**Perception as Controlled Hallucination**

“Perception is controlled hallucination,” say leading proponents of *predictive processing* accounts of vision (discussed in §1).[[1]](#footnote-1) I believe they are right that something like this is a consequence of the view, but wrong in how they have thus far developed the idea (§2). In what follows, I advance my own analysis. The focus of my proposal is the *causal theory of perception* (§3). The causal theory famously allows for *veridical hallucination*, where a subject’s visual experience accurately represents the world but fails to qualify as successful perception because the causal condition on perception is not satisfied. This is how I propose to understand the line that “perception is controlled hallucination”: if predictive processing accounts are right, then what otherwise seem to be paradigmatic instances of successful perception are instead cases of controlled, veridical hallucination.[[2]](#footnote-2) Other, more deflationary understandings of the slogan are available, as are ways of developing the predictive processing view that drop the slogan altogether. But I think there is a case to be made for an account of perception according to which the slogan turns out to be literally true if the predictive processing account is right. In what follows, I make the case for such an account.

Central to my case is a distinction drawn by Ned Hall and others between the *difference-making* concept of causation and the *productive* concept(§4).[[3]](#footnote-3) I defend a productive version of the causal theory of perception, and so reject the difference-making approach. One point that distinguishes the two is that difference-making accounts allow *absences* to enter into causal relations while productive accounts do not. This is key to our discussion because of the central role absences play within predictive processing accounts of vision. The guiding idea of the paper will be that predictive processing accounts entail that the causal condition on perception is unsatisfied in certain cases—cases that involve absences—once the condition is understood in productive rather than difference-making terms.

I offer two interrelated arguments for favoring a productive theory of perception over a difference-making theory (§5). The first says that the productive theory better captures what distinguishes predictive processing views from their competitors, namely a kind of physical disconnection from the external world. The argument is loosely Kantian, focusing on the roles that a kind of receptivity and spontaneity play in perception. My second argument says that the productive theory better captures what putative instances of successful perception have in common with cases of uncontrolled hallucination, a point emphasized by defenders of the predictive processing view. After setting out these arguments, I conclude with a series of objections and replies to my view (§6).

This paper is meant as a contribution to the *metaphysics of science*, a naturalist project that involves using tools from metaphysics to help clarify and work out the implications of our leading scientific accounts, thereby advancing our understanding of those accounts.[[4]](#footnote-4) Traditional metaphysics sometimes relies too heavily on intuitions as evidence. A metaphysical project driven instead by science is likely to more often yield counterintuitive results, in the way that science sometimes does. This is a point I will say more about near the end of the paper, where I consider the intuitive costs of the view of perception I defend.

**1. Predictive Processing**

We begin by reviewing predictive processing accounts of vision, focusing on those elements most relevant to the discussion that follows.[[5]](#footnote-5) Predictive processing accounts can be understood as involving a hierarchical Bayesian computational model of vision implemented by a predictive coding algorithm. The computational models are *hierarchical* in that they attribute to the visual system a series of module-like stages, with different stages representing different features that distal objects might possess. Each stage directly interacts only with adjacent stages in the hierarchy, either above or below or at the same level. So, for example, a stage at a relatively higher level in the hierarchy might represent the category to which a distal object belongs, while lower-level stages might represent the object’s shape or color. The overall conscious percept is taken to be distributed across the hierarchy rather than located at any particular stage—a point to which we shall return (§6).

The models are *Bayesian* in that they draw on resources from Bayesian statistics to provide a computational-level account of the task the visual system performs.[[6]](#footnote-6) For example, a Bayesian model might attribute to the visual system (or to some particular stage within it) a hypothesis space consisting of a set of propositions describing properties of distal objects, together with a prior probability distribution over that space. And then the model might say that when the visual system receives new evidence in the form of a fresh incoming sensory signal, this probability distribution is updated in accordance with Bayes’ rule, which states that the posterior probability of a hypothesis is proportional to its prior probability together with the likelihood of the new evidence given the hypothesis in question.[[7]](#footnote-7)

On David Marr’s familiar framework, there are distinct computational, algorithmic, and implementational levels of analysis for information-processing systems.[[8]](#footnote-8) Hierarchical Bayesian models are pitched at the computational level, which concerns *what* a given system is doing, what its *task* is. A given computational model can be realized in different ways at the lower algorithmic level, which concerns *how* the given task is accomplished, including how the inputs and outputs specified at the computational level are represented by a system, and what the specific algorithm is that is used for transforming input to output. Defenders of the predictive processing approach maintain that the visual system makes use of *predictive coding* at the algorithmic level.[[9]](#footnote-9)

The core idea is that instead of representing an input directly, it is often more efficient for an information-processing system to represent just a *prediction error—*that is, the difference (or perhaps ratio) between an input and a prior prediction of what the input would be.[[10]](#footnote-10) Andy Clark illustrates the thought using an example of image transmission.[[11]](#footnote-11) For many naturally occurring images, the value of one pixel (e.g., suppose its color is black) is predictive of the value of its neighboring pixels (e.g., they too are likely to be black), with exceptions to this general trend often reflecting important features (e.g., boundaries between distinct objects). This creates room for a data compression strategy that would be attractive to our thrifty brains: explicitly encode just the unexpected variation in an image rather than the predicted variation.

Putting all the pieces together, here is a schematic version of the view that emerges. Let *H* be a comparatively higher-level stage in the visual hierarchy and let *L* be a directly adjacent lower-level stage. *H* makes a prediction about some feature in the environment to be perceived, where this prediction has implications for what the incoming sensory signal should be. This prediction relies on Bayesian elements such as a prior probability distribution. *H* passes this prediction down to *L*, which compares it to the actual sensory signal received. If the prediction matches the sensory signal, *L does nothing*. It does not pass the sensory signal along to *H*; there is no need, since *H* got it right. On the other hand, if the prediction is incorrect, *L* sends an *error signal* back to *H*, in effect passing along just that portion of the sensory signal that *H* was wrong about; again, there is no need to pass along what *H* got right. Upon receiving this error signal from *L*, *H* revises its probability distribution in accordance with Bayes’s rule, then uses the resulting posterior probability distribution to make a new prediction, starting the process all over again.

**2. But How is this Hallucination?**

Suppose the preceding account is correct. How is it supposed to follow that perception is controlled hallucination, as the slogan has it? According to Chris Frith, on the predictive processing account,

Our brains build models of the world and continuously modify these models on the basis of the signals that reach our senses. So, *what we actually perceive are our brain’s models of the world*. They are not the world itself, but for us, they are as good as. You could say that our perceptions our fantasies that coincide with reality.[[12]](#footnote-12)

Similarly, Jakob Hohwy writes “One important and, probably, unfashionable thing that this [predictive processing] theory tells us about the world is that *perception is indirect*… what we perceive is the brain’s best hypothesis… about the causes in the outer world.”[[13]](#footnote-13) Hohwy adds that “conscious experience is like a fantasy or virtual reality,” in that it is “at one remove from the real world it is representing.”[[14]](#footnote-14)

There are different ways to develop the thought being expressed here, with my account below representing one attempt. But perhaps the most common way to pursue it, made fairly explicit by Hohwy, proceeds along these lines.

