 'smooth' perceptual experience.

The imagined machine, then, cannot just provide an 'inbound' stimulus pattern to the two optic nerves, it must also be able somehow to monitor the 'outbound' saccadic impulses carried along the various cranial nerves to the six orbital muscles and to the ciliary muscles in each eye, and to adjust the stimulus pattern accordingly. Given the foregoing constraints, then, it becomes much more plausible that the machine (or the demon) provides some kind of *standing structure*, which the subject can explore via different possible saccadic impulses. Given that the machine cannot just provide momentary sensory input to a passive subject, but must also be responsive to the subject's active impulses and responses over a non-negligible stretch of time, different subjects hooked-up to the machine might visually explore the lemon-looking scene in different ways. I might scan left to right across the 'lemon', whilst you scan top to bottom. That different subjects could all agree on the way that things looked, despite having directed their attention/focus in different patterns, and so enjoyed differing experiences, lends much plausibility, I suggest, to the idea that they all had genuine relational awareness of the one same mind-independent feature — some kind of standing structure in the machine.

Moreover, in a perfect hallucination case, unlike in cases of real-life hallucination, dreaming, or being hit on the head, etc., the visual system is *by hypothesis* functioning perfectly normally, exactly as it would in the corresponding perceptual situation (from an internal point of view anyway). So we also have anatomical grounds for grouping perfect hallucinations with perceptions rather than with such

non-object-involving episodes as real-life hallucinations, dreams, and injuries to the head, etc.

### 5. Perceptual Relativity

How could having a visual experience indistinguishable from seeing a lemon possibly count as having (direct) visual awareness of some complicated neuroscientific device? The machine (presumably) simply would not look anything like a lemon!

The answer here is to bear in mind the relativity of perceptual appearances: whilst, of course, the hypothesized machine will not look anything like a lemon under most (the vast majority of) conditions of visual awareness, given one very specific manner of viewing, some specific part/feature of the machine *does indeed* look just like a lemon.

Any theory of perceptual experience must account for the fact that the same particular item can look different ways when experienced in different circumstances, i.e. it must account for perceptual relativity due to different angles of viewing, different lighting conditions, different conditions of the visual system, etc. A standard move made by a relational theorist here is to insist that perceptual experience is not just a simple two-place relation between conscious subject and object. It is a three-place relation between the subject, object, and the manner/circumstances in which the object is perceived.<sup>10</sup> The way the object appears is the joint upshot of the nature of the object and the nature of the perceptual circumstances — the angle and distance of viewing, lighting conditions, the condition of the subject's eyes and visual system, etc.

A penny looks a certain way when viewed from a certain angle (it looks the same way that a particular elliptical shape viewed from head-on looks). And the penny looks a different way when viewed from a different angle. Likewise, then, the idea is that the machine looks a specific way — the way a lemon looks viewed from some standard viewpoint — when viewed in a very specific manner — i.e. when the perceiving subject is 'hooked up' to the machine in the intended way. Presumably the exterior of the machine will not look like a lemon when viewed from any angle or under any lighting conditions, etc. But there will be some internal feature or aspect of the machine, a 'standing structure', that must have been designed by the hypothetical brain-scientists *precisely so as to appear* just like a lemon when it is

---

[10] See Campbell (2002; 2009) or Kennedy (2007).

experienced from a very particular 'viewpoint' — i.e. in the very specific manner and circumstances of awareness that obtains when the subject is hooked up to the machine in just the right way. But if this feature or structure in the machine is viewed from any other 'perspective', it will look nothing like a lemon.

Obviously the viability of this response turns on the viability of treating the subject's being hooked up to the machine as being just one more possible manner of gaining visual awareness of the machine — another viewpoint or perspective onto the machine. To repeat: if talk of the subject seeing the machine sticks in your throat, this relational strategy does not depend on any particular view as to the semantics of the word 'see'. All the naïve realist needs to maintain is that the subject's visual experience, whether or not you want to call it 'seeing', would be a genuine relation to a feature/aspect of the machine, and so would have the same fundamental metaphysical structure as other episodes of visual *perceptual* experience.

