Modelling Empty Representations: The Case of Computational Models of Hallucination

Marcin Miłkowski

Abstract I argue that there are no plausible non-representational explanations of episodes of hallucination. To make the discussion more specific, I focus on visual hallucinations in Charles Bonnet syndrome. I claim that the character of such hallucinatory experiences cannot be explained away non-representationally, for they cannot be taken as simple failures of cognizing or as failures of contact with external reality—such failures being the only genuinely non-representational explanations of hallucinations and cognitive errors in general. I briefly introduce a recent computational model of hallucination, which relies on generative models in the brain, and argue that the model is a prime example of a representational explanation referring to representational mechanisms. The notion of the representational mechanism is elucidated, and it is argued that hallucinations—and other kinds of representations—cannot be exorcised from the cognitive sciences.

1 Introduction

Contemporary discussion on the use of the notion of representation for explanatory purposes in cognitive science focusses mainly on the controversy between representationalism and anti-representationalism. Representationalism claims that cognitive representations are at least sometimes relevant for cognition, explanatorily and causally. There are numerous criticisms of positing representations in a hasty manner. Critics point out that some representations are not supposed to fulfil any particular representational role [1], can be treated merely instrumentally [2, 3], or lack naturalistic credentials, especially when it comes to showing how they can have satisfaction conditions [4]. Although I argue for representationalism, the case for representationalism requires that we consider relevant alternatives. Hence, discussing anti-representational alternatives is important. Anti-representationalism claims that, at least in some domains, positing representations that have satisfaction

M. Miłkowski (✉)
Institute of Philosophy and Sociology, Polish Academy of Sciences, Warszawa, Poland
e-mail: mmilkows@ifispan.waw.pl

G. Dodig-Crnkovic and R. Giovagnoli (eds.), *Representation and Reality in Humans, Other Living Organisms and Intelligent Machines*, Studies in Applied Philosophy, Epistemology and Rational Ethics 28, DOI 10.1007/978-3-319-43784-2_2

conditions is unnecessary (however, they rarely deny all kinds of representations; see [5] for more detail).

In this paper, I argue that there are no relevant non-representational explanations of episodes of hallucination. These are usually explained by positing empty representations, or representations that have no actual reference. To make the discussion more specific, I focus on visual hallucinations occurring in people with Charles Bonnet syndrome. I claim that the character of such hallucinatory experiences cannot be explained away in a non-representational manner, for they cannot be taken as simple failures of cognizing or as failures of contact with external reality —such failures being the only genuinely non-representational explanations of hallucinations and cognitive errors in general. Then I briefly introduce a recent computational model of hallucination, which relies on generative models in the brain, and argue that the model is a prime example of a representational explanation referring to representational mechanisms for which there is simply no non-representational alternative. The notion of the representational mechanism is elucidated, and it is argued that hallucinations—and other kinds of empty representations—cannot be exorcised from the cognitive sciences.

2 Charles Bonnet Syndrome and Representations

There are numerous kinds of hallucinatory episodes, and the number of studies on these phenomena is virtually countless (for a comprehensive review, see [6]; for an accessible introduction, see [7]). To make my discussion more specific, in this paper I discuss Charles Bonnet syndrome (henceforth: CBS). CBS is usually a complex visual hallucination in people with some impairment of vision. CBS hallucinations are frequently bizarre in nature: they include figures in elaborate costumes, human beings of non-natural size, fantastic creatures, or extreme colours, which may partly overlap with real visual perception. Yet there is nothing to which these hallucinations correspond; in other words, they are not visual illusions, even if, owing to the impairment of vision, these hallucinations occur at the same time as various other abnormal phenomena.

Importantly, CBS subjects usually (but not always) realize that the visual episodes they are experiencing are not real, and sometime may even think that their unusual perception is a result of their hallucinations. One reason it is easy for them to understand that they are experiencing a hallucination is that CBS is merely visual, and there is usually discrepancy with auditory or tactile perception.

There are two main competing neurophysiological explanations of CBS. The first classifies CBS as release hallucinations, i.e. ‘hallucinations mediated by spontaneous electrophysiological activity originating from subcortical brain areas such as the thalamus, the pedunculus cerebri and the limbic system’ [6, p. 93]. Another attributes CBS to increased excitability of the visual pathways or the visual cortex, owing to a lack of inhibitory afferent impulses. Brain regions considered capable of mediating spontaneous visual percepts include the retina, the lateral geniculate nucleus, the primary visual cortex and the visual association cortex [6, p. 94].

These features of CBS syndrome make it difficult to explain away in a non-representational manner, because the content of hallucinations is apparently decoupled from perception. At the same time, these hallucinatory representations thereby satisfy the requirement proposed by several theorists as particularly important for representations, namely decouplability or detachment of the representation from its target [8, 9]. While Andy Clark does not see decouplability as a necessary feature of representation, it is reliably present in such hallucinations. In this case, detachment may actually justify the use of representational talk. Note that a generic move recommended by proponents of anti-representationalism in response to Clark's decouplability argument, namely to introduce time-extended perceptual processes instead of decoupling [10], will not work for CBS. Simply put, there was no point of contact between a hallucinated entity—for example, a fantastic creature—and the CBS subject in the past, so extending the perceptual process still cannot reach the hallucinated creature.