(P1): If predictive processing accounts are correct then we do not have direct perceptual awareness of the external world.

(P2): If we do not have direct perceptual awareness of the external world then perception is a form of hallucination**.\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

(C): If predictive processing accounts are correct then perception is a form of hallucination.

Now, (P1) is very much open to dispute, but I will grant it here.[[15]](#footnote-15) Instead I reject (P2). There are, after all, *indirect realist* views of perception, according to which we indirectly but successfully (i.e., non-hallucinatorily) perceive external objects by virtue of being directly aware of internal entities—entities often taken to be sense data, but which we might instead take to be the brain’s best models or hypotheses.[[16]](#footnote-16) If the philosophical import of predictive processing is just a kind of indirect realism, this would be something worth noting, but it would not at all justify the striking claim that “perception is controlled hallucination.”[[17]](#footnote-17)

Now, I can envision an objection at this point. What we are running up against here are simply the limits of a slogan, says the objection. The line that “perception is controlled hallucination” is meant to be suggestive, it is meant as a loose metaphor that guides you in the right direction of a view. Nothing more. And so, it is no deep problem with the present analysis that it does not render the slogan literally true. In response to this objection, I have two points.

First, if the philosophical import of predictive processing really is just indirect realism, then the view is much less philosophically interesting or novel than we might have expected or hoped. Even as a loose metaphor, the line that “perception is controlled hallucination” seems to suggest something radically new, something revolutionary. But if it is just indirect realism, we could with equal justice say that Descartes and Locke and many of the traditional views of perception defended throughout the history of philosophy all took perception to be controlled hallucination, given their commitment to indirect realism.[[18]](#footnote-18) In short, the present analysis makes predictive processing *philosophically boring*. I think this indicates that the analysis is off track, it is failing to put its finger on what is interesting and novel. My alternative proposal below does better on this front, I claim. My view is sufficiently radical.

Second, there is reason to construe the talk of controlled hallucination as more than just loose metaphor. Part of the empirical case for predictive processing accounts is that they seem to shed light on various actual hallucinations experienced by subjects, for example in cases of schizophrenia or post-traumatic stress disorder, or in people suffering from Alzheimer’s or Parkinson’s diseases.[[19]](#footnote-19) In such cases, the idea is that hallucinations are the result of “strong priors,” or systems that assign too much weight to top-down predictions, so that they fail to adjust properly when these predictions do not match the incoming sensory signal.[[20]](#footnote-20) This is taken to show that veridical perception and hallucination fall along a continuum, involving much of the same underlying psychological machinery. As psychiatrist Philip Corlett and colleagues write, on the predictive coding view “the distinction between perception and hallucination becomes less clear,” a result that they hope “renders hallucinations more understandable and less stigmatizing.”[[21]](#footnote-21) For our purposes the point is we should not be too quick to dismiss the controlled hallucination talk as mere metaphor, and so it perhaps really is an advantage of my account that I am able to construe such talk literally.

**3. The Causal Theory of Perception**

At the heart of my counterproposal is the causal theory of perception, according to which a subject *S* sees that an object *o* has the property *F* only if (i) *S* has a visual experience as of *o* being *F*, (ii) *o* is *F*, and (iii) *o*’s being *F* causes *S*’s visual experience.[[22]](#footnote-22) The final clause is the distinctive claim of the theory. I will refer to it as the *Causal Condition*.

The most influential defenses of the causal theory appeal to intuitions elicited by thought experiments. Here is one such thought experiment, due to H. P. Grice.[[23]](#footnote-23)

**Clock**: A subject *S* has a visual experience as a of a clock on the shelf. There is in fact a clock on the shelf. But *S* is wearing an apparatus that stimulates her brain and causes her visual experience as of the clock, and so the clock on the shelf does not cause it.

According to Grice, *S* does not successfully see the clock on the shelf because the clock does not cause her visual experience; the Causal Condition is unsatisfied. Many philosophers have found this and similar defenses of the causal view compelling, so that the causal theory of perception has now reached the status of philosophical orthodoxy (even while there continue to be some heterodox opponents).

In this paper, I will simply be assuming that the causal theory of perception is correct, treating it as an unargued premise in order to work out certain consequences that follow. To be sure, the predictive processing literature is shot through with causal claims that seem to fit comfortably with the causal theory of perception. To pick just one example: “Following Helmholtz, we view the human perceptual system as a statistical inference engine whose function is to infer the probable *causes* of the sensory input,” write Peter Dayan, Geoffrey Hinton, Radford Neal, and Richard Zemel, in the opening line of their pathbreaking paper on generative models in machine perception, a work treated as foundational by proponents of predictive processing.[[24]](#footnote-24) However, my aim in this paper is not to use empirical claims from the predictive processing literature to support the causal theory of perception against its non-causal competitors. Rather, my aim is use such empirical claims to support one particular way of developing the causal theory against competing ways of developing the causal theory.

To set up my argument on that front, let us consider what have come to be known as cases of *veridical hallucination*. In such cases, conditions (i) and (ii) of the causal theory of perception are satisfied—and so the content of the given visual experience is accurate—but the Causal Condition (iii) is not. This is what occurs in **Clock**. *S*’s visual experience accurately represents the world and yet she does not count as successfully perceiving the clock since it does not cause her experience. She does not see the clock, she merely veridically hallucinates it.

My argument below says that if predictive processing accounts are correct, distal objects often fail to cause our visual experiences in the sense relevant to the Causal Condition. Open your eyes and attend to the world around you. There is a good chance that your visual experience is just like that of the subject in Grice’s **Clock** case, in the relevant causal respects. And so there is a good chance that you are undergoing your own veridical hallucination right now. Before I can make my case though, we need one last place in piece.

**4. Two Concepts of Causation**

We have two distinct concepts of causation, Ned Hall has argued.[[25]](#footnote-25) On the first, causes are understood as events that *make a difference* to their effects. Analyses of difference-making might appeal to counterfactuals, probability-raising, interventionist approaches that make use of structural equation models and directed acyclic graphs, or something else.[[26]](#footnote-26) It will be helpful to have a specific proposal on the table, so consider David Lewis’s 1973 counterfactual theory.[[27]](#footnote-27) Lewis holds that when *c* and *e* are distinct events, *e causally depends* on *c* just in case if *c* were not to occur then *e* would not occur. He then analyzes *causation* itself in terms of the ancestral of causal dependence: *c* causes *e* just in case *e* causally depends on *c*, or *e* causally depends on an intermediate event *d* which causally depends on *c*, or etc.

