**Perception as Controlled Hallucination**

“Perception is controlled hallucination,” according to certain proponents of *predictive processing* accounts of vision (§1).[[1]](#footnote-1) I think they are right that something like this is a consequence of their view but wrong in how they have developed the idea (§2). In what follows, I advance my own analysis, focusing on the *causal theory of perception* (§3). I argue that the causal theory should be understood in terms of a *productive* concept of causation, as opposed to a *difference-making* concept (§4). On my analysis, predictive processing accounts entail that various putative instances of successful perception are instead cases of *veridical hallucination* because of the role they assign to absences, which cannot enter into productive causal relations (§5).[[2]](#footnote-2) In defending a productive theory of perception, I advance two interrelated lines of argument. The first appeals to a series of thought experiments, of the sort familiar from H. P. Grice’s original defense of the causal theory of perception (§6). The second appeals to Kant’s view that perception requires both a kind of spontaneity and receptivity (§7). I conclude by sketching some potential responses to my arguments (§8).

**1. Predictive Processing**

We begin by reviewing predictive processing accounts of vision, focusing especially on the elements most relevant to the discussion below.[[3]](#footnote-3) 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 modular-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 of the hierarchy might represent the category to which a distal object belongs, stages at intermediate levels might represent the object’s shape and color, and stages at the lowest level might just encode proximal sensory stimulation. The overall conscious percept is taken to be distributed across the hierarchy rather than located at any particular stage—a point we will return to below.

The models are *Bayesian* in that they draw on resources taken from Bayesian statistics to provide a computational-level account of the task the visual system performs.[[4]](#footnote-4) 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 will be 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.[[5]](#footnote-5)

On David Marr’s familiar framework, there are distinct computational, algorithmic, and implementational levels of analysis for information-processing systems.[[6]](#footnote-6) Hierarchical Bayesian models of the sort we have been describing are pitched at the computational level, which concerns *what* a given system is doing, what its *task* is. At least roughly in this case, that task is to create a probabilistic, generative model of the environment, from sensory inputs. A given computational model can be realized in different ways at the lower algorithmic level, which concerns *how* the given task is accomplished, including for instance how the inputs and outputs specified at the computational level are represented by a system, and what the algorithm used for the transformation from input to output is. Defenders of the predictive processing approach maintain that the visual system makes use of *predictive coding* at the algorithmic level.[[7]](#footnote-7)

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.[[8]](#footnote-8) Andy Clark illustrates the thought using the example of image transmission.[[9]](#footnote-9) 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 figures to 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 an arbitrary, comparatively higher-level stage in the visual hierarchy and let *L* be a directly adjacent lower-level stage. *H* will make a prediction about some feature in the environment to be perceived, where this prediction has implications for what the incoming sensory signal will 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 to, since *H*’s prediction 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 for the sake of argument that the predictive processing account is correct. How is it supposed to follow that perception is a (controlled) form of hallucination, as the slogan puts it? According to Chris Frith,

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.[[10]](#footnote-10)

Similarly, Jakob Hohwy writes that “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.”[[11]](#footnote-11) 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.”[[12]](#footnote-12) Here is a way to formalize the Frith/Hohwy argument as I understand it.

(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 challenge, but I am willing to grant it here.[[13]](#footnote-13) 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 suppose are the brain’s best models or hypotheses.[[14]](#footnote-14) If the philosophical import of predictive processing is that the view supports a kind of indirect realism, this would not at all justify the striking conclusion that perception is controlled hallucination.[[15]](#footnote-15)

. In fact, if the import is just indirect realism, then all the recent talk about perception as controlled hallucination turns out to be much less philosophically interesting than we might have hoped or expected. We could with equal justice say that Descartes and Locke and many of the traditional views of perception throughout the history of philosophy have taken perception to be controlled hallucination, if all this slogan means is an embrace of indirect realism.[[16]](#footnote-16) I think this suggests that the analysis is off track. That perception is controlled hallucination should be getting at something revolutionary, something new. My counterproposal manages this.

**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.[[17]](#footnote-17) The final clause is the distinctive claim of the causal theory. I will refer to it as the *Causal Condition*.

The most influential defenses of the causal theory appeal to thought experiments. Here I will recount two famous ones due to H. P. Grice.[[18]](#footnote-18)

**Pillar**: A subject *S* has a visual experience as of a pillar at a certain distance away and at a certain location. There is in fact a pillar at that location. But, unbeknownst to *S*, a mirror has been interposed between her and this pillar, a mirror that reflects a second, numerically distinct but similar pillar. It is this second pillar, and not the first, that is the cause of *S*’s visual experience.

Which pillar does *S* see? The second rather than the first, says Grice, because the second satisfies the Causal Condition while the first does not.

**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.

In that case, Grice claims, *S* does not see the clock on the shelf because the clock does not cause her visual experience; the Causal Condition is unsatisfied.

The causal theory allows for what have come to be known as cases of veridical hallucination, where conditions (i) and (ii) of the theory are satisfied—and so the content of the given visual experience is accurate—but the Causal Condition (iii) is not. **Clock** is an example. *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, and so the Causal Condition is unsatisfied. The subject does not see the clock, she merely veridically hallucinates it.

I say that if predictive processing accounts are correct, external objects often fail to cause our visual experiences, in the sense of cause relevant to the Causal Condition. This applies even in cases of what otherwise seem to be paradigmatic successful perception. 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 the case of Grice’s clock, in the relevant respects. You are not seeing the world around you; you are merely veridically hallucinating it, because the Causal Condition on perception is not satisfied. To support this claim, we turn now to theories of causation.