This, however, is not enough to show that representations cannot be avoided. First, anti-representationalism can appeal to an empiricist argument that all ideas stem from sensual impressions; in a Humean manner, they can appeal to elementary building blocks of perception that are recombined to build a complex hallucination. For example, you could combine an elementary perception of a white horse, and a perception of a horn, and get an image of the unicorn. The problem of course is that the 'solution' sounds fairly incompatible with the general approach of most anti-representationalists, which is the dynamical account of cognition. Dynamicists seem to embrace the claim that there are no elementary primitives of concepts or perception at all. The assumption of elementary building blocks in the traditional symbolic approach (or cognitivism) was criticized by Dreyfus [11]. In other words, the price of the classical empiricist move may be too high for most anti-representationalists to pay. But the recombination approach, with some time-extended perception, might seem to work if one believes that there is a credible empiricist answer to Berkeley's puzzle of how one can imagine things one has never seen. While CBS subjects have not perceived miniature people, they have perceived people, and they have perceived small entities, so it is possible for them to combine the two.

For the sake of argument, let me suppose that some solution like this might be put to work. However, the perceptual *recombination* would still seem to produce a representation—something semantic, something about something else—even if particular elementary perceptions are held to be non-representational points of contact with reality. Their recombination is additionally non-veridical for the CBS subject. Simply put, it is easier to eliminate veridical perception and replace it with 'direct' or 'representationally unmediated' contact with reality than to replace content-rich hallucinations with time-extended contact with reality. It is the recombination, if it actually occurs, that drives a wedge between the representation and reality. Hence, the recombination 'solution' is merely verbal—representation has been merely rebranded as recombined perception but retains the essential features of representation: aboutness and satisfaction conditions (which are never satisfied in the case of hallucinations).

But there is another option still open for anti-representationalists: to apply a neo-Gibsonian analysis of errors in cognition [12]. James J. Gibson, the founder of ecological psychology, insisted that hallucinations differ from imaginations in that they are passively experienced, and from perceptions in that they are not 'made of the same stuff' [13], p. 425). Namely, '*a person can always tell the difference between a mental image and a percept when a perceptual system is active over time*' (ibid., italics in original).[1] His general rule for distinguishing perception is as follows:

> Whenever adjustment of the perceptual organs yields a corresponding change of stimulation there exists an external source of stimulation and one is *perceiving*. Whenever adjustment of the perceptual organs yields *no* corresponding change of stimulation there exists no external source of stimulation and one is imagining, dreaming or hallucinating. [13], p. 426)

Gibson may be roughly right about some kinds of hallucination,[2] including most cases of CBS: subjects usually discover CBS hallucinations just because they are not accompanied by proper adjustments of auditory or tactile stimulation. At the same time, in another passage, he seems to contradict himself by claiming:

> One perceptual system does not *validate* another. Seeing and touching are two ways of getting much the same information about the world. [14, pp. 257–8]

In CBS, it is exactly the case that discrepancy between different perceptual systems allows hallucinators to understand that their visual experiences are not entirely veridical; and touching does provide *different* information to the subject. All in all, Gibson seems to ignore the complex role of multimodal integration, for example, the role of the vestibular system in seeing [15].

However, for our purposes, what is important is the question of whether hallucinations are to be understood as representations. To this question, unfortunately, Gibson gives no clear answer, but his theory is usually interpreted as a form of direct realism (however, see [16] for a representational reading of Gibson). In direct realism, error is understood as a failure to cognize in some way, or a failure to cognize that one fails to cognize [12]. This latter, hierarchical solution is an account of false beliefs and similar errors, so it should work for bizarre CBS hallucinations as well. What this account claims, basically, is that people who have CBS, during hallucinatory episodes, fail to cognize that they do not have perceptual states. But the truth is exactly the opposite, and Gibson himself stresses that one discovers that

[1] Note that this is an extreme empirical claim, and a false one, and CBS subjects can not only take hallucination to be veridical but also sometimes mistake veridical perception for hallucination [7]. Such a mistake might go easily undetected forever.

[2] Only in some hallucinations can a person tell the difference between the hallucination and perception. There might be scenic, multimodal and persistent hallucinations [6], which can be confused with perceptions by a subject.

hallucinations are not perceptions. They usually *know* that they are hallucinating, so this is not a failure to cognize that the subject fails to cognize. Neither can the contents of their hallucination be naturally accounted for in terms of a simple failure to perceive, which may happen when you cannot find your keys in a drawer. It cannot be accounted for by stipulating a hierarchy of failures to cognize, as there is no explanation of the rich *content* of bizarre CBS hallucinations (subjects stress the richness as striking; cf. [7, pp. 4–5]). For this reason, a neo-Gibsonian alternative is doomed to fail.