On the second concept of causation, as least as I will understand it, causes are events that *produce* their effects. Production is often spelled out in physical terms, as on Phil Dowe’s theory, which says that cause and effect are linked by a process involving some quantity that figures in a conservation law of physics, like mass-energy, linear momentum, charge, and so on.[[28]](#footnote-28) The causal processes that figure in vision often involve the transference of energy, as when light from the sun is reflected by the table in front of you and reaches your retina. Because of this, I will often put things in terms of energy transference in what follows.[[29]](#footnote-29)

Now, in the last half century or so, philosophers have advanced causal theories of a great many philosophically interesting notions. There are causal theories of reference, mental content, knowledge, justification, action, explanation, *perception*, and much more. However, because difference-making and production often coincide, Hall contends, philosophers have not been careful to assess just which causal concept should be operative in a given analysis. Do we want a difference-making theory of reference, or a productive theory? A difference-making theory of knowledge, or a productive theory? Different topics might call out for different concepts.[[30]](#footnote-30)

One place where the two concepts of causation come apart is in cases involving *absences*.[[31]](#footnote-31) In the stock example, the gardener fails to water a plant that subsequently dies. Was the gardener’s failure to water it—an absence—a cause of the plant’s death? *Yes*, say typical difference-making accounts. After all, there is counterfactual dependence: if the gardener had watered it, the plant would still be alive. *No,* say typical productive accounts. After all, there is no physical process involving the transference of energy connecting the gardener to the plant. The gardener, we can even imagine, is hundreds of miles away, napping.

In the following section, I defend a *productive theory of perception*. That is, I argue that the Causal Condition should be understood in terms of the productive concept of causation, not the difference-making concept. My arguments for this proposal will draw on the central role that absences play within predictive processing accounts.

**5. Productive Perception & Veridical Hallucination**

To see that central role, consider cases where the prediction that some higher-level stage, *H*, passes down to a lower-level stage, *L*, matches the incoming sensory signal that *L* receives. In that event, *L* *does nothing*—an absence. Nico Orlandi and Geoff Lee make this element of the predictive processing view explicit in their discussion of Clark’s work, asking us to consider

the case where predictions from above match the lower stage representations, so there is no error signal. The *absence* of error signal means that the lower stage is *prevented from causally influencing the higher stage*. If you’re thinking of “information flow” as a partly causal notion, it’s natural to read this as meaning that there is no information flow from the lower stage to the higher stage.[[32]](#footnote-32)

I want to embrace the causal claims made in this passage, and on that basis deny that the Causal Condition from the causal theory of perception is satisfied in such cases. This will require operating with the productive concept of causation rather than the difference-making one. Let me spell out in some detail how this would go.

Suppose that some distal object *o* is *F*. For concreteness, we can follow Grice and suppose that a clock (*o*) is on the shelf (*F*). Suppose further that *H*, a higher-level stage within *S*’s visual system, predicts (correctly) that the incoming sensory signal will indicate that a clock is on the shelf, and it sends this prediction down to *L*, a lower-level stage. Light reflected by the clock on the shelf reaches *S*’s retina and the resulting incoming sensory signal makes its way to *L*. So far, then, we have a physical process that connects the clock on the shelf to *L*. However, at this point the physical process comes to a stop. *L* does *not* pass the sensory signal along to *H*. It does not need to—it would be inefficient to do so—since *H* got its prediction right. So instead, *L* does nothing. As a result, *H* never comes to be connected to the distal object, the clock on the shelf, by any physical process involving the transference of energy.

Given that (i) *H* represents that a clock is on the shelf, and (ii) a clock is on the shelf, the first two conditions of the causal theory of perception are satisfied. However, the third Causal Condition is not, at least not if we understand it in terms of the productive concept of causation, since there is no process involving the transference of energy or any other conserved physical quantities that ever connects *H*’s representation that a clock is on the shelf to the actual presence of the clock on the shelf, just as there is no physical process that connects the gardener to the dying plant in the stock example of absence causation (§4). If this is right, and the Causal Condition is not satisfied, then what we have here is not a case of successful perception, but instead one of mere veridical hallucination, just as in Grice’s **Clock**.

Defenders of the difference-making version of the causal theory of perception can and should deny this. They should hold that the Causal Condition is satisfied in the case, since *H*’s representation that a clock is on the shelf counterfactually depends on there being a clock on the shelf, even if no physical process connects them. After all, if the clock had not been on the shelf, then *L* would have received a different incoming sensory signal, and would have detected a mismatch between that signal and *H*’s prediction, in which case *L* would have sent *H* an error signal, nudging it away from representing that a clock is on the shelf via a process of Bayesian updating. So then, since the Causal Condition is satisfied—if construed in difference-making terms—proponents of the difference-making theory of perception should regard this as a case of successful perception rather than one of mere veridical hallucination. The question then is, which view should we favor here.

I have two interrelated arguments for preferring the productive theory of perception. To set up the first, it will be useful to help ourselves to some suggestive terminology from Kant’s account of perception, even if we don’t put in the work needed to develop a nuanced Kantian understanding of them.[[33]](#footnote-33) Perception, Kant held, requires both the exercise of a kind of mental *receptivity*, or a capacity for the mind to be affected by external stimuli, and also a kind of mental *spontaneity*, or a capacity for the mind to initiate activity of its own, not directly under the control of external stimuli.[[34]](#footnote-34) The receptivity component is something that almost all competing views of perception will share. In vision, your eyes *receive* light; in hearing, your ears *receive* energy in the form of sound waves; and so on. Indeed, traditional empiricist or otherwise non-Kantian views have often thought of perception as involving something like *pure* *receptivity*, an entirely passive mental faculty whereby causal processes originating in the external world make their way into the mind and leave their mark. The mind acts as a kind of “mirror of nature,” in Richard Rorty’s phrase.[[35]](#footnote-35)

On the opposing Kantian view, perception involves more than just such receptivity, it also involves a kind of constructive activity carried out by the mind—spontaneity. On this point, many empirical researchers today follow Kant’s lead, assigning the mind a fairly active role in constructing perceptual states. As neuroscientist Andreas Engel and colleagues write,

Classical theories of sensory processing view the brain as a passive, stimulus-driven device. By contrast, more recent approaches emphasize the constructive nature of perception, viewing it as an active and highly selective process. Indeed, there is ample evidence that the processing of stimuli is controlled by top-down influences that strongly shape the intrinsic dynamics of thalamocortical networks and constantly create predictions about forthcoming sensory events.[[36]](#footnote-36)

However, to get to the key point, in an important sense, predictive processing goes well beyond this. What is distinctive about it is not that it assigns *some* *role* in perception to constructive activity, some role for a kind of Kantian spontaneity. Numerous psychological views do this, including those described by Engel and colleagues, who were writing before the advent of predictive processing. Rather, what is distinctive about predictive processing is that it assigns *no role* in perception to Kantian receptivity, at least in those cases that have been our focus. When the prediction that *H* passes down to *L* turns out to be accurate, *H* ends up exercising no receptivity at all, for it *receives* *nothing* back from the external world—no energy, no physical something else, just absence. This feature of the view is not shared by alternative constructivist views. Instead, it is the result of the particular way that predictive coding algorithms use absences as signals.