An analogy might help here: 'random dot' autostereograms are a kind of novelty image that enjoyed a burst of popularity in the 1990s in the 'Magic Eye' series of books.<sup>11</sup> When viewed normally these images appear to be just a fuzzy array of 'randomly' coloured dots. But if one learns to defocus one's eyes in just the right way (to focus as if on a point behind the image) then one sees what looks like a three-dimensional shape/scene. One can then focus on and visually explore different parts of this apparently 3D image just as one would when looking at an object normally — one can even view apparently 3D animation films. This manifest way that the pattern of dots can look does not require a particular angle of viewing, it requires that the visual system is 'geared' in the right way.

So the manner and circumstances of viewing, which contribute to the manifest way that things look, are not exhausted by the angle and distance of viewing and the lighting conditions; the way that the subject's visual system is 'set up' to engage with the environment is also a factor. The 3D way that a stereogram picture can look is an aspect of that picture that is open, in principle, for anyone to see so long as they can get their visual system 'geared' in the required way.<sup>12</sup> There is surely no pressing reason for a relational theorist to conceive of the subject's experience of the 3D look of the stereogram as 'inner' or hallucinatory. The subject is simply looking at the stereogram in a new

---

[11] Random dot stereograms were first discovered by Bela Julesz in 1959 and made into single images (autostereograms) in 1979 by Julesz's student Christopher Tyler.

[12] In fact 5–10% of people are unable to achieve this for various reasons.

manner and so is aware of the stereogram looking a new way (a way similar to the way that 3D objects can look). The stereogram is a mind-independent part of the environment, and the particular 3D way that it looks can be seen, in principle, by anyone who gets their visual system ‘set up’ in the required manner. The subject does not, according to the direct-relational theorist, begin to hallucinate when she defocuses her eyes in the required way — it is still the physical picture that is looking that way to her and which she is (directly) visually aware of.

I suggest that relational theorists should think of the subject getting hooked up in the right way to the perfect-hallucination-machine as being akin to the defocusing of the eyes required to see the stereogram’s hidden image; it is an instance of getting the visual system set up in a specific way so as to provide awareness of a specific mind-independent feature looking a specific way. The lemon-looking aspect of the machine will only be seen when the subject’s visual system is set up in the required manner — that is, say, having the device connected to the optic nerves in the right way.

The analogy with the autostereogram might also help to allay a further worry one might have: which ‘interior’ part or region of the machine exactly is the subject supposed to be visually aware of (when things are looking to them just the way that a lemon usually looks)? Well, consider the stereogram again: a certain pattern has been ingeniously incorporated into the design of the random dots. It requires looking at the dots in a particular way to see this pattern looking the intended 3D way. My suggestion is that there must be something like this pattern, a precisely constructed structure (e.g. a computer program determining the exact stimulus patterns) that is a designed feature of the machine. It requires getting one’s visual system hooked up in just the right way to experience this designed structure looking the intended lemon-like way.

There will be many parts and features of the machine that causally enable this designed structure to be revealed to the subject’s visual awareness but which are not themselves revealed. Just as one does not have visual awareness of one’s own eye nor of the photons of light that are reflected from the objects that one does see,<sup>13</sup> so there will be parts of the machine that play analogous roles — they are causally efficacious but invisible intermediaries. So, for instance, the subject is not supposed to be visually aware of the wires that dangle from her optic

[13] I assume that nearly everyone agrees that we do not normally see photons of light in visual perception. Though see O’Shaughnessy (1984) for a defence of the idea that we are constantly seeing light when we visually perceive.

nerves. The suggestion is that the subject is visually aware of some standing structured pattern that has been built into the machine — a mind-independent feature that can be visually explored, via indefinitely many different patterns of saccading, by any subject who gets their visual system hooked up in the required manner.