A third explanation of hallucination in broadly non-representational terms appeals to a sense of real presence, which is supposedly present during perception or hallucination, and absent in imagining [17]. The experienced presence of objects, even when they are occluded but accessible on further exploration, is supposed to help explain hallucination's content non-representationally. The illusory presence of hallucinatory objects is supposed to stem from the skilful exercise of perceptual skills: "The hallucinator acts out the same sensorimotor repertoire as the perceiver" [17, p. 249].[3] But this does not explain why CBS subjects describe their hallucinations as involving bizarre figures. There are no real sensorimotor skills that involved the hallucinator's previous perception of bizarre figures (and if you suppose that bizarre figures are recombinations of previous skills, you presuppose experiential primitives that the enactive approach rejects explicitly). The sensorimotor account seems, rather, to presuppose intentionality of these visual episodes:

> [T]he approach actually makes it easier to envisage brain mechanisms that engender convincing sensory experiences without any sensory input, since the sensation of richness and presence and ongoingness can be produced in the absence of sensory input merely by the brain being in a state such that the dreamer implicitly "supposes" (in point of fact incorrectly) that if the eyes were to move, say, they would encounter more detail. [18, pp. 66]

The word 'supposes' is obviously intentional, and it is not easy to eliminate or explain it away from this passage. For this reason, the sensomotoric account is actually representational, and cannot be considered an alternative to the representational explanation. Any talk of sensory experiences being about anything implies representationalism.

Critics of representationalism sometimes complain that illusions, hallucinations and misperceptions are the focus of representational explanations in psychology, and that correct perception could be understood in terms of contact. A milder non-representational position might be that of disjunctivism,[4] or the claim that perceptual processes are essentially different from misperceptual processes. Of course, in hallucinations, perceptual processes are different in that hallucinations are

[3] There are further problems with the sensorimotor account of CBS; it may be accompanied with partial or total paralysis, which makes any exercise of motor skills simply impossible.

[4] Note that there are representationalist versions of disjunctivism. I discuss only a possibility of disjunctivist anti representationalism above.

not perceptions. But that's trivially true.[5] The problem for the disjunctivist is that the underlying brain and bodily machinery (including active exploration) recruited for perception are the same [20]; it is not just that the subjective experience may be the same. Additionally, the difference between perception and hallucination is likely a matter of degree rather than of quality [21]. For the agent and its subpersonal processes, until discrepancy is detected between various sources of information, hallucinatory episodes may seem perceptual. When discrepancy is detected, though, non-veridical representations are considered to be such, so that the cognitive process is actually different but shares a common core with the standard perceptual process, as can be witnessed in the model analyzed below. So while there is a grain of truth that these processes differ for the subject (at least when the subject is not delusional or not confabulating), it is not the case that the differences can substantiate a non-representational position. But they may seem the same for the subject, in which case the disjunctivist claim has no explanatory role to play at all. Non-veridical hallucinations can explain behaviour just as well as veridical perception can; when the deluded subject takes hallucinations to be real, disjunctivism has yet another fact to explain: why a completely different process leads to the same behaviour as perception would. In brief, the presupposition that there is a single process underlying hallucination and perception is more parsimonious than a disjunctivist proposal that adds unnecessary complexity.

All three anti-representational alternative explanations of CBS therefore fail, and disjunctivism is unnecessarily complex. I know of no successful non-representational attempt to explain the content of hallucinations away, but my argument is based only on negative induction. Representational explanations seem to be much more plausible. Let me turn to a computational model of CBS, which will be used to offer a representationalist explanation.

3 Computational Modelling of Hallucination

Existing computational models of hallucination are all representational. In general, recent work on predictive coding in the brain [22–24] includes also some suggestions regarding hallucinations and psychosis [25]. Here, I will focus on a similar but non-Bayesian model of CBS [26], according to which there are hierarchical, generative models in the brain that control perception. These models work homeostatically; i.e. the bias varies proportionally to input strength (as such, the model assumes a neurophysiological explanation close to the second type mentioned in Sect. 2, which appeals to the increased excitability of the visual system). There are two pieces of empirical evidence that support the bias hypothesis. One is synaptic

[5] Only representationalists who endorse methodological solipsism [19] might deny this, as they cannot include the relationship with the environment in their explanations. But methodological solipsism is long dead. Content does *not* locally supervene on the brain alone.

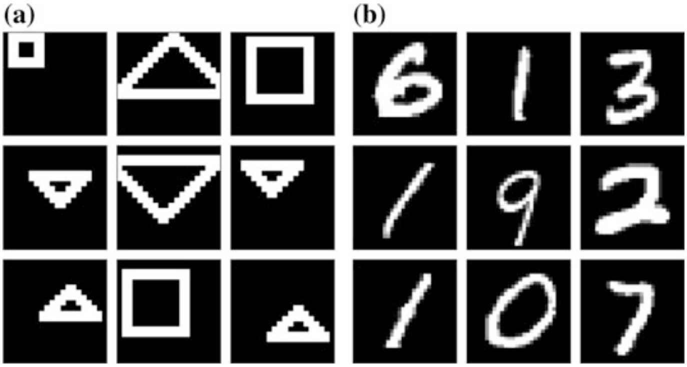

Fig. 1 **a** A custom data set of simple shapes at various positions. **b** The MNIST data set of handwritten digits, a standard benchmark in machine learning [26] doi:10.1371/journal.pcbi.1003134.g003

scaling, which is a change in synaptic efficacy that is thought to affect all synapses in a neuron together, keeping their relative strengths intact. Synaptic scaling is known to occur in the neural system. Another is the fact that a neuron's intrinsic excitability can be regulated by changing the distribution of ion channels in its membrane.