We can reinforce the present point even just by considering predictive processing accounts of those cases in which *L* detects that *H*’s prediction is wrong. In such cases, *H* *does* receive something, namely an error signal from *L*. And by receiving this error signal, *H* comes to be hooked up to the distal object, it comes to be causally connected to it by a physical process involving the transference of energy. Now, upon receiving this error signal, *H* revises its subsequent predictions in accordance with Bayesian principles, and in this way continues to exercise a form of spontaneity. However, this is *in addition to* the receptivity it has exercised. In contrast, in the cases we have made our focus, where *H*’s prediction is right, it receives nothing back. In such cases, receptivity drops entirely out of the equation, leaving nothing but the mind’s exercise of spontaneity.

Consider an analogy. You are arrested and sent to prison. The guards prohibit contact with the external world, with the exception that they allow you to write predictions of the letters your friends and family will send, and they promise to pass along whatever portions of those letters you predict wrongly. You go to work, writing out your prediction of your mother’s letter, which you then hand to the guard. The guard responds with an icy silence. Congratulations: your prediction was entirely correct. You *receive* nothing from your mother in this scenario; you remain utterly disconnected from the outside world. And yet according to the predictive processing view, something like this is the lonely life of quiet desperation that the higher-level stages of our visual systems find themselves leading, housed within the prisons of our skulls.

My argument on this point can be connected back to §2’s discussion of indirect realism. Suppose that Frith and Hohwy are right and that what we directly have access to in perception is not the external world but merely the brain’s model or best hypothesis. Still, if such a model or hypothesis were physically connected to the external world, as the last step in a causal chain that involves the transference of energy from some distal object, the view again would not be so different from Descartes’ or Locke’s indirect realism. Instead, what is novel here about predictive processing is the suggestion that such a model or hypothesis would be disconnected from the external world in something like the way you as a prisoner are disconnected from your mother. The issue is not that what you are directly aware of is just the brain’s model or best hypothesis, but that this model or best hypothesis seems insufficiently plugged into what is going on outside your head.

In contrast with the argument regarding indirect realism set out in §2, then, consider the loosely Kantian argument suggested by our discussion in this section.

(P1): Successful perception requires the exercise of a certain form of receptivity (not just spontaneity).

(P2): If predictive processing accounts are correct, then various instances of what otherwise seems like successful perception involve no exercise of receptivity.

(C): If predictive processing accounts are correct, then various instances of what otherwise seem like successful perception are not cases of successful perception.

These instances are instead cases of veridical, controlled hallucination. This argument gets at what makes the predictive processing account radical, what sets it apart from alternative views, including even alternative constructive ones. The productive theory of perception that I am defending can be understood as supporting (P1). It says that perception requires receiving energy from the distal object—this is the kind of receptivity at play. That the productive theory of perception directs attention to the key issue, and in doing so delivers the result that predictive processing advocates themselves endorse—that “perception is controlled hallucination”—is, I claim, a substantial advantage of the productive theory over the difference-making theory of perception.

We turn now to my second argument, which says that the productive theory better captures the sense in which putative cases of successful perception are like undisputed cases of hallucination, cases of “uncontrolled” hallucination. To make my case here, begin by considering a scenario in which *H* represents that a clock is on the shelf, and passes a prediction along these lines down to *L*. *L* sends no error signal back, not because *H*’s prediction is correct, but because *L* is malfunctioning and so is insensitive to the incoming sensory signal it receives. Since *H* receives no error signal back, it continues with its unrevised representation that a clock is on the shelf, and makes further predictions on this basis, even though this representation does not match the scene before the subject’s eyes. In effect, this is a case of someone hallucinating a clock. You will recall that predictive processing advocates understand various actual cases of hallucination along at least roughly these lines (§2). In this scenario, *H* is entirely disconnected from the external world, and it is this disconnection that marks it as a case of hallucination.

Now, tweak the case so that *L* *would* pass the sensory signal along to *H* if it were needed, but in fact this does not occur because the signal matches *H*’s prediction. In other words, tweak the case so that *L* is no longer malfunctioning. Then *H* seems *just as disconnected* from the external world as it was in the original case of undisputed hallucination, with the only difference being that now we have an assurance that *H* *would* *be* put into a real connection with the external world if, counterfactually, it were needed. That is, if *H*’s prediction did not match the sensory signal, then *L* would take the trouble to hook *H* up to the external world by passing along an error signal, but since *H*’s prediction was in fact right this is not needed, and so *H* can remain unplugged. What is common between the original case of uncontrolled hallucination and this tweaked case of putatively successful perception is this form of physical disconnection from the external world. I claim it is an advantage of the productive theory of perception that it captures this common element, where again the similarity between uncontrolled hallucination and supposedly successful perception is something that predictive processing advocates have often emphasized.

To the extent the difference-making approach seems in any way superficially attractive here, advocates of the productive theory can offer a counterfactual analysis of their own. They can analyze the *controlled* portion of “controlled hallucination” in counterfactual terms, so that a hallucination is *controlled* just in case it would be corrected by an error signal if it were wrong, and *uncontrolled* just in case it would not be so corrected if it were wrong. This sort of difference-making theory of control would not undermine in any way the productive theory of perception itself.[[37]](#footnote-37) In slogan form: “perception is controlled (understood counterfactually) hallucination (understood productively).” In this way, proponents of the productive theory of perception can get at whatever might seem right about a difference-making approach here while staying with their productive analysis of the Causal Condition.

The preceding two arguments get at important advantages of the productive theory of perception, or at least important advantages that hold conditional on the predictive processing account being correct. We should prefer the metaphysical view of perception that best fits with how our leading science understands perception, after all. Of course, empirical doubts about predictive processing would undercut the case I have made here for the productive theory of perception. But so it goes with projects in the metaphysics of science generally, where the aim is not to build empirically unassailable fortresses of speculation, but to use the tools of metaphysics to clarify and work out the implications of our leading scientific views.

**6. Objections and Replies**

In this concluding section, I further clarify and defend the positions set out in the preceding section by responding to a series of objections.

*Objection 1*. The first objection says that on the predictive processing view, although the entirety of the percept, distributed across the visual hierarchy, is not productively connected to the distal object, *part* of the percept is, and this should be good enough to satisfy the Causal Condition properly understood. When a baseball makes contact with a bat, not every square inch of the ball makes contact—the backside of the baseball does not. Analogously, when a mind makes contact with the world in perception, it is good enough that the lower-level stage *L* is productively connected to the distal object when it receives the incoming sensory signal; there is no additional requirement that *H*, the higher-level stage that passes its predictions down to *L*, must also be connected. *H* is like the “backside” of the percept.