**4. Two Concepts of Causation**

We have two distinct concepts of causation, Ned Hall has argued.[[19]](#footnote-19) 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.[[20]](#footnote-20) It will be helpful to have a specific proposal on the table, so consider David Lewis’s 1973 counterfactual theory.[[21]](#footnote-21) 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, causes are understood as events that *produce* their effects. Production is often spelled out in physical terms, as on Phil Dowe’s theory, which entails that cause and effect are linked by a process involving some quantity that figures in the conservation laws of physics, like mass-energy, linear momentum, charge, and so on.[[22]](#footnote-22) 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.

My discussion in this paper will adopt Hall’s “two concepts” framing, but this is not absolutely essential to my argument. You might hold instead that causation just is difference-making, while allowing that a theoretically interesting subset of difference-making relations are grounded in physical processes involving conserved quantities. Or you might hold that all genuine causation involves such physical processes, while allowing that difference-making relations can be theoretically interesting even when they are not grounded in such processes, and so even when they do not track causal relations.[[23]](#footnote-23) I could reformulate the arguments to come within either of these alternative approaches without sacrificing anything strictly necessary to my case.

In recent decades, 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. Because difference-making and production often coincide, Hall suggests, philosophers have often not been careful to assess which causal concept should be operative in a given analysis. Different topics might yield different answers. For example, Hall suggests that difference-making is what is needed within causal decision theory, while production is called for within causal theories of persistence.[[24]](#footnote-24) Our question is, which concept should figure within the causal theory of perception? Do we want a *productive theory of perception*, or a *difference-making theory*?[[25]](#footnote-25)

**5. Absences**

Although the two concepts of causation often coincide, one place where they come apart involves *absences*.[[26]](#footnote-26) 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 the plant, it would still be alive. *No,* say typical productive accounts.[[27]](#footnote-27) After all, there is no physical process involving the transference of a conserved physical process connecting the gardener to the plant. The gardener, we can even imagine, is hundreds of miles away, taking a nap or playing with his kids or whatever.

Absences play a central role within predictive processing accounts of vision. This is clearest in cases where the prediction that some higher-level stage, *H*, passes down to a lower-level stage, *L*, matches the incoming sensory signal *L* receives. In that event, *L* *does nothing*. It does not pass the sensory signal along to *H*, and because it does not, *H is never connected to the distal object by a physical process involving energy transference*. In a popular presentation of his work, Clark provides a helpful analogy:

Suppose you make a plan with your friend Duke by saying that if you *don’t* call him, then all is “as expected” and that he should therefore meet your plane at Miami airport next Wednesday at 9 a.m. local time. Your *failure* to call is then (technically speaking) a tiny little one-bit signal that conveys a large amount of neatly compressed information![[28]](#footnote-28)

Is your failure to call Duke a cause of his subsequently meeting your plane in Miami? *Yes*, say typical difference-making accounts of causation. *No*, say typical productive accounts.

Orlandi and Lee make the key point even more explicitly in a critical 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.[[29]](#footnote-29)

But what sort of “causal notion” is being invoked here, productive or difference-making? The claims being made in the passage seem correct if causation is understood as production, but not if it is understood as difference-making.

In the two sections that follow, I set out arguments in favor of the productive theory of perception. If the productive theory is correct, then higher-level stages within the visual system are not causally connected to distal objects, and so the Causal Condition on perception appears to be unsatisfied. Representing the idea schematically, suppose some external object *o* is *F*. Suppose *H*, a higher-level stage within *S*’s visual system predicts that *o* is *F*, and sends this prediction to *L*, a lower-level stage. Light reflected by *o*, as a result of its being *F*, reaches *S*’s retina and the resulting incoming sensory signal reaches *L*. At this point, *L* does not pass the signal along to *H*, given that *H*’s prediction was correct. But thus it follows that *H* is not physically connected to *o*’s being *F* by a causal process involving the transference of energy (or any other conserved quantity). If we continue to suppose that percepts are distributed across the hierarchy, it seems to follow that the overall percept is not causally connected to the distal object (we will question this step of the argument in §8, but for now I will adopt it uncritically). The Causal Condition is thus unsatisfied. *S*’s visual experience representing that *o* is *F* is therefore a veridical hallucination rather than a case of successful perception.

At this point we can add that the hallucination is *controlled* in the sense that if *o* had not been *F*, then *L* would have detected the mismatch with *H*’s prediction and sent *H* an error signal, nudging it back toward accuracy. That is, we can analyze the “controlled” in counterfactual terms, even as we analyze the Causal Condition in productive ones.[[30]](#footnote-30) We thus arrive at the predictive processing slogan: perception is controlled hallucination.

**6. Productive Thought Experiments**

At this point I have articulated the predictive theory of perception and explained what I take to be its consequences for predictive processing accounts of vision. I now turn to the positive case for the theory. My first line of argument appeals to the same sort of thought experiments used to motivate the causal theory of perception (§3), but now designed to tease apart the two concepts of causation. Consider the following twist on Grice’s **Clock** case, inspired by an example of Lewis’s.