The model has been implemented by a deep Boltzman machine (DBM), which is a connectionist architecture supporting deep learning. This means that there is no need for biologically implausible learning algorithms such as backpropagation [27]; the machine can find the model from the data. As the name implies, DBM is also probabilistic, but not classically Bayesian.

The network contains three hidden layers, and has been trained to recognize input figures. Notably, these input figures were not geometrically complex (simple figures and numbers, see Fig. 1), which might seem a huge idealization vis-à-vis the real phenomenon. However, in mild cases of CBS, hallucinations may consist in simpler shapes; at the same time, the model is unable to reconstruct complex, bizarre images typical of CBS. After the training, the input was clipped.

After the loss of input, owing to the homeostatic nature of the network, the output was restored. This happened as early as after 20 cycles of operation. But this result is not the only reason to believe that the model is correct. There is also a structural match of the model to the neural process (in other words, the model can be said to be structurally valid in terms of [28]).[6] The evidence for the structural match is that clamping the first hidden layer to zeros stopped hallucinations appearing homeostatically, and that this same behaviour has been observed in humans; namely applying transcranial magnetic stimulation (TMS) to early areas

[6] For this reason, the model is a prime example of the computational mechanistic explanation: the computational process is thought to correspond to the actual neural process (for more on computational explanation and mechanism, see [29]).

leads to a temporary cessation of the hallucinations. Additionally, the hallucinatory episodes are more likely to appear in states of drowsiness or low arousal. The authors believe that this is related to acetylcholine (Ach) dysfunction; and this is modelled as the balance factor between the feedforward and feedback flows of information.

All in all, while this model is an idealization it nevertheless suggests that interfering with cortical homeostatic mechanisms might prevent the emergence of hallucinations in CBS. This means that it provides novel predictions. DBM also uses top-down interactions during inference, not just learning, which fulfils an important requirement for modelling the role of hierarchical bottom-up and top-down processing in hallucination.

What are the specific representational features of this model? First of all, DBM learns the patterns in the input data instead of memorizing them: "rather than just memorizing patterns, BMs can learn internal representations of sensory data" [26]. These patterns are amplified by bias owing to homeostasis, which leads to hallucinations in drowsiness or low arousal. The model, however, does not explain why subjects realize that the experienced images are hallucinatory rather than perceptual. This is not included in the intended scope of the model—i.e. the model simply does not represent such perceptual evaluation processes at all. In other words, it is not a model of all the relevant processes responsible for CBS hallucinations but of one critical phenomenon, namely the mechanism of varied sensory excitability that may lead to visual hallucinations in case of no visual input.

4 Representational Mechanisms and Hallucinations

The discussion so far has not referred to any substantial theory of representation; I only mentioned in passing that there are two features that have been considered essential: decouplability and satisfaction conditions. Although I do not think that potential decoupling is necessary for representation as such, it is certainly one of the essential properties of hallucinatory representations. But more needs to be said here.

I do not assume that all computational models are representational; on the contrary, I think that proving that a computational model is a model of representation is difficult (see [29], Chap. 4). There are obvious necessary conditions that the DBM satisfies, such as having some tokens (in the generative model in the DBM) that play the role of representation vehicles.[7] This is satisfied just because DBM is a computational mechanism, and such mechanisms process information, which means that they manipulate certain information vehicles. Another satisfied condition is that we talk of representation targets in this case: the simple figures are not just vehicles of information; they are supposed to refer to perceptually given

[7] Note that I do not claim that the input layer contains any representational vehicles at all. They are merely input information, not representation. For more on this distinction, see [30, 29].

entities, while there are no relevant perceptual inputs in reality, just as in CBS episodes. Representational targets would be found in reality were this a perceptual process—if only in a person's visual field, there would be entities with visual characteristics furnished by the visual system. In traditional terminology, one could say that there is some content of visual representations, or some intension of these representations. In the DBM, these characteristics are represented as the configuration of the network output layer.

These conditions—referring to targets (which may fail), having characteristics, having satisfaction conditions, containing vehicles of information—do not suffice for something to count as a representation, however. In general, without any reference to agency and evaluation of the representation, it is still unclear whether these vehicles of information are *mental* representations at all [31]. I suggest that one approach to this problem is to apply the neo-mechanistic account of explanation to see what is lacking in the model. The advantage of this approach is that the mechanism always posits mechanisms in the context of other mechanisms, so that representation won't float freely without being part of an organized cognitive system.