*Reply*. I begin by noting that the objection concedes the productive theory of perception I have been defending: the objection claims not that the difference-making relation that *H* enters into with the distal object is sufficient for successful perception, but rather that the productive relation that *L* enters into is sufficient. Beyond this, if the critic pressing the objection will concede that the lack of a productive connection between *H* and the distal object is what makes predictive processing so radically different from its traditional competitors, that this is what makes perception seem rather hallucination­-*like* on predictive processing accounts, then we are in agreement on what matters most on my view, even if we disagree on whether perception “really is” controlled hallucination. If higher-level stages like *H* are part of what determine the representational content or phenomenology of our percepts, and such stages are physically disconnected from distal objects, this is a striking fact about perception, and one that is illuminated by attending to productive causal relations rather than difference-making ones.

*Objection 2*. The next objection argues that we can use those cases in which higher-level stages get their predictions wrong to secure productive causal relations for the cases in which they get them right. Suppose that at time *t*, *H* predicts that *o* is *F* and passes this prediction down to *L*; the incoming sensory signal indicates that *o* is not *F* but rather is *G*; *L* sends an error signal back to *H* and *H* undergoes Bayesian revision. Then at time *t*+1, *H* predicts that *o* is *G*; when it passes this prediction down to *L*, the prediction matches the sensory signal and so *L* sends nothing back. In that case there is a productive causal chain linking *o*’s being *G* at time *t* with *H*’s accurately representing that *o* is *G* at time *t*+1, and so the Causal Condition is satisfied after all, even when it is construed in productive terms.

*Reply*. I again begin by noting that the objection concedes the productive theory of perception: the objection claims not that a difference-making relation is sufficient for successful perception, but rather that a productive relation extended across time is sufficient. Beyond this, if the critic pressing the objection is willing to concede that the lack of a more immediate productive relation between *H* and the object perceived is what makes predictive processing so different from its traditional competitors, what makes perception seem rather hallucination*-like* on the view, then once again I think we are in agreement on what matters most, even if we disagree on just what the Causal Condition says regarding time. In that case, the critic should still allow that attending to productive causal relations rather than difference-making ones helps illuminate what is distinctive about predictive processing.

*Objection 3*. The next objection says that the productive theory of perception has implications beyond the cases we have been considering, and these implications are unacceptable. In particular, the productive theory entails that we cannot perceive absences. Sartre gives a famous example of supposedly seeing the absence of his friend Pierre in the café, an example that has since been used by many other philosophers.

**Pierre:** A subject *S* [for Sartre] has a visual experience that we can contentiously describe as an experience as of Pierre’s absence from the café. Pierre is in fact absent from the café. If Pierre had been present, *S* would not have had the visual experience he did.[[38]](#footnote-38)

The description of the experience is contentious in that some deny we have visual experiences as of absences. If we momentarily assume that we do, **Pierre** satisfies the first two conditions of the causal theory of perception. Whether it also satisfies the Causal Condition depends on which concept of causation we employ. Pierre’s absence makes a difference to Sartre’s visual experience, but is not connected to it via some physical process. Going beyond **Pierre**, the productive theory seems to entail that we cannot see holes, or shadows, or black holes, or perfectly black objects, if we assume such entities do not reflect or emit light. Several authors have regarded these results as intolerable and defended a difference-making theory of perception on just this basis.[[39]](#footnote-39)

*Reply*. Let me start with **Pierre**. Sartre’s visual experience would have been different if Pierre had been present, yes, but the same is true if the Duke of Wellington had been, or Paul Valéry, or really anyone in the entire world who was not in fact there in the café.[[40]](#footnote-40) Accordingly, each of these other absences also qualifies as a cause of Sartre’s visual experience on standard difference-making theories of causation: Sartre’s visual experience counterfactually depends on these other absences, and so is caused by them. If we suppose in addition that Sartre’s visual experience can be described with just as much justice as an experience of these other absences—after all, it’s at least difficult to say how Pierre’s absence *looks* different from Wellington’s or Valéry’s; it’s not as if Pierre’s absence has a distinctive color or shape that the others lack[[41]](#footnote-41)—it follows that all such absences satisfy all three conditions of the difference-making version of the causal theory of perception. But it is absurd to hold that Sartre or the rest of us see the absence of every single person who is not in front of us. The difference-making theory of perception seems to be trapped into this absurdity. The productive theory of perception avoids it by simply denying that any absence satisfied the Causal Condition, and so by denying that any absence is ever seen.[[42]](#footnote-42)

In addition, the productive theory can be combined with what I think is an attractive view of absences as *unobservables*, akin to electrons or genes. That is, defenders of the productive theory of perception need not deny that absences exist, or that they make a difference to various effects. Indeed, they can allow that in many domains other than perception, the difference-making concept of causation is more fitting than the productive one. What they must deny is just that we can see them. When you direct your eyes at a piece of Swiss cheese, you do not see the holes in the cheese, since they transfer no energy to your eyes. Rather, you see *through* the holes to whatever objects are on the other side, and then infer the existence of the holes, just as in other cases we infer from our observations the existence of unobservable entities. Or at least so the defender of the productive theory of perception can maintain.

Along similar lines, perhaps proponents of the productive theory of perception can maintain that we do not see shadows, which are “holes in light,” but rather we see through them, often to objects on which such shadows fall.[[43]](#footnote-43) And perhaps they can take a hardline on black holes and perfectly black objects, denying we can see them at all. “By its very nature, a black hole cannot be seen,” says NASA’s website, reporting on the Event Horizon Telescope taking what was sometimes described as the first photograph of a black hole.[[44]](#footnote-44) “Scientists develop a material so dark that you can’t see it,” reads the headline of the online newspaper *The Independent*, describing Vantablack, a material that absorbs 99.6% of visible light.[[45]](#footnote-45) Perhaps there is genuine metaphysical wisdom in these snippets.

To develop these responses at length would take us beyond the scope of the present paper. It is in the spirit of the naturalist project being pursued here to allow that further work would need to be done to show that the productive theory of perception can fit with our comprehensive empirical understanding of perception, and to concede that if this work cannot be done, we may have empirical reason to go back and reject the productive theory after all. Again, the aim is not to build empirically unassailable fortresses of speculation.

*Objection 4*. The final objection says that the result that a great many cases of what seems like successful perception instead get classified as cases of veridical, controlled hallucination is a reductio of the productive theory of perception. It is a Moorean fact that we often successfully perceive the world. I *see* one hand, and then another; I do not *veridically hallucinate* one hand, and then another. If the difference-making theory of perception accommodates this Moorean fact while the productive theory does not, then this by itself outweighs whatever other theoretical benefits the productive theory might possess.