**Censor**: A subject *S* has a visual experience as a of a clock on the shelf. There is in fact a clock on the shelf. *S* is wearing the very same brain-stimulating apparatus mentioned in Grice’s original **Clock** case, but the apparatus is now connected via Bluetooth to a clock-detector built into the shelf. The detector is such that when there is no clock on the shelf, the detector registers this and just sits there doing nothing, allowing the apparatus to simulate *S*’s brain so that *S* has a visual experience as of a clock. And when there is a clock on the shelf—as in the case we are considering—the detector registers this and sends a signal that delays the process by which the apparatus stimulates *S*’s brain, so that by the time the process runs to completion, *S* is already having a visual experience as of a clock in response to the light reflected by the clock and received by her eyes.[[31]](#footnote-31)

**Censor** is an instance of what is known as *late causal preemption* in the causation literature—“late” in that the preempted causal process (involving the apparatus, or would-be “censor”) runs all the way to completion.[[32]](#footnote-32) In the stock example, Billy and Suzy are throwing rocks at a bottle. Suzy’s rock gets there first, shattering the bottle. Productive theories of causation have no trouble delivering the intuitively correct verdict on the case: Suzy’s rock transfers energy and momentum to the bottle while Billy’s rock does not, and so Suzy’s throw counts as causing the shattering while Billy’s throw does not. But difference-making theories notoriously struggle with late preemption. If Suzy had not thrown her rock, the bottle still would have shattered (given Billy’s throw), and so Suzy’s throw apparently makes no difference to the bottle shattering. We seem to be forced into the absurd result that Suzy’s throw does not qualify as a cause of the bottle shattering.

Difference-making theorists have explored various solutions to the problem of late preemption, but none of their solutions is widely accepted or even regarded as very promising.[[33]](#footnote-33) My take on the state of the debate is that difference-making theorists should (and often do) simply concede that productive theories have an advantage on this point, at least for now (maybe a satisfying solution will be someday be found), while arguing that difference-making approaches have other, greater advantages. Philosophical theory selection can be understood as a kind of cost-benefit analysis, where the costs and benefits are reasons that speak against or in favor of a given view. The treatment of late preemption is a cost,

but there are benefits to a difference-making approach that outweigh this cost, defenders of the approach can try to argue.

Applying this general line to the case at hand, I take it to be intuitively obvious that the subject in **Censor** sees the clock, since the apparatus in this case (unlike in **Clock**) is causally preempted by the process of normal vision.[[34]](#footnote-34) I anticipate this verdict will be widely shared.[[35]](#footnote-35) Operating on the premise that *S* in **Censor** sees, it follows that the Causal Condition must be satisfied in the case. The productive theory of perception says that it is, since (we can suppose) energy is transferred from the clock to *S*’s retina to the neural state that realizes or is identical with *S*’s visual experience. And so the productive theory of perception gets **Censor** right. In contrast, the difference-making theory of perception seems to entail the Causal Condition is not satisfied, since even if there had been no clock on the shelf, *S* would have had a visual experience as of a clock on the shelf (since in that case the apparatus would be stimulating *S*’s brain, and would be doing so without having been delayed by the detector). Or at least, there is presently no widely accepted or very promising difference-making account of causation that yields the right result on **Censor**. I thus take late preemption cases like **Censor** to constitute a theoretical advantage of the productive theory of perception.

Again, I think proponents of the difference-making theory of perception should concede this point and try to argue that their view delivers benefits that outweigh this cost. What are these benefits? Well, in domains other than perception, difference-making theorists often point to their ability to handle absence causation in this context, so let’s consider a perceptual case involving absences.

**Detector**: A subject *S* has a visual experience as a of a clock on the shelf. There is in fact a clock on the shelf. *S* is wearing the very same brain-stimulating apparatus mentioned in Grice’s original **Clock** case, but the apparatus is now connected via Bluetooth to a clock-detector built into the shelf. The detector is such that when there is no clock on the shelf, the detector registers this and sends a signal that shuts off entirely the apparatus *S* is wearing, in which case *S* has no visual experience as of a clock. And when there is a clock on the shelf—as in the case we are considering—the detector registers the clock’s presence but sends no such signal, it just sits there doing nothing, allowing the apparatus to stimulate *S*’s brain so that *S* has a visual experience as of a clock.

**Detector** is the opposite of **Censor**, in that whereas **Censor** involved production without difference-making, **Detector** offers difference-making without production. There is counterfactual dependence in **Detector**: if the clock had not been on the shelf, *S* would not have had the visual experience as of a clock (since in that case the detector would have sent the shutoff signal). But there is no energy transferred from the clock to *S*’s visual experience, since the detector in this case operates by *not* sending the apparatus a signal—an absence.

I take it to be intuitively obvious that the subject in **Detector** does *not* see the clock. If more of an argument is wanted, start by assuming with Grice and others that the subject in the original **Clock** case does not see the clock on the shelf. But then, notice that **Detector** is just like **Clock** except we have added a detector that registers the presence of the clock on the shelf and… *does nothing* with that information. The detector does not send a signal to the apparatus or otherwise alter *S*’s brain. Indeed, we can suppose that *S* in **Detector** is intrinsically indiscernible from the hallucinating subject in **Clock**. The detector also does not alter the clock in **Detector**, which we can suppose is intrinsically indiscernible from the one in **Clock**. Finally, the detector does not create new spatiotemporal or other physical relations between *S* and the clock in **Detector** that do not obtain in **Clock**.[[36]](#footnote-36) But in that case, the addition of the detector seems entirely *extraneous* to the question of whether *S* sees the clock in **Detector**. The addition of the detector that does nothing just does not seem like the kind of thing that could transform the hallucinating subject from **Clock** into a successfully perceiving one in **Detector**. Therefore, the subject in **Detector** does not see the clock.