Before I systematically introduce the notion of a representational mechanism, the notion of a mechanism needs to be elucidated. While definitions of mechanisms offered by various authors accentuate different aspects, the main idea can be summarized as follows: mechanisms are complex structures, involving organized components and interacting processes (or activities) that contribute jointly to a capacity of the structure. Mechanistic explanation is a species of causal explanation, and interactions of components are framed in causal terms (for the main proponents, see [32–35]).

What is important is that there are no mechanisms per se; there are only *mechanisms of something*. In other words, it is critically important to specify the phenomenon to be explained, or the capacity of the mechanism. The mechanism is defined by its capacity, or individuated with respect to its capacities. For example, the capacity of the mechanism of a mousetrap is to catch mice. The mousetrap may be physically connected to a table, but as long as this connection makes no difference to its exercising the capacity to catch mice, the connection to the table does not make the table a component of the mechanism. Briefly, only those activities and components that contribute to the mechanism's exercising its capacity count as belonging to the mechanism. Hence, mere spatiotemporal co-occurrence does not make anything a component of the mechanism; the notion of the mechanism, which is a spatiotemporal entity, is defined via its capacity (or function). For this reason, it is also theory dependent [36, 37].

A representational mechanism will be one that has the above-mentioned necessary properties and that makes the information contained in the vehicles available to the cognitive system by modifying the system's readiness to act. The notion of information used here requires more elucidation. DBM states can be treated as states of the physical medium; and as long as the medium has different physical states, as distinguished by the machinery of the DBM, it contains information. In this case, one can call this information *structural* (following [38]). Minimally, the physical medium needs to have at least one degree of freedom distinguished by the whole system. Now, the structural information becomes *semantic* as soon as it modifies the system's readiness to act, more precisely as soon as the conditional probabilities of actions of the system are changed accordingly when the information vehicles change [38].

However, mere semantic information in the above sense does *not* make a representation. The representation needs to play a representational role in the system, and for that it needs to have satisfaction conditions—truth or veridicality conditions for descriptive representations, and success or failure conditions for directive representations (some representations may have both kinds of satisfaction conditions, in particular pushmi-pullyu representations sensu Millikan; cf. [39]). These satisfaction conditions, additionally, have to be evaluable by the cognitive system itself. Only then may the contents be said to be available for the system. Such evaluation requires more than negative feedback in the information-processing mechanism: negative feedback might merely modify the system's input value. What is required instead is that the error is detected by the system; note that this is what CBS subjects normally do, although error detection was not included in the DBM model. For this reason, the DBM is not a complete model of a representational mechanism. The idea that system-detectable error gives rise to genuine representationality is by no means new (for an extended argument, see [30]).

By framing the capacity of the representational mechanism as the modification of the readiness to act, based on the information available to the cognitive system, this framework is committed to a claim that representation is essentially action-oriented. However, this orientation does not mean that all representations directly activate effectors of the system, or that representation simply controls the motor activity of the system. There might be content that is not exploited in action; what is altered is just the readiness to act. The notion of action is to be understood liberally to include cognitive operations.

Summing up, we can define the complex capacity of the representational mechanism as follows:

1. Having information vehicles that modify the cognitive system's readiness to act
2. Referring to the target (if any) of the representation
3. Identifying the characteristics of the target
4. Having satisfaction conditions based on these characteristics
5. Evaluating the epistemic value of information, or checking the satisfaction conditions

In research practice in cognitive sciences, a variety of different representational mechanisms have been posited, and the current proposal is neutral with regard to empirical questions such as whether there might exist mechanisms that only deal with the "language of thought" (usually dubbed "symbolic"), or whether there are also imagistic formats of representation; it does not decide the nature of concepts, or non-conceptual thinking, either. In other words, the account is very liberal. However, just because it requires more than negative feedback, it will not license representations in a thermostat or, Watt governor or any other simple control system [40]. It also will not license content attributions to creatures capable only of taxes, such as phonotaxis, because taxes do not imply any satisfaction conditions available to the system itself [29, 41]. But it does ascribe content to theories that talk of sensomotoric contingencies as modifying readiness to act [42].[8] This, contra Hutto [43], is not a disadvantage of the sensomotoric account of vision. Hutto is right that the account is merely *verbally* non-representational, if we apply his minimal understanding of representation as having content with satisfaction conditions. Noë [17], however, presupposes that the representationalist claim is that it is perceptual experience *as a whole* that is representational and fully detailed. But in reality there are many other options available for a representationalist, and representations can be sketchy and highly action-oriented. Experience can be considered an ongoing interaction that creates multiple representational states; it may not be reducible to any particular representation among them.

Indeed, it need not be presupposed that the goal of perception is to create detailed, rich visual representations [44]. But saying that the goal of perception is not to build detailed percepts but to drive further exploration and modify the readiness to act is not the same as saying that vision does not require representation or that percepts do not exist (*contra* [45]). The latter does *not* follow from the former. However, adversaries of the representational account of perception seem to presuppose that this account implies a kind of detailed mental image, sense data and so forth, and that such entities are end products of perception. But it does not imply anything like that. Why should it? It implies much less; that there are entities that are perceived (targets), that they have perceived properties (characteristics), that they can be veridical or not, and that they can be evaluated, sometimes successfully, by the cognitive system, just by acting and exploring further.