So far then: I have argued that when we consider the class of hypothetical cases where a *perfect* hallucination has been caused by some ingenious external device, the relational theorist is quite within her rights to treat such situations as instances of relational, world-involving experience, fundamentally akin to normal perception rather than to dreams or to imperfect, real-life hallucinations. In these situations the visual system would be operating normally and succeeding in detecting and exploring a real, standing mind-independent feature, something that other subjects in the same situation could likewise gain relational awareness of. Of course, I do not intend any of the foregoing to be persuasive to opponents of the relational theory. My aim has only been to convince those who are antecedently sympathetic to the relational view that this 'radical' approach to perfect hallucinations is not as crazy as it first seems. Moreover, we should recall here the dialectical lesson of the opening section — we should *prefer* having to revise our intuitions concerning the rare, strange, and so-far entirely hypothetical class of perfect-hallucination-experiences to revising our 'natural conception' of the familiar, everyday class of perceptual experiences.

### 6. Perfect Chaotic Hallucinations

Presumably it is in some sense possible, though *massively* unlikely, that the photoreceptors in the eye — or the ganglion cells in the optic nerve — might begin to fire (generate electrochemical signals) even though they have received no readily identifiable prior stimulus, but due only to random quantum fluctuation or some such freak circumstance. And so presumably it is possible, though *even more massively* unlikely, that such random firings precisely match the pattern of firings that would occur were the subject to see a lemon.<sup>14</sup> But in such a hypothetical case there would be no equivalent of the machine or demon to be a candidate object of visual awareness — the subject, or the disembodied brain, is perhaps just floating in empty space. So there would be, by hypothesis, exactly the same neural activity as in a perceptual visual experience but with absolutely no candidate object

[14] Remember, these randomly caused firings would have to maintain a pattern for a non-negligible period of time, and would miraculously have to *accord with the subject's impulses to saccade!*

of awareness in the environment. Here, then, it would seem that the CK assumption could only be employed in an argument from hallucination and not in an argument from perception as there just is no candidate feature to be the object relatum in a relational experience. David Chalmers has called this sort of possible situation the ‘Chaos Hypothesis’:

**Chaos Hypothesis:** I do not receive inputs from anywhere in the world. Instead, I have random, uncaused experiences. Through a huge coincidence, they are exactly the sort of regular, structured experiences with which I am familiar. The Chaos Hypothesis is an extraordinarily unlikely hypothesis... But it is still one that could in principle obtain, even if it has minuscule probability. (Chalmers, 2005, p. 23)

More recently, David Papineau, in a nod to Davidson’s Swampman example, uses the label ‘cosmic swampbrain’:

*Cosmic swampbrain:* Two cases of subjects with intrinsically identical brains, but where (a) one is a normal Earthly subject with normally representing sensory states and (b) the other is a ‘cosmic swampbrain’ that has randomly assembled itself along with supporting vat in outer space and so arguably isn’t representing anything at all. (Papineau, 2014, p. 8)

I can imagine a few different kinds of move that a relational theorist could try to make in response to the Perfect Chaotic Hallucination scenario:

(a) One might try to make trouble for the assumption that the floating-in-a-void brain really is behaving in exactly the same physical manner as in the perceptual case. After all, presumably at some highly microscopic scale or level of magnification there are, by hypothesis, supremely unlikely quantum events taking place somewhere inside the neurons of the brain in the chaos scenario that are absent in the perceptual case. For in the perceptual case, energy is transferred from some external stimulus, rather than generated by sheer random, uncaused quantum fluke. And after all, there have been theorists (e.g. Penrose, 1989; Lockwood, 2003) who have argued that quantum-level activity is somehow essential to explaining consciousness.

(b) One might try to deny that such a chaotic case is really possible — or, less ambitiously, one might try to insist that such a ‘merely logical’ possibility is not something that a theory of consciousness needs to deal with. Compare: in response to Davidson’s Swampman example, a friend of the teleo-functional theory of content might protest that they are giving a theory of content *as it naturally occurs in the real world*. It does not aim or need to deal with merely logically possible cases.



(c) It remains theoretically open for the relational theorist to go disjunctivist *just* about these chaotic perfect hallucinations — i.e. to deny the CK assumption *only* for these chaotic kinds of perfect hallucination.