This result is predicted by the productive theory of perception but not by the difference-making theory. That is, if the Causal Condition is understood in productive terms, it is *not* satisfied in **Detector**, entailing that the subject does not see the clock—the intuitively correct result. In contrast, if the Causal Condition is understood in difference-making terms, then it *is* satisfied in **Detector**, and so either we will need to embrace the unpalatable result that the subject in the scenario successfully sees the clock or else we will need to find some other, yet to be determined basis for resisting the conclusion.[[37]](#footnote-37) That the productive theory predicts and neatly explains the intuitively correct verdict on the case is a theoretical (abductive) advantage of the theory, I say.[[38]](#footnote-38)

Thought experiments have their limits, perhaps. Happily, my present line of argument can be bolstered in a way that does not rely on them. **Detector** involves an event structure, (*c*, ~*d*, *e*), in which an initial positive event, *c*, makes a difference to a terminal positive event, *e*, via an intermediary event that is an absence, ~*d* (the negation symbol marks that it is an absence or “negative event”). Specifically, the clock on the shelf, *c*, makes a difference to *S*’s visual experience*, e*, via the intermediary of the detector not sending a shutoff signal, *~d*. This very same type of (*c*, ~*d*, *e*) event structure is found in those predictive processing cases that are our focus, in which a higher-level stage within the visual system makes a correct prediction and so receives no error signal back. In those cases, a distal object having the feature it does, *c*, makes a difference to the probability distribution represented by the higher-level stage, *e*, via the intermediary event that is the absence of an error signal, *~d*.

That scientists describe predictive processing models with the slogan that “perception is controlled hallucination” is, I claim, further abductive support for the productive theory of perception and for its treatment of (*c*, ~*d*, *e*) event structures. Productive theorists are able to look at such models and agree that Yes, they do make putative instances of perception seem rather hallucination-like—they do so by violating the Causal Condition, productively construed. Difference-making theorists owe us some sort of alternative account. In what sense is perception hallucination-like, if by their lights the Causal Condition is satisfied in those predictive processing scenarios in which the absence of an error signal is acting as a causal intermediary? Again, the productive theory of perception has a theoretical (abductive) advantage here, I say.

Now, Jonathan Schaffer has compellingly argued that event structures of the (*c*, ~*d*, *e*) type are often regarded as causal in various scientific and commonsense domains, and so productive theories that reject such “causation by disconnection” in such domains are forced into an embarrassing causal revisionism.[[39]](#footnote-39) For example, they are forced to deny that you can cause a victim to die by shooting them in the heart, given the role that various intermediary absences play in the process (e.g., the absence of oxygenated blood reaching the victim’s brain), or to deny that HIV causes death, given again the role of intermediary absences (an absence of functioning CD4+ T-cells).

As a causal concept pluralist, I am happy to grant much of Schaffer’s argument. For many domains, what we want is a difference-making concept of causation, I am willing to suppose. What I deny is that perception is such a domain. Perception is where mind makes contact with the world, requiring a *real connection* between perceptual experience and perceived object, a connection more robust and physically real than, say, the relation that obtains between the dead plant and the gardener hundreds of miles away, napping or playing with his kids or whatever. Or so says the productive theory of perception that I am in the process of defending.

Framing this in cost-benefit terms, I am willing to suppose that for many domains, the theoretical cost of not having a satisfying treatment of late preemption is more than offset by the theoretical benefits that difference-making approaches to causation have to offer, including that of being able to (*c*, ~*d*, *e*) event structures as causal.[[40]](#footnote-40) But in the domain of perception, this particular feature seems like no benefit at all, it seems like yet another cost of the difference-making approach. The worry for difference-making theorists is that when it comes to perception, they retain the familiar costs of difference-making accounts of causation while losing the benefits they hold in other domains.

Next consider the supposed perception of absences, which we might think of as involving event structures of the form (~*c*, *e*), where ~*c* is an absence that is putatively perceived while *e* is a subject’s perceptual experience. Sartre gives a famous example of supposedly seeing the absence of his friend Pierre in the café.

**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 absent from the café. If Pierre had been present, *S* would not have had the visual experience he did.[[41]](#footnote-41)

The description of Sartre’s experience in **Pierre** is “contentious” in that some deny—I deny—

That we have perceptual experiences of such absences. On my preferred view, absences exist and enter into (difference-making) causal relations, but they are *unobservables*, like electrons or genes. There may be a phenomenology attached to having one’s *desire* to meet Pierre frustrated, or to recognizing that one’s *belief* that Pierre would be here is false, or to some other non-perceptual mental state, but there is no distinctive visual phenomenology that consists in *seeing* Pierre’s absence from the café. The description in **Pierre** is meant to be neutral on this matter.

Pierre’s absence makes a difference to Sartre’s visual experience but does not produce it—it transfers no energy to Sartre. More generally, the productive theory of perception entails that absences are never perceived, while the difference-making theory can allow for absence perception. The question is which view has the upper hand on this point. One argument that favors the productive theory derives from the *problem of profligate causation*: once you admit some absences as causes, there is no principled way to avoid being stuck with lots of absences as causes, in ways that seem counterintuitive.[[42]](#footnote-42)

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 Joseph Stalin. And so each of their absences qualifies as a cause of Sartre’s visual experience according to standard counterfactual theories of causation. If you suppose in addition that Sartre’s visual experience can be described with just as much justice as an experience as of these absences as it can be described as an experience of Pierre’s absence—after all, it’s not as though Pierre’s absence has a certain color or shape the others lack[[43]](#footnote-43)—it follows that each of these other absences satisfy all three conditions of the difference-making theory of perception. But it is absurd to hold that Sartre sees these absences, or, generalizing more widely yet, the absence of every single person who is not present in the café. The productive theory of perception avoids this absurdity by denying that any absences satisfies the Causal Condition. With the difference-making theory of perception, in contrast, it is not clear how to avoid the result.