Radical enactivism proposes that the notion of perceptual representation can be replaced with one of contact [4, 17]. In this respect, it shares the presupposition of crude causal theories of reference.[9] However, under the assumption that content is reducible to causal contact, hallucinations or misperceptions could not exist, as they are not *about* the entities that caused them. They are non-veridical or at least not entirely veridical (some of their contents may match reality). Hutto and Myin are right in saying that neither causation, correlation, or similarity constitute content; of course falsity would be impossible under such a model of representation. But hallucinations are about something, and they are false, so they do have satisfaction conditions, which can be known to the hallucinators. So how do such satisfaction conditions arise? For Hutto and Myin, this is the hard problem of content, which presumably cannot be solved. They think that there is no naturalistically kosher explanation of how content with satisfaction conditions may emerge from

[8] O'Regan and Noë even appeal to MacKay's [38] theory of information in their account of vision.

[9] Tom Froese suggested in his review of the previous version of this paper that direct realism does not share this trouble, as the world "directly shapes experience like a mold shapes clay". This is still causation if anything is, and direct realism faces the same troubles as crude causal theories of reference [46].

non-cultural and non-social processes. In particular, they think that there is no naturalistic explanation of the emergence of content from information (as constituted by causation, correlation, similarity, or learning).

Before I go on, it is important to note that the main argument for anti-representationalism given by Hutto and Myin [4] is a striking case of a straw man. Beside the crudest version of the causal account of reference—Fodor [47] half-jokingly attributes it to B.F. Skinner—no current account in naturalized semantics actually claims that content is constituted merely by a tracking (causation or covariation) or similarity relation. But content is *not* constituted by tracking or similarity. If a relation (in a strict logical sense) between the vehicle and the representation's target had constituted content, then false content would have been impossible. Relations obtain only when relata exist, and in the case of intentionality, the targets, or what the representation is about, might not exist.

However, Dretske, Millikan, Fodor and other proponents of naturalized semantics do not treat intentionality as a relation. For this reason, in their accounts, intentionality is not reduced to tracking or similarity relationships. First of all, the problem of the impossibility of falsehoods would reappear. In addition, we know that not all tracking or similarity relationships constitute mental representations. They are necessary but not sufficient for representation. For Dretske and Millikan, another crucial factor of content determination is the notion of teleological function; for Fodor, the important role is assigned to counterfactual considerations (in this paper, I barely touch upon the notion of function; see however [48] for a full account of a complex notion of observer-independent teleological function). Briefly, according to Dretske's account, a certain activation of neurons in the visual pathway has the function of indicating the properties of the perceived scene. In the case of biological dysfunction (such as in people with visual impairment), the visual system may still seem to indicate bizarre figures even though there is nothing in the visual field that corresponds to them. But then of course there is no real indication; the system uses the visual pathway *as though* it were indicating visual properties. The content is not determined by mere indication but by a *function* of indication.

One fact that is frequently missed in polemics against teleofunctional theories of content is that indication is for Dretske a basic form of predication. Let us see how Dretske defines functional meaning (meaning$_f$):

> (M_f) *d*'s being *G* means$_f$ that *w* is $F = d$'s function is to indicate the condition of *w*, and the way it performs this function is, in part, by indicating that *w* is *F* by its (*d*'s) being *G*. [49, p. 22]

Indication is truth-functional; a property *F* is ascribed to *w*, and this can be spelled out in basic logical terms as ascribing a predicate to a subject. Hence, indication has satisfaction conditions. At the same time, indication cannot be false; it cannot fail to indicate that *w* is *F*. To make this possible, Dretske makes falsehood asymmetrically dependent on truth by introducing the notion of function. The entity *d* has the function of indicating that *w* is *F,* but as soon as it malfunctions, the indication is false. But the content is not lost; if it were an indicator, it would truly indicate that *w* is *F*.

There might be various problems with Dretske's account of content, but it solves the hard problem of content—at least in principle. The satisfaction conditions are determined by the indication relation *cum* teleological function, and there is nothing non-naturalistic about the account. While various accounts of naturalized semantics differ in many regards, they usually recruit a similar solution. What is particularly interesting is that the solution does not treat truth and falsehood symmetrically: falsehood is dependent on truth but not vice versa. For example, an account closer in spirit to the account of representational mechanisms is the interactivist model [50, 51]. In interactivism, information relationships—such as those constituted by causation, correlation, or similarity—are recruited for action, and they are used to build indications of possible actions. These indications say that such-and-such an action would be successful in such-and-such circumstances, while the circumstances are determined, inter alia, by information relationships with the environment.[10]

To sum up, there are several ways that one could analyze the hallucinatory episodes simulated by the DBM model, and there is no real difficulty with solving the hard problem (see also [5]). On the contrary, as soon as the DBM model is framed in terms of representational mechanisms, the representational role of some states of the network becomes clear; these states have satisfaction conditions just because they fail to refer. The representational explanation given by the DBM model of CBS is not complete with regard to the requirements specified by the current proposal: the scope of the model does not include evaluation processes, as I have stressed several times, so it does not fully license representational explanations. However, it clearly conforms to the general scheme proposed here: it relies on vehicles of information, and they are about (non-existing) targets, identified via visual characteristics; in the verbal gloss on the model, researchers add that subjects are usually aware that these representations are not veridical, so they are evaluated as such, and for that, they need to have satisfaction conditions. Note that, even though I rephrased the description of the model to show that it involves representational mechanisms, I do not see a plausible way to rephrase it in a non-representational way, by supposing that there is a failure to cognize along neo-Gibsonian lines (all other non-representational explanations turn out to be only verbally non-representational, so they are not even candidates for paraphrasing). Put simply, a non-representational explanation of CBS episodes does not seem to be forthcoming at all.