(d) One might try to turn the tables on the proponent of the chaos scenario:

If we are allowed to imagine that any external stimulus is absent and replaced by sheer random quantum events, why can't we imagine the same occurring inside the brain? For example, whatever neural parts or processes are alleged to provide the minimum supervenience base for the conscious experience, we imagine a scenario in which various arbitrary portions of the brain are missing but in which the activity of these missing, supposedly vital, portions is miraculously compensated for by sheer random quantum-level events. Why not just a single neuron floating in the void and firing in just the same pattern as in the perceptual case 'by sheer chance' with the rest of the brain's activity recreated by the random quantum flux? Indeed, why insist on the brain at all? Consider the following 'possible but massively unlikely' scenario — a human body floating in the void, but whose skull cavity is totally empty, no brain inside. Yet within this empty cranial space, by sheer chance, random quantum level events occur which just happen to recreate the required stimulation to all the muscles in the face and larynx and arms and hands, etc. so that the body produces fluent conversation and expressive gestures just as if it did possess a functioning brain.

Of course these are not meant as serious suggestions. I mention them merely to show that 'two can play at that game'. Anyone who proposes a minimum supervenience base for an experience — whether it is brain + external scene, or the whole brain alone, or just some sub-portion of the brain — faces a potential challenge of this form. We can always imagine that some of the proposed supervenience base is missing but due to miraculous random chance the absence, as it were, makes no difference to the continued normal functioning of the remainder. So some other principled reason has to be given for why one's preferred supervenience base is just the right size — not too small that it fails to support consciousness, not so large as to include inessential elements.

Perhaps in the end none of these four are adequate responses.

But the important *dialectical* point to make about these perfect chaotic cases is that they seem to pose *just as much of a problem* for stan-

dard representational theories of experience.<sup>15</sup> If there is just nothing else in the universe that is causally related to the lonely brain, floating in the void, enjoying random, massively unlikely, uncaused neural stimulations, then there is presumably nothing that the brain's states could *represent*, at least on standard theories of representational content. And indeed a number of prominent representational theorists have been prepared to bite the bullet here and insist that an intrinsic duplicate of a normal perceiver's brain, but which lacks any causal contact or history with anything else external to it, would indeed lack representational states and hence would not have any phenomenal consciousness (e.g. Dretske, 1996; Tye, 1995; Lycan, 2001; Byrne and Tye, 2006).

In summary: the only kind of hallucinations that really put pressure on the relational theorist to accept the CK assumption are *perfect* hallucinations (neurally matching, hence phenomenally indistinguishable). I have suggested that someone who is, for whatever reason, attracted to the relational view in the case of perceptions can simply *accept* the CK assumption just for these perfect hallucinations (though not for common-or-garden, non-matching, real-life hallucinations), by claiming that such 'hallucinations' would in fact be relational experiences of something mind-independent. It should be admitted that this strategy seems to founder when faced with *perfect chaotic* hallucinations — where, by hypothesis, there *is* no candidate mind-independent feature to be related to. But this sort of perfect chaotic hallucination case is just as much of a problem for the main rival position — the representational theory. Moreover, it is not clear how much dialectical weight our intuitions about such a far-fetched case should be allowed to bear. David Papineau writes:

Many will take it to be simply obvious that the contrasting pairs of subjects must be consciously identical, given their intrinsic identity. Perhaps we should not be too quick to side with this intuition in the case of cosmic swampbrain. It is a strange and unfamiliar case, and maybe little weight should be accorded to intuitions about such far-off possibilities. (Papineau, 2014, p. 20)

So even *without* appealing to the subtle strategies of disjunctivism, perfect hallucinations — supposedly the worst-case scenario for naïve

---

[15] To be fair: a representational theory does seem to fare better with the case where a brain that was previously housed in a normal human subject, then somehow finds itself having random uncaused perfect hallucinations. In this sort of case the brain's states during the hallucination could still be ascribed representational content. It is only with a *genuine* 'swampbrain', which for its whole life has lacked causal contact with anything external to it, that standard representational theories are in the same boat as the relational theory.

realism — should cause no more anxiety for the relational theorist than is felt by her representational rivals.