The problem of profligate causation is not unique to the domain of perception. It arises wherever absences are taken to be causes: the plant would not have died if the gardener had watered it, Yes, but the same is true if the Queen of England had, and so her omission too qualifies as a cause of death if causation is difference-making. However, the problem is worse for perception than it is for other domains. Perhaps the leading response to the problem, defended by both Lewis and Schaffer, concedes that absence causation is profligate but then appeals to linguistic pragmatics to explain why we cite only some absences as causes.[[44]](#footnote-44) For instance, we mention the gardener’s failure to water the plant as a cause of death but not the Queen’s failure because there is no expectation the Queen would water it, and so to speak of her omission imparts no information that is not already presupposed.

To deploy this strategy as a response to the problem for perception I have posed would mean conceding that, for instance, Sartre and all the rest of us really are perceiving Stalin’s absence, it’s just that we don’t talk about it. This is totalitarianism. Or, if not that, it is at least an implausible phenomenology. Nobody arrives at such a view as a result of attending to the phenomena themselves; you could only end up there if you are in the grips of some (causal) theory. In short, the problem of profligate causation carries phenomenological burdens in the perceptual domain that don’t have clear analogues in other domains. Indeed, if I were a difference-making theorist responding to the plant case, I would emphasize that the considerations that lead us to posit the Queen’s omission as a cause of death are abstract and theoretical in much the way you would expect for an unobservable entity. At any rate, the upshot is that within the domain of perception, allowing (~*c*, *e*) event structures to be causal is not obviously a great theoretical advantage, as it is perhaps in other domains.

To be sure, there are also some problem cases for the productive theory of perception on this front. At least on its face, the theory seems to entail that we do not see holes, or shadows, or black holes, or perfectly black objects, or other “perceptual ephemera” that seem not to reflect light.[[45]](#footnote-45) There may be moves available to handle these cases. For instance, perhaps productive theorists can maintain that we do not see holes, we see *through* them, or alternatively perhaps we can identify holes with material objects (e.g., hole-linings) that are capable of transmitting energy after all.[[46]](#footnote-46) And perhaps we can treat shadows in analogous fashion, on the basis that they are “holes in light.”[[47]](#footnote-47) And perhaps we can take a hard line on black holes, insisting that they really are invisible, recent reports of one having been photographed notwithstanding.[[48]](#footnote-48) And perhaps we can point out that the ordinary black objects we see in everyday life are not perfect blackbodies but instead reflect at least some incident light, while adding that perfectly black objects are at least not obviously visible.[[49]](#footnote-49)

I will not try to develop any of these lines of response here. My point is only that none of these problem cases is obviously fatal to the productive theory of perception. Running the cost-benefit analysis, I still am inclined to say that the topic of absence perception counts as a net benefit for the productive theory, but at the very least I claim that it incurs no cost on the theory so great that it outweighs the benefits of the theory regarding late preemption cases and (*c*, *~d*, *e*) event structures, including those at the heart of predictive processing accounts. Looking at the balance of reasons, I say that the productive theory of perception is coming out ahead.

**7. Receptivity & Spontaneity**

My second, related argument for the productive theory of perception is broadly Kantian, although the central idea can be detached from its Kantian inspiration. Kant took the mind to possess two fundamental capacities.[[50]](#footnote-50) First, there is its *receptivity*, or capacity to be affected by external objects. Second, there is its *spontaneity*, or capacity to initiate activity on its own, where this especially includes (but is not limited to) rational activity.

Traditional, non-Kantian views have often though of perception as something like pure receptivity, an entirely passive exercise whereby causal processes originating with external objects make their way into the mind and leave their mark. In perception, the mind acts as a “mirror of nature,” in Rorty’s phrase.[[51]](#footnote-51) This is supposedly in contrast with inference, understood as a paradigmatic exercise of spontaneity. In inference, the mind moves from certain beliefs taken as premises to new beliefs arrived at as conclusions, where this rational transition in thought need not be in response to any particular external trigger.

The problem with such traditional views of perception, according to the Kantian line of thought, is they miss the constructive role the mind plays in perception. Perception involves spontaneity, it involves (unconscious) inference, it is not just pure receptivity. Predictive processing accounts provide a way of trying to capture this constructive, inferential, Kantian element to perception.[[52]](#footnote-52) When a higher-level stage, *H*, makes its prediction about what the incoming sensory signal will be, this occurs prior to *H* being affected by the signal (otherwise it would not count as a *prediction*). This is the mind actively getting out ahead of the sensory signal rather than just passively receiving it. This is spontaneity. Add in that *H*’s prediction is made in accordance with rational, Bayesian principles, and the predictive processing view takes on a distinctively Kantian feel, something its proponents have often explicitly acknowledged.[[53]](#footnote-53) Neuroscientist Andreas Engel and his coauthors do not cite Kant, but in the following passage they put things in a way that fits the Kantian framework well:

Classical theories of sensory processing view the brain as a passive, stimulus-drive 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.[[54]](#footnote-54)

At this point, however, a new worry arises. In moving so sharply away from the traditional, passive, purely receptive understanding of perception, do predictive processing accounts represent a kind of overcorrection, at least from the Kantian view? That is, do predictive processing models build in so much spontaneity that receptivity drops out entirely? Again, in those cases in which *H* correctly predicts the incoming sensory signal, it *receives nothing*. No error signal from the lower-level stage, and so in turn no energy (or anything else) from the external world. But if *H* receives nothing, then in a very literal sense it cannot be manifesting a capacity for receptivity, just as a sugar cube cannot manifest its disposition of water-solubility in a scenario where it does not dissolve in water. And if there is no receptivity, there is no successful perception—at best there is veridical hallucination. Or at least this is what seems to follow if we take seriously the Kantian thought that perception requires *both* spontaneity *and* receptivity.