*Reply*. My view is that this type of consideration is often assigned too much weight in traditional metaphysics, and that part of what is exciting about the metaphysics of science project is its appeal to forms of evidence beyond intuitions and supposed Moorean facts, opening space for results that might require us to radically revise our understanding of the world, as science often does. Perhaps the visual system operates by generating experiences that involve less successful perception than common sense assumes, and more active prediction that operates in advance of or even in the absence of successful seeing. Provided there are mechanisms that ensure such predictions do not go too far off track—mechanisms that keep the hallucinations controlled and veridical rather than uncontrolled—just what would be lost this way? It may be that such a view is mistaken, but if so, supposed Moorean certainties are unlikely to go far in telling us how. For that, we will need science. Insofar as there is good empirical reason to accept the predictive processing account, and insofar as the account supports the productive theory of perception, we should be open to accepting the radical consequences that might follow.

REFERENCES

Aronson, J. (1979). “On the Grammar of ‘Cause’,” *Synthese*, 22: 414-430.

Casati, R. & Varzi, A. (1994). *Holes and Other Superficialities.* Cambridge, MA: The MIT Press.

Casati, R. & Cavanaugh, P. (2019). *The Visual World of Shadows*. Cambridge, MA: The MIT

Press.

Clark, A. (2013). “Whatever Next? Predictive Brains, Situated Agents, and the Future of

Cognitive Science,” *Behavioral and Brain Sciences*, 36.3: 181-204.

Clark, A. (2016). *Surfing Uncertainty: Prediction, Action, and the Embodied Mind*. Oxford:

Oxford University Press.

Corlett, P. R., Frith, C. D., & Fletcher, P. C. (2009). “From Drugs to Deprivation: A Bayesian

Framework for Understanding Models of Psychosis,” *Psychopharmacology*, 206.4: 515-530.

Corlett, P. R., Horga, G., Fletcher, P. C., Alderson-Day, B., Schmack, K., & Powers, A. R. (2019).

“Hallucinations and Strong Priors,” *Trends in Cognitive Science*, 23.2: 114-127.

Dowe, P. (1992). “Wesley Salmon’s Process Theory of Causality and the Conserved Quantity

Theory,” *Philosophy of Science*, 59.2: 195-216.

Dowe, P. (2000). *Physical Causation*. New York: Cambridge University Press.

Dowe, P. (2001). “A Counterfactual Theory of Prevention and ‘Causation’ by Omission,”

*Australasian Journal of Philosophy*, 79.2: 216-226.

Drayson, Z. (2018). “Direct Perception and the Predictive Mind,” *Philosophical Studies*, 175.12:

3145-3164.  
Eells, E. (1991). *Probabilistic Causality.* Cambridge: Cambridge University Press.

Engel, A. K., Fries, P. & Singer, W. (2001). “Dynamic Predictions: Oscillations and

Synchrony in Top-Down Processing,”. *Nature Reviews*, 2: 704-716.

Fair, D. (1979). “Causation and the Flow of Energy,” *Erkenntnis*, 14: 219-250.

Farennikova, A. (2013). “Seeing Absences,” *Philosophical Studies*, 166.3: 429-454.

Fletcher, P. & Frith, C. (2009). “Perceiving is Believing: A Bayesian Approach to Explaining

the Positive Symptoms of Schizophrenia,” *Nature Reviews: Neuroscience*, 10: 48-58.

French, S. (2014). *The Structure of the World: Metaphysics and Representation*. Oxford

University Press.

Friston, K. (2005). “A Theory of Cortical Responses,” *Philos. Trans. R. Soc. Land. B. Biol.*

*Sci.,* 360.1456: 815-836.

Friston, K. (2009). “The Free-Energy Principle: A rough guide to the brain?” *Trends in Cognitive*

*Sciences*, 13: 293-301.

Friston, K. & Kiebel, S. (2009). “Predictive Coding Under the Free-Energy Principle,” *Philos.*

*Trans. R. Soc. Land. B. Biol. Sci.,* 364.1521: 1211-1221.

Frith, C. (2007). *Making Up the Mind: How the Brain Creates Our Mental World*. Oxford: Wiley-

Blackwell.

Goldman, A. I. (1977). “Perceptual Objects,” *Synthese*, 35.3: 257-284.

Grice, H. P. (1961). ““The Causal Theory of Perception”, Proceedings of the Aristotelian Society

Supplementary Volume 35: 121–52.

Griffiths, T. L., Kemp, C. & Tenenbaum, J. B. (2008). “Bayesian Models of Cognition,” in *The*

*Cambridge Handbook of Computational Psychology*, ed. R. Sun, pp. 59-100. New York: Cambridge University Press.

Grush, R. (2004). “The Emulation Theory of Representation: Motor Control, Imagery, and

Perception,” *Behavioral and Brain Sciences*, 27.3: 377-397.

Hall, N. (2004). “Two Concepts of Causation,” in *Causation and Counterfactuals*, eds. J.

Collins, N. Hall, & L. A. Paul, pp. 225-276. Cambridge, MA: The MIT Press.

Helmholtz, H. (1925). “Treatise on Physiological Optics,” translated from the 3rd German

Edition. Edited by Southall, J. P. C. Rochester, NY: Optical Society of America.

Hitchcock, C. (2018). “Causal Models,” *Stanford Encyclopedia of Philosophy*, URL =

https://plato.stanford.edu/entries/causal-models/

Hohwy, J. (2007). “Functional Integration and the Mind,” *Synthese*, 159.3: 315-328.

Hohwy, J. (2013). *The Predictive Mind*. Oxford: Oxford University Press.

Hohwy, J. (2016). “The Self-Evidencing Brain,” *Noûs*, 50.2: 259-285.

Horn, B. K. P. (1980). “Derivation of Invariant Scene Characteristics from Images,” *Proceedings of*

*the May 19-22, 1980, National Computer Conference:* 371-367.

Johnston, I. (2014). “Blackest is the New Black: Scientists Develop a Material So Dark that You

Can’t See It,” *Independent*, URL = <https://www.independent.co.uk/news/science/blackest-is-the-new-black-scientists-have-developed-a-material-so-dark-that-you-cant-see-it-9602504.html>

Kistler, M. (2020). “Introduction: New Trends in the Metaphysics of Science,” *Synthese*,

197: 1841-1846.

Knill, D. C. & A. Pouget. (2004). “The Bayesian Brain: The Role of Uncertainty in Neural

Coding and Computation,” *Trends in Neurosciences*, 27.12: 712–719.

Ladyman, J. and Ross, D., with Spurrett, D. & Collier, J. (2007). *Everything Must Go:*

*Metaphysics Naturalised*. Oxford University Press.

Lawson, R. P., Rees, G., & Friston, K. J. (2014). “An Aberrant Precision Account of Autism,”

*Frontiers in Human Neuroscience*, 8: 302.