And even if one is not interested in defending the relational theory *per se*, the fact that it is at least a *live option* to endorse the CK assumption, but to advance an 'argument from perception' rather than the argument from hallucination, quite significantly alters the dialectic of the traditional problem of perception and the role/importance of hallucinations. For if it is a live option to advance an argument from perception rather than from hallucination, then the traditional opposition of 'direct' vs. 'indirect' theories of perceptual experience *cannot be resolved* by appeal to the possibility of perfectly matching hallucinations nor by defence of the CK assumption. We would need *independent* reasons to prefer an internalist theory of perceptual consciousness<sup>16</sup> to an essentially world-involving account of such experiences.<sup>17</sup>

### References

- Alston, W.P. (1999) Back to the theory of appearing, *Noûs*, **33** (Supplement: Philosophical Perspectives, 13, Epistemology), pp. 181–203.
- Bouwsma, O.K. (1949) Descartes' evil genius, *Philosophical Review*, **58** (2), pp. 141–151.
- Brewer, B. (2006) Perception and content, *European Journal of Philosophy*, **14** (2), pp. 165–181.
- Brewer, B. (2011) *Perception and Its Objects*, Oxford: Oxford University Press.
- Butchvarov, P. (1994) Direct realism without materialism, *Midwest Studies in Philosophy*, **XIX**, pp. 1–21.
- Button, T. (2013) *The Limits of Realism*, Oxford: Oxford University Press.
- Byrne, A. & Tye, M. (2006) Qualia ain't in the head, *Noûs*, **40**, pp. 241–255.
- Campbell, J. (2002) *Reference and Consciousness*, Oxford: Oxford University Press.
- Campbell, J. (2009) Consciousness and reference, in McLaughlin, B., Beckermann, A. & Walter, S. (eds.) *Oxford Handbook of Philosophy of Mind*, pp. 648–662, Oxford: Oxford University Press.
- Chalmers, D. (2005) The Matrix as metaphysics, from the philosophy section of the official Matrix website, [Online], [http://whatisthematrix.warnerbros.com/rl\\_cmp/phi.html](http://whatisthematrix.warnerbros.com/rl_cmp/phi.html). Also in Grau, C. (ed.) *Philosophers Explore the Matrix*, Oxford: Oxford University Press.
- Davidson, D. (1986) A coherence theory of truth and knowledge, in Lepore, E. (ed.) *Truth and Interpretation, Perspectives on the Philosophy of Donald Davidson*, Oxford: Blackwell.
- Dretske, F. (1996) Phenomenal externalism, in Villanueva, E. (ed.) *Philosophical Issues 7, Perception*, Atascadero, CA: Ridgeview Publishing.

[16] A recent argument against phenomenal externalism that does not appeal to traditional arguments from hallucination is Pautz (2013).

[17] An ancestral version of this paper was presented at King's College London — my thanks to the audience on that occasion for helpful questions and comments. Thanks also to David Papineau and Jonathan Knowles for conversation and comments.