Imagine you are arrested and sent to prison. The guards cut off your contact with the external world, with the exception that they allow you to predict the letters your heartbroken family sends to the prison, and they promise to pass along whatever portions of those letters you predict wrongly. You get to work, writing out the letter you predict your poor mother would send, and then you hand your work to the guard, who responds with icy silence and a blank stare—congratulations, your prediction was entirely right. You *receive* nothing from your mother; you remain utterly disconnected from her and the outside world. And yet this is the lonely life of quiet desperation that higher-level stages of your visual system find themselves leading, imprisoned within your skull, if predictive processing accounts are correct.

In a few different writings, Hohwy has argued that a distinct feature of predictive processing accounts is that they open the door to global skepticism, an implication he accepts.[[55]](#footnote-55) I’m not sure this is right, but at any rate it is not the point I mean to press with the preceding analogy. Imagine you know that the prison’s message system is perfectly reliable, and so you know with complete confidence that the content of your mother’s undelivered letter really is just what you predicted. It still is the case that you have received nothing from here, that you remain utterly disconnected from her. Similarly, even if there were some sort of guarantee that your visual experience is veridically representing the world—suppose skepticism has finally been refuted somehow—it still would be the case that your experience is disconnected from the external world, your experience is a spontaneous anticipation of how the world will be rather than a receptive reaction to how it is.

No small part of the empirical case for predictive processing accounts is the light they seem to shed on various real-life cases of hallucination (as opposed to purely philosophical examples), for example in cases of schizophrenia or post-traumatic stress disorder, or in people suffering from Alzheimer’s or Parkinson’s diseases.[[56]](#footnote-56) The idea is that such hallucinations are the result of “strong priors,” or visual systems that assign too much weight to top-down predictions, and as a result fail to adjust properly when those predictions fail to match the incoming sensory signal—prediction errors are unable to keep the system on track, veridically representing the world.[[57]](#footnote-57)

Hallucinations so conceived are a form of spontaneity without receptivity. Perception is different in that in perception, prediction errors would nudge the system back on track if it were needed. But in those cases where the higher-level predictions are correct, which have been our focus, this seems not to be an assurance that perceptual experiences are in fact connected to the world in a way hallucinations are not, but rather an assurance that, counterfactually, perceptual experiences *would be* put in touch with the world if things had gotten out of control. Perception is therefore controlled hallucination.

Making explicit the argument for the productive theory, my claim is that successful perception requires a kind of Kantian receptivity and not just spontaneity. The manifestation of such receptivity requires that something literally has to be transmitted by the perceived object and received by the perceiver. To put it in the terms used in §6, there needs to be a real connection between the object and the perceiver, and not just a counterfactual assurance that a real connection would have been established if things had gone wrong. This demand that something be received rules out a role for absences, and so the Kantian view feeds naturally into the productive theory of perception rather than the difference-making theory.

**8. Responses**

In this concluding section, I consider some possible responses to the preceding arguments. First, you could simply regard the result that perception is controlled hallucination as a reductio of the productive theory of perception. The difference-making theory is able to allow that what we regard as paradigmatic cases of perception are in fact instances of (successful) perception rather than mere veridical, controlled hallucination, and you might think this benefit alone trumps whatever costs of the difference-making theory I managed to set out in §§6-7.

I am actually open to this response. Perhaps when reaching reflective equilibrium in our thinking about perception, we will settle on the difference-making theory. In that case, I would still regard this paper’s arguments as largely successful if they successfully captured what seems right *prima facie* about the predictive processing slogan that perception is controlled hallucination. In general, you can settle on the belief that *P* even as you continue to feel the pull of defeasible (and by your lights, defeated) reasons that favor *~P*. If, even as we were to embrace the difference-making theory of perception, we were to continue to feel the pull of the productive theory, and if in addition it would be because of this pull that we find it tempting to say the perception is controlled hallucination according to predictive processing accounts (although we resist this temptation), I would declare this at least a partial victory.

Second, you could go back and reject the assumption we made (in §5) that because percepts are distributed across the hierarchical arrangement of the visual systems rather than being located at any one stage, it follows that if some higher-level stage is not causally connected to the distal object, the percept itself is not so connected. When a baseball makes contact with a bat, not every square inch of it makes contact—the backside of the baseball does not. When a mind makes contact with the world in perception, perhaps not every stage within the visual hierarchy needs to enter into a productive causal relation with the perceived object. Perhaps it is good enough if the lower-level stage *L* when it receives the incoming sensory signal; it is not required in addition that *H* does.

This response aims to avoid the result that putative instances of perception are in fact but mere veridical, controlled hallucination, but it could be developed while granting that what seems prima facie right about the predictive processing slogan is the lack of a productive causal relation between the distal object and higher-level stages of the visual system. Substantial argument would be needed to establish that a satisfying account of perception can allow that, say, a percept’s representation of an object’s shape and color (which might be represented by a comparatively lower-level stage within the visual hierarchy) stand in a relation of real connection with the object, while the percept’s representation of the object’s category (which might be represented by a comparatively higher-level stage) does not but instead is disconnected from the object. But in principle I am open to this option and would regard it as vindicating my position in this paper.