Lee, T. S. & Mumford, D. (2003). “Hierarchical Bayesian Inference in the Visual Cortex,”

*Journal of Optical Society of America*, A 20.7: 1434-1448.

Lewis, D. K. (1973). “Causation,” *Journal of Philosophy*, 70: 556-67.

Lewis, D. K. (1980). “Veridical Hallucination and Prosthetic Vision,” *Australasian Journal of*

*Philosophy*, 58: 239-249.

Lewis, D. K. (2000). “Causation as Influence,” *Journal of Philosophy*, 97: 182-197.

Marr, D. (1982). *Vision: A Computational Investigation into the Human Representation of Visual*

*Information*. Cambridge, MA: The MIT Press.

Menzies, P. (2004). “Difference-making in Context,” in *Causation and Counterfactuals*,

eds. J. Collins, N. Hall, and L. A. Paul, pp. 139-80.. Cambridge, MA: MIT Press.

NASA. (2019). “Black Hole Image Makes History,” URL =

<https://www.jpl.nasa.gov/news/news.php?feature=7372>

Newman, L. (2009). “Ideas, Pictures, and the Directness of Perception in Descartes and Locke,”

*Philosophy Compass*, 4.1: 134-154.

Orlandi, N. & Lee, G. (2019). “How Radical is Predictive Processing?” in *Andy Clark and His*

*Critics*, eds. M. Colombo, E. Irvine, and M. Stapleton, pp.. Oxford: Oxford University Press.

Paul, L. A. & Hall, N. (2013). *Causation: A User’s Guide*. Oxford: Oxford University Press.

# Rao, R. P. & Ballard, D. H. (1999). “Predictive Coding in the Visual Cortex: A Functional

# Interpretation of Some Extra-Classical Receptive-Field Effects,” *Nature Neuroscience,* 2.1:79-87.

Rescorla, M. (2015). “Bayesian Perceptual Psychology,” in *The Oxford Handbook of Philosophy*

*of Perception*, ed. M. Matthen, pp. 694-716.. Oxford: Oxford University Press.

Rescorla, M. (2017). “Review of *Surfing Uncertainty: Prediction, Action, and the Embodied*

*Mind* by Andy Clark,” *Notre Dame Philosophical Reviews*. URL =

<https://ndpr.nd.edu/news/surfing-uncertainty-prediction-action-and-the-embodied-mind/>.

Salmon, W. (1984). *Scientific Explanation and the Causal Structure of the World.* Princeton:

Princeton University Press.

Sartre, J. P. (1956). *Being and Nothingness*, trans. H. E. Barnes, New York: Washington Square

Press.

Schaffer, J. (2000). “Causation by Disconnection,” *Philosophy of Science*, 67.2: 285-300.

Schaffer, J. (2004). “Causes Need Not be Physically Connected to their Effects: The Case

for Negative Causation,” in *Contemporary Debates in Philosophy of Science*, ed. C. R. Hitchcock, pp. 197-216.. Malden, MA: Blackwell**.**

Sider, T. (2020). *The Tools of Metaphysics and the Metaphysics of Science*. Oxford University

Press.

Sorensen, R. (2008). *Seeing Dark Things: The Philosophy of Shadows*, New York: Oxford

University Press.

Strawson, P. F. (1979). “Perception and Its Objects,” in *Perception and Identity*, ed. G. F.

Macdonald, pp. 41-60. London: Palgrave.

Suppes, P. (1971). *A Probabilistic Theory of Causality*. Amsterdam: North Holland Publishing

Company.

Swanson, L. R. (2016). “The Predictive Processing Paradigm has Roots in Kant,” *Frontiers in*

*Systems Neuroscience*, 10: 79.

Tenenbaum, J. B., Kemp, C., Griffiths, T. L., & Goodman, N. D. (2011). “How to Grow a Mind:

Statistics, Structure, and Abstraction,” *Science*, 331: 1279-1285.

Tye, M. (1982). “A Causal Analysis of Seeing,” *Philosophy and Phenomenological Resarch*, 42: 311-

325.

Wacongne, C., Labyt, E., van Wassenhove, V., Bekinschtein, T., Naccache, L. & Dehaene, S.

(2011). “Evidence for a Hierarchy of Predictions and Prediction Errors in Human Cortex,” *Proc. Natl. Acad. Sci. U S A*, 108.51: 20754-20759.

Wiese, W. & Metzinger, T. (2017). “Vanilla PP for Philosophers: A Primer on Predictive

Processing,” *Philosophy and Predictive Processing*, URL = < https://predictive-mind.net/papers/vanilla-pp-for-philosophers-a-primer-on-predictive-processing>.

Woodward, J. (2003). *Making Things Happen: A Theory of Causal Explanation*. Oxford:

Oxford University Press.