[10] Note that the content that emerges first in the interactivist model is—just like the notion of affordance in Gibson [14]—egocentric, and it involves an indication about the agent. However, via a hierarchy of differentiations—information relationships recruited for action—it is supposed to provide other kinds of information. It is not entirely clear, however, if this model can supply allocentric representation, i.e. representation that does not relate immediately to a cognitive system or to its actions. In contrast, the account of representational mechanism does not claim that the basic form of content needs to be egocentric.

5 Conclusion

Real content does involve satisfaction conditions, and hallucinations have them. They fail to be veridical, and having them does not require one to speak a language. Hence, they are empirical counterexamples to the claim of radical enactivism, namely that rich content with satisfaction conditions requires language [4]. Additionally, there is currently no viable non-representational explanation of hallucinations, and offering one remains an open problem for non-representationalism.

Visual hallucinations of the kind experienced under CBS have now only representational explanations. This means that they involve the work of representational mechanisms that are responsible for the hallucinatory episodes. As for other perceptual impairments, an ideal explanatory text would contain all relevant causal factors, including the detail of neural mechanisms, relevant computational processes and the environmental context. The DBM model included epistemic evaluation only as a verbal gloss. Thus, it is definitely not an ideal explanation but a partial one.

Representations cannot be easily exorcised from the cognitive sciences, and there is no need to exorcise them unless one is ready to defend extreme reductionism. Explanations that involve representations are parsimonious and general; changing one's mental representations is the easiest way to intervene in one's own behavior; representations also play essential heuristic roles [52]. Put simply, representations are here to stay, and the news of representationalism's death was greatly exaggerated.

Acknowledgements The work on this paper was financed by the National Science Centre under the program OPUS, grant no. 2011/03/B/HS1/04563. The author wishes to thank reviewers, in particular Tom Froese, and other careful readers of the draft, in particular Joe Dewhurst, Krzysztof Dołęga, Gualtiero Piccinini and Paweł Zięba for their very helpful comments.

References

1. Ramsey, W.M.: Representation Reconsidered. Cambridge University Press, Cambridge (2007)
2. Chemero, A.: Anti-representationalism and the dynamical stance. Philos. Sci. **67**(4), 625–647 (2000)
3. Chemero, A.: Radical Embodied Cognitive Science. The MIT Press, Cambridge, Mass. (2009)
4. Hutto, D.D., Myin, E.: Radicalizing Enactivism: Basic Minds Without Content. MIT Press, Cambridge Mass. (2013)
5. Miłkowski, M.: The hard problem of content: solved (Long ago). Stud. Grammar, Logic Rhetoric **41**(54), 73–88 (2015)
6. Blom, J.: A Dictionary of Hallucinations. Springer, New York (2010)
7. Sacks, O.: Hallucinations. Alfred A. Knopf, New York (2012)
8. Clark, A.: Being There: Putting Brain, Body, and World Together Again. MIT Press, Cambridge, Mass. (1997)