Third, you could try to appeal to cases in which higher-level stages get their predictions wrong to secure productive causal relations in cases where they get their predictions right. Imagine a case in which *H* receives an error signal upon making a prediction it gets wrong. Receiving this error signal puts *H* in a productive causal relation with the distal object—there is a physical process linking the two via the transference of energy. *H* then revises its probability distribution in response to this error signal and makes a new prediction, which turns out to be correct and so elicits no new error signal. The idea is then that even in this later case *H* stands in productive causal relation with the distal object, since its (accurate) prediction is made in response to its receiving an error signal that productively connects it to the object.

Argument would be needed to establish that a satisfying account of perception can allow for this kind of temporal lag in receptivity. In addition, the proposal is hostage to empirical fortune, in that it would need the empirical details to work out right if the goal is to dodge the conclusion that perception is controlled hallucination. But again I am open to the option in principle. It concedes that the productive theory of perception is correct, and once again would appear to be compatible with granting that my discussion captures what seems at least prima facie right about the claim that perception is controlled hallucination.

Finally, you could reject predictive processing accounts of perception. I also can be open to this, provided it is combined with an acceptance of the productive theory of perception and an agreement that the correctness of the theory explains why the predictive processing account entails that perception is controlled hallucination.

REFERENCES

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

Clark, A. 2012. “Do Thrift Brains Make Better Minds,” *The New York Times*, URL =

<https://opinionator.blogs.nytimes.com/2012/01/15/do-thrifty-brains-make-better-minds/>

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.

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.

Dowe, P. 2004. “Causes are Physically Connected to their Effects: Why Preventers and

Omissions are not Causes,” in *Contemporary Debates in Philosophy of Science*, ed. C. Hitchcock, pp. 189-196. Oxford: Blackwell Publishers.

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.

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.

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

Supplementary Volume 35: 121–52.

Griffiths, T. L., Kemp, C. and 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, and L. A. Paul, pp. 225-276. Cambridge, MA: The MIT Press.

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,” *Nous*, 50.2: 259-285.

Kim, J. 2007. “Causation and Mental Causation,” in *Contemporary Debates in*

*Philosophy of Mind*, eds. B. P. McLaughlin and J. Cohen, pp. 227-42. Oxford: Blackwell Publishers.

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

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

Le Catt, B. 1982. “Censored Vision,” *Australasian Journal of Philosophy*, 60.2: 158-162.

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. 1986. “Postscripts to Causation,” in *Philosophical Papers: Volume II*, pp.

172-213. Oxford: Oxford University Press.

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

Lewis, D. K. 2004. “Void and Object,” in *Causation and Counterfactuals*, eds. J. Collins, N.

Hall, and L.A. Paul, pp. 277-290. Cambridge: MIT Press.

Loewer, B. 2007. “Mental Causation, or Something Near Enough,” in *Contemporary*

*Debates in Philosophy of Mind*, eds. B. P. McLaughlin and J. Cohen, pp. 243-64. Oxford: Blackwell Publishers.

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

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

McLaughlin, B. P. 1996. “Lewis on What Distinguishes Perception from Hallucination,” in

*Perception*, ed. K. Akins, pp. 198-231. Oxford: Oxford University Press.

Menzies, P. 1996. “Probabilistic Causation and the Pre-emption Problem,” *Mind*, 105: 85–117.

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.

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

*Philosophy Compass*, 4.1: 134-154.

Orlandi, N. and G. Lee. 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. and N. Hall. 2013. *Causation: A User’s Guide*. Oxford: Oxford University Press.

# Rao, R. P. and 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**.**

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. *A Probabilistic Theory of Causality*. Amsterdam: North Holland Publishing Company.