1. Hohwy (2013: 208), Clark (2016: 14), and Wiese and Metzinger (2017: 3). Horn (1980) attributes the slogan to artificial intelligence researcher Max Clowes, but Wiese and Metzinger (2017) report that no published remarks by Clowes seem to match. Grush (2004: 395) attributes it to Ramesh Jain, while Clark (2016: 308, n.3) suggests it may have been neuroscientist Rodolfo Llinas. The slogan has filtered down to popular culture and has been the topic of popular media articles, TED talks, Sam Harris podcasts, and more [↑](#footnote-ref-1)
2. Here and throughout, I assume that if a mental state is an instance of hallucination then it cannot literally be an instance of (successful) perception. I will still make free use of the slogan that “perception is controlled hallucination,” but this should be understood as shorthand for the more careful formulation just given in the text. [↑](#footnote-ref-2)
3. Hall (2004). [↑](#footnote-ref-3)
4. Representative work from this project includes for instance Ladyman & Ross (2009), French (2014), Sider (2020), and Kistler (2020). [↑](#footnote-ref-4)
5. My discussion draws on Hohwy (2013), Clark (2013) and (2016), Wiese and Metzinger (2017), and Orlandi and Lee (2019). I self-consciously adopt the bad habit of equating perception with vision alone, ignoring the other senses. My excuse is that leading discussions of the slogan “perception is controlled hallucination” focus primarily on vision, and it will be helpful at points to put things in their terms. [↑](#footnote-ref-5)
6. On Bayesian models, see for instance Knill and Pouget (2004), Griffiths et al. (2008), and Tenenbaum et al. (2011). Rescorla (2015) provides a philosophical overview of Bayesian perceptual psychology. [↑](#footnote-ref-6)
7. *Bayes’s Theorem* states that P(*h*|*e*) = (P(*e*|*h*) × P(*h*))/P(*e*), where P(*e*) is assumed to be greater than 0. ‘P’ here is a probability function, assigning propositions values between 0 and 1. P(*h*|*e*) and (P(*e*|*h*) are conditional probabilities, so that for instance P(*h*|*e*) is the probability of *h* conditional on *e*. *Bayes’s Rule* is the normative prescription that when one obtains evidence *e*, one should update the probability one assigns to *h* so that it is equal to P(*h*|*e*). [↑](#footnote-ref-7)
8. Marr (1982). [↑](#footnote-ref-8)
9. Rescorla (2017) objects to Clark (2016) on the basis that the evidence for the algorithmic-level predictive coding account is weaker and more speculative than the evidence for the computational-level hierarchical Bayesian account; see also Orlandi and Lee (2019). This is a line of objection we will set aside, focusing on what follows *if* predictive processing accounts are correct. We will also set aside questions about Marr’s bottom, implementational level, but for work exploring the possible neural implementation of predictive processing, see for instance Engel et al. (2001), Friston (2005), Wacongne et al. (2011). [↑](#footnote-ref-9)
10. On predictive coding, see for instance Rao and Ballard (1999), Lee and Mumford (2003), Friston (2009), and Friston & Kiebel (2009). [↑](#footnote-ref-10)
11. Clark (2016: 26). [↑](#footnote-ref-11)
12. Frith (2007: 135), emphasis added. [↑](#footnote-ref-12)
13. Hohwy (2007: 322), emphasis added. [↑](#footnote-ref-13)
14. Hohwy (2013: 138). Hohwy (2016) pursues a related line of thought to argue that predictive processing accounts invite skepticism, a conclusion he embraces (with qualifications). [↑](#footnote-ref-14)
15. Clark (2016: 195) says that predictive processing allow for what he calls “not-indirect perception,” which I take to be a rejection of (P1). Drayson (2018) distinguishes between different senses of directness in order to reject (P1). [↑](#footnote-ref-15)
16. This is in contrast with *direct realism*, which takes subjects to be directly aware of external objects in perception. [↑](#footnote-ref-16)
17. There is a familiar *argument from hallucination* for indirect realism, but it is a distraction here. What it says is not that perception is in fact hallucination, but that perception and hallucination share a common factor (e.g., sense data), and this common factor is the object of direct awareness. [↑](#footnote-ref-17)
18. For discussion of Descartes’s and Locke’s indirect realism, see for instance Newman (2009). [↑](#footnote-ref-18)
19. Friston (2005), Corlett et al. (2009), Fletcher and Frith (2009), Corlett et al. (2019). [↑](#footnote-ref-19)
20. And on the flipside, Lawson et al. (2014) propose that autism spectrum disorder involves assigning too much weight to prediction errors and not enough to top-down predictions, resulting in subjects who lose the incoming signal for the noise. [↑](#footnote-ref-20)
21. Corlett et al. (2019: 111). [↑](#footnote-ref-21)
22. Classic defenses of the causal theory include Grice (1961), Strawson (1979), and Lewis (1980). The three conditions are meant to be individually necessary but not jointly sufficient. [↑](#footnote-ref-22)
23. Grice (1961). [↑](#footnote-ref-23)
24. Dayan, Hinton, Neal, & Zemel (1995): 889, emphasis added. The authors are alluding to Helmholtz (1925), which is itself a foundational work for predictive processing accounts; see for instance Wiese & Metzinger (2017). [↑](#footnote-ref-24)
25. Hall (2004). [↑](#footnote-ref-25)
26. On counterfactual theories, see Lewis (1973) and (2000); on probability-raising, Suppes (1970) and Eells (1991); on interventionist approaches, Woodward (2003), and on causal models, Hitchcock (2018). [↑](#footnote-ref-26)
27. Lewis (1973). [↑](#footnote-ref-27)
28. Dowe (1992), (2000). See also Aronson (1971) and Fair (1979) for earlier views in the vicinity. Dowe’s work partly draws on and critically responds to Salmon’s (1984) causal process theory. [↑](#footnote-ref-28)
29. Here and throughout, I adopt Hall’s “two concepts” framing, but this is not absolutely crucial to my argument. You might instead hold that all real causation just is difference-making, while granting that a theoretically interesting subset of difference-making relations are grounded in physical processes involved conserved quantities. Or you might hold instead that all real causation involves such physical processes, while granting that difference-making relations can be theoretically interesting even when not grounded in such processes. If need be, I could reframe my argument using either of these alternative views. [↑](#footnote-ref-29)
30. For example, Hall (2004: 268) suggests that the difference-making concept should be used within causal decision theory while the productive concept should be used in the causal theory of persistence. [↑](#footnote-ref-30)
31. On whether absences can be causes, see for instance Lewis (2000) and (2004); Dowe (2000) and (2004); Schaffer (2000) and (2004); Hall (2004), Paul and Hall (2013); Menzies (2004). [↑](#footnote-ref-31)
32. Lee and Orlandi (2019). In the passage, the authors are describing a gloss on predictive processing models that they do not themselves endorse. [↑](#footnote-ref-32)
33. For an argument that there is a deep connection between predictive processing and Kant’s view of perception, see Swanson (2016). [↑](#footnote-ref-33)
34. Kant (1999). McDowell (1994) invokes Kantian spontaneity and receptivity in his argument that perceptual states have conceptual rather than nonconceptual content, but that is a topic I don’t mean to engage here at all. [↑](#footnote-ref-34)
35. Rorty (1979). Incidentally, both Hohwy (2013: Ch. 11) and Clark (2016: 305) use Rorty’s phrase to describe the sort of views they mean to oppose in defending predictive processing. [↑](#footnote-ref-35)
36. Engel et al. (2001: 704). [↑](#footnote-ref-36)
37. Cf. Dowe (2001), who defends a counterfactual theory of prevention and “quasi causation” without giving up his physical theory of causation. [↑](#footnote-ref-37)
38. Sartre (1956: 40-41). See Farrenikova (2013) and Sorensen (2008) for recent defenses of absence perception that discuss Sartre’s case. [↑](#footnote-ref-38)
39. On the perception of holes see Casati and Varzi (1994: Ch. 11); on shadows, Sorensen (2008); on black holes, Goldman (1977); and on perfectly black objects, Tye (1982). Each of these authors appeals to a difference-making account of causation to make sense of how we can perceive such entities. [↑](#footnote-ref-39)
40. The examples of Wellington and Valéry are taken from Sartre’s own discussion. [↑](#footnote-ref-40)
41. Suppose the contents of visual experience are sets of centered worlds. The set that is the content of Sartre’s experience presumably includes no world in which Wellington or Valéry are present in the café, just as it includes no world in which Pierre is. [↑](#footnote-ref-41)
42. In effect, this reply exploits what is known as the *problem of profligate* causation for absences—if there is any absence causation at all, there is bound to be a lot of it. For discussion see Menzies (2004). [↑](#footnote-ref-42)
43. The thought that shadows are “holes in light” is taken from Casati & Kavanaugh (2019). [↑](#footnote-ref-43)
44. NASA (2019). [↑](#footnote-ref-44)
45. Johnson (2014). [↑](#footnote-ref-45)