9. Smith, B.C.: On the Origin of Objects. English. MIT Press, Cambridge, Mass. (1996)
10. Garzon, F.C.: Towards a general theory of antirepresentationalism. Br. J. Philos. Sci. **59**(3), 259–292 (2008)
11. Dreyfus, H.: What Computers Can't Do: A Critique of Artificial Reason. Harper & Row, New York (1972)
12. Rantzen, A.J.: Constructivism, Direct Realism and the Nature of Error. Theory & Psychology **3**(2), 147–171 (1993)
13. Gibson, J.J.: On the relation between hallucination and perception. Leonardo **3**(4), 425–427 (1970)
14. Gibson, J.J.: The Ecological Approach to Visual Perception. Psychology Press, Hove (1986)
15. Berthoz, A.: The Brain's Sense of Movement (G. Weiss, Trans.), Harvard University Press, Cambridge, Mass (2000)
16. Bickhard, M.H., Richie, D.M.: On the nature of representation: a case study of James Gibson's theory of perception. Praeger, New York (1983)
17. Noë, A.: Real Presence. Philosophical Topics **33**(1), 235–264 (2005)
18. O'Regan, J.K.: Why Red Doesn't Sound like a bell: Understanding the feel of Consciousness. Oxford University Press, New York (2011)
19. Fodor, J.A.: Methodological solipsism considered as a research strategy in cognitive psychology. Behav. Brain Sci. **3**(01), 63 (1980)
20. Ffytche, D.H.: The hodology of hallucinations. Cortex **44**(8), 1067–1083 (2008)
21. Collerton, D., Mosimann, U.P.: Visual hallucinations. Wiley Interdisc. Rev. Cognit. Sci. **1**(6), 781–786 (2010)
22. Clark, A.: Whatever next? Predictive brains, situated agents, and the future of cognitive science. Behav. Brain Sci. **36**(3), 181–204 (2013)
23. Friston, K., Kilner, J.M., Harrison, L.: A free energy principle for the brain. J. Physiol. Paris **100**(1–3), 70–87 (2006)
24. Hohwy, J.: The Predictive Mind. Oxford University Press, New York (2013)
25. Adams, R.A., Stephan, K.E., Brown, H.R., Frith, C.D., Friston, K.J.: The computational anatomy of psychosis. Front. Psychiatry **4**, 47 (2013). doi:10.3389/fpsyt.2013.00047
26. Reichert, D.P., Seriès, P., Storkey, A.J.: Charles Bonnet syndrome: evidence for a generative model in the cortex? PLoS Comput. Biol. **9**(7), e1003134 (2013)
27. Crick, F.: The recent excitement about neural networks. Nature **337**(6203), 129–132 (1989)
28. Zeigler, B.P.: Theory of Modelling and Simulation. Wiley, New York (1976)
29. Miłkowski, M.: Explaining the Computational Mind. MIT Press, Cambridge, Mass. (2013)
30. Bickhard, M.H.: Representational content in humans and machines. J. Exp. Theor. Artif. Intell. **5**(4), 285–333 (1993). doi:10.1080/09528139308953775
31. Morgan, A.: Representations gone mental. Synthese **191**(2), 213–244 (2013)
32. Bechtel, W.: Mental Mechanisms. Routledge, New York (2008)
33. Craver, C.F.: Explaining the Brain. Mechanisms and the mosaic unity of neuroscience. Oxford University Press, Oxford (2007)
34. Glennan, S.S.: Modeling mechanisms. Stud Hist. Philos. Sci. Part C Stud. Hist. Philos. Biol. Biomed. Sci. **36**(2), 443–464 (2005)
35. Machamer, P., Darden, L., Craver, C.F.: Thinking About Mechanisms. Philos. Sci. **67**(1), 1–25 (2000)
36. Craver, C.F.: Functions and mechanisms: a perspectivalist view. In: Hunemann, P. (ed.) Functions: Selection And Mechanisms, pp. 133–158. Springer, Dordrecht (2013)
37. Pöyhönen, S.: Carving the mind by its joints: culture-bound psychiatric disorders as natural kinds. In: Miłkowski, M., Talmont-Kaminski, K. (eds.) Regarding the Mind, Naturally: Naturalist Approaches to the Sciences of the Mental, pp. 30–48. Cambridge Scholars Publishing, Newcastle upon Tyne (2013)
38. MacKay, D.M.: Information, Mechanism and Meaning. MIT Press, Cambridge (1969)
39. Millikan, R.G.: Pushmi-pullyu representations. Philos. Perspect. **9**, 185–200 (1995)
40. Miłkowski, M.: Satisfaction conditions in anticipatory mechanisms. Biol. Philos. **30**(5), 709–728 (2015). doi:10.1007/s10539-015-9481-3

41. Burge, T.: Origins of Objectivity. Oxford University Press, Oxford (2010)
42. O'Regan, J.K., Noë, A.: A sensorimotor account of vision and visual consciousness. Behav. Brain Sci. **24**(5), 939–73; discussion 973–1031 (2001)
43. Hutto, D.D.: Knowing what? Radical versus conservative enactivism. Phenomenol. Cognit. Sci. **4**(4), 389–405 (2006)
44. Churchland, P.S., Ramachandran, V.S., Sejnowski, T.J.: A critique of pure vision. In: Large-Scale Neuronal Theories of the Brain, pp. 23–60. MIT Press, Cambridge, Mass (1994)
45. Neisser, U.: Cognition and Reality: Principles and Implications of Cognitive Psychology. W. H. Freeman, San Francisco (1976)
46. Bielecka, K.: Spread mind and causal theories of Content. Avant **V**(2), 87–97 (2014). doi:10.12849/50202014.0109.0004
47. Fodor, J.A.: A Theory of Content and Other Essays. MIT Press, Cambridge, Mass. (1992)
48. Miłkowski, M.: Function and causal relevance of content. New Ideas Psychol. **40**, 94–102 (2016)
49. Dretske, F.I.: Misrepresentation. In: Bogdan, R. (ed.) Belief: Form, Content, and Function, pp. 17–37. Clarendon Press, Oxford (1986)
50. Bickhard, M.H.: The interactivist model. Synthese **166**(3), 547–591 (2008)
51. Campbell, R.J.: The Concept of Truth. Palgrave Macmillan, Houndmills, Basingstoke, New York (2011)
52. Bechtel, W.: Investigating neural representations: the tale of place cells. Synthese (2014). doi:10.1007/s11229-014-0480-8