- Fish, W. (2009) *Perception, Hallucination, and Illusion*, Oxford: Oxford University Press.
- Hellie, B. (2010) An externalist's guide to inner experience, in Nanay, B. (ed.) *Perceiving the World*, Oxford: Oxford University Press.
- Hinton, J.M. (1967) Visual experiences, *Mind*, **76**, pp. 217–227.
- Hinton, J.M. (1973) *Experiences: An Inquiry Into Some Ambiguities*, Oxford: Clarendon Press.
- Hubel, D.H. (1988) *Eye, Brain and Vision*, New York: Scientific American Library.
- Hurley, S. (1998) *Consciousness in Action*, Cambridge, MA: Harvard University Press.
- Ings, S. (2007) *The Eye: A Natural History*, London: Bloomsbury.
- Johnston, M. (2004) The obscure object of hallucination, *Philosophical Studies*, **103**, pp. 113–183.
- Johnston, M. (2006) Better than mere knowledge? The function of sensory awareness, in Gendler, T. & Hawthorne, J. (eds.) *Perceptual Experience*, Oxford: Oxford University Press.
- Kennedy, M. (2007) Visual awareness of properties, *Philosophy and Phenomenological Research*, **LXXV** (2), pp. 298–325.
- Langsam, H. (1997) The theory of appearing defended, *Philosophical Studies*, **87**, pp. 33–59.
- Langsam, H. (2011) *The Wonder of Consciousness*, Cambridge, MA: MIT Press.
- Lockwood, M. (2003) Consciousness and the quantum world: Putting qualia on the map, in Smith, Q. & Jovic, A. (eds.) *Consciousness: New Philosophical Perspectives*, Oxford: Oxford University Press.
- Logue, H. (2011) The skeptic and the naïve realist, *Philosophical Issues*, **21** (1), pp. 268–288.
- Logue, H. (2012) What should the naïve realist say about total hallucinations?, *Philosophical Perspectives*, **26** (1), pp. 173–199.
- Lycan, W. (2001) The case for phenomenal externalism, in Tomberlin, J.E. (ed.) *Philosophical Perspectives*, **15**, pp. 17–35, Atascadero, CA: Ridgeview Publishing.
- Martin, M.G.F. (1998) Setting things before the mind, in O'Hear, A. (ed.) *Contemporary Issues in the Philosophy of Mind*, pp. 157–179, Cambridge: Cambridge University Press.
- Martin, M.G.F. (2002) The transparency of experience, *Mind and Language*, **17**, pp. 376–425.
- Martin, M.G.F. (2003) Particular thoughts and singular thought, in O'Hear, A. (ed.) *Thought and Language*, pp. 173–214, Cambridge: Cambridge University Press.
- Martin, M.G.F. (2004) The limits of self-awareness, *Philosophical Studies*, **120**, pp. 37–89.
- Martin, M.G.F. (2006) On being alienated, in Gendler, T.S. & Hawthorne, J. (eds.) *Perceptual Experience*, pp. 354–410, Oxford: Oxford University Press.
- McDowell, J. (1982) Criteria, defeasibility and knowledge, *Proceedings of the British Academy*, **68**, pp. 455–479.
- McDowell, J. (1994) *Mind and World*, Cambridge, MA: Harvard University Press.
- Noë, A. (2004) *Action in Perception*, Cambridge, MA: MIT Press.
- O'Regan, J.K. (2011) *Why Red Doesn't Sound Like a Bell: Understanding the Feel of Consciousness*, Oxford: Oxford University Press.
- O'Shaughnessy, B. (1984) Seeing the light, *Proceedings of the Aristotelian Society*, **85**, pp. 193–218.

- Papineau, D. (2014) Sensory experience and representational properties, *Proceedings of the Aristotelian Society*, **114** (1), pp. 1–33.
- Pautz, A. (2013) The real trouble for phenomenal externalists, in Brown, R. (ed.) *Consciousness Inside and Out: Phenomenology, Neuroscience and the Nature of Experience*, Springer: Dordrecht.
- Penrose, R. (1989) *The Emperor's New Mind*, Oxford: Oxford University Press.
- Putnam, H. (1981) *Reason, Truth and History*, Cambridge: Cambridge University Press.
- Robinson, H. (1985) The general form of the argument for Berkeleyan idealism, in Foster, J. & Robinson, H. (ed.) *Essays on Berkeley*, Oxford: Clarendon Press.
- Robinson, H. (1994) *Perception*, London: Routledge.
- Smith, A.D. (2002) *The Problem of Perception*, Cambridge, MA: Harvard University Press.
- Snowdon, P. (1980) Perception, vision and causation, *Proceedings of the Aristotelian Society*, **81**.
- Snowdon, P. (2005) Some reflections on an argument from hallucination, *Philosophical Topics*, **33** (1), pp. 285–306.
- Thompson, E. & Varela, F.J. (2001) Radical embodiment: Neural dynamics and consciousness, *Trends in Cognitive Sciences*, **5**, pp. 418–425.
- Tye, M. (1995) *Ten Problems of Consciousness*, Cambridge, MA: MIT Press.

Paper received: February 2014; revised July 2014.