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

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

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. and T. Metzinger. 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. See for instance Hohwy (2013: 208), Clark (2016: 14), and Wiese and Metzinger (2017: 3). Wiese and Metzinger (2017: 3, n. 6) note that Horn (1980) attributes the slogan to artificial intelligence researcher Max Clowes, but report that no published remarks by Clowes seem to match. Grush (2004: 395) attributes the line to neuroscientist Ramesh Jain; Clark (2016: 308, n. 3) writes that it is also attributed to 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, etc. [↑](#footnote-ref-1)
2. Here and throughout, I assume that if a mental state is an instance of hallucination then it is *not literally* an instance of (successful) perception. I still will make free use of the slogan that “perception is controlled hallucination,” but I mean this as shorthand for the more careful formulation just given in the text: what seem to be paradigmatic instances of perception are instead cases of controlled hallucination. [↑](#footnote-ref-2)
3. The discussion that follows draws on Hohwy (2013), Clark (2013) and (2016), Wiese and Metzinger (2017), and Orlandi and Lee (2019). My discussion self-consciously adopts the bad philosophical habit of equating perception with vision, ignoring the other senses. My excuse is that leading discussions of the slogan that “perception is controlled hallucination” focus primarily on vision, and it will be helpful at points to put things in their terms. [↑](#footnote-ref-3)
4. 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-4)
5. *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-5)
6. Marr (1982: 19-29). [↑](#footnote-ref-6)
7. Rescorla (2017) objects to Clark (2016) on the basis that the evidence for the algorithmic-level predictive coding account is much weaker and more speculative than the evidence for the computational-level hierarchical Bayesian account; see also Orlandi and Lee (forthcoming). This is a point we will largely 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-7)
8. On predictive coding, see for instance Rao and Ballard (1999), Lee and Mumford (2003), Friston (2009), and Friston & Kiebel (2009). [↑](#footnote-ref-8)
9. Clark (2016: 26). [↑](#footnote-ref-9)
10. Frith (2007: 135), emphasis added. [↑](#footnote-ref-10)
11. Hohwy (2007: 322), emphasis added. [↑](#footnote-ref-11)
12. Hohwy (2013: 138). Hohwy (2016) and (2017) pursues a related line of thought to argue that predictive processing accounts invite skepticism, a conclusion he embraces (with qualifications). [↑](#footnote-ref-12)
13. Clark (2016: 195) insists that predictive processing accounts 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 to resist views like those advanced by Frith and Hohwy. [↑](#footnote-ref-13)
14. This is in contrast with *direct realism*, which takes us to be directly aware of external objects in perception. [↑](#footnote-ref-14)
15. There is of course an *argument from hallucination* for indirect realism, but its claim is not that perception is in fact hallucination; its claim is that perception and hallucination share a common factor (e.g., sense data, or perhaps the brain’s models or best hypothesis) , and that this common fact is the object of direct awareness. [↑](#footnote-ref-15)
16. For discussion of Descartes’s and Locke’s indirect realism, see Newman (2009). [↑](#footnote-ref-16)
17. Classic defenses of the causal theory include Grice (1961), Strawson (1974), and Lewis (1980); for a recent discussion. Note that the three conditions are meant to be individually necessary but not jointly sufficient for successful perception. On the challenges of developing the theory further so that it specifies necessary and sufficient conditions, see Arstila and Pihlainen (2009). [↑](#footnote-ref-17)
18. Grice (1961). [↑](#footnote-ref-18)
19. Hall (2004). [↑](#footnote-ref-19)
20. On counterfactual theories, see Lewis (1973) and (2000); on probability-raising theories, Suppes (1970) and Eels (1991); on interventionist approaches, Woodward (2003), on causal models, Hitchcock (2018). [↑](#footnote-ref-20)
21. Lewis (1973). Lewis (2000) is a significant revision to the original 1973 theory. [↑](#footnote-ref-21)
22. 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) influential causal process theory. [↑](#footnote-ref-22)
23. Dowe (2001) applies the term “quasi-causation” to such difference-making relations. [↑](#footnote-ref-23)
24. Hall (2004: 268). [↑](#footnote-ref-24)
25. There is a parallel question in the mental causation debate. Kim (2007) argues that what we want is productive mental causation, while Lower maintains that difference-making mental causation is good enough. Their focus is cases in which mental events cause physical effects (action), while my focus is on cases in which physical events cause mental effects (perception). [↑](#footnote-ref-25)
26. On whether absences can be causes, see for instance Hart and Honoree (1985: 38); Lewis (1986a), (2000) and (2004); Dowe (2000) and (2004); Schaffer (2000) and (2004); Hall (2004), Paul and Hall (2013); Menzies (2004). [↑](#footnote-ref-26)
27. Fair (1979) is an exception. [↑](#footnote-ref-27)
28. Clark (2012), emphasis added. [↑](#footnote-ref-28)
29. Lee and Orlandi (forthcoming). In the passage, the authors are describing a gloss on predictive processing models that they do not themselves accept. [↑](#footnote-ref-29)
30. This is no contradiction or embarrassment for the productive theory of perception. Cf. Dowe (2001), who defends a counterfactual theory of prevention and “quasi causation” without giving up his physical theory of causation. [↑](#footnote-ref-30)
31. Lewis (1980: 248). I use Lewis’s name for the case. [↑](#footnote-ref-31)
32. For discussion of late preemption, see Lewis (1986a), Hall (2004), Paul and Hall (2013: 99-143). [↑](#footnote-ref-32)
33. According to Paul and Hall (2013: 143) of all the problems they consider in their book, late preemption “appears to present the most stubborn obstacles to constructing an adequate philosophical account of causation.” [↑](#footnote-ref-33)
34. *Objection*: Have you not heard? Perception is controlled hallucination, per predictive processing accounts. If those accounts are right then *S* in **Censor** does not see the clock, but at most veridically hallucinates it. *Reply*: Ignore predictive processing accounts for now! Or, suppose this is a case in which the higher-level stages within *S*’s visual experience make incorrect predictions, and so the lower-level stages send them error signals, ensuring that the higher-level stages are connected to the distal object by a physical process involving the transference of energy. [↑](#footnote-ref-34)
35. Lewis (1980) denied that the subject in his version of **Censor** sees, but this is widely regarded as an embarrassment of his view; for critical discussion, see McLaughlin (1996). Lewis later acknowledged the point by writing a follow-up piece in which he adopted his cat’s name as a pseudonym and, after acknowledging that the “great majority” of philosophers reject his original verdict on the case, tinkered with his counterfactual theory to try to show that a version could secure the verdict that the subject in the case successfully sees; this is in Bruce Le Catt (1982). Lewis (1986b) subsequently rejected his cat’s theory. Without engaging in the sort of detailed analysis that would be needed to show that Le Catt’s account is still subject to late preemption counterexamples, I want to suggest that cats authoring papers that their owners distance themselves from does not seem like a sign that the difference-making theory of perception is a healthy, thriving research program. [↑](#footnote-ref-35)
36. Compare the discussion in Menzies (1996) of whether causation is an *intrinsic relation* in the sense that it is determined by the intrinsic properties of its relata together with the natural relations that hold between them. [↑](#footnote-ref-36)
37. Notice that the mere fact that **Detector** involves a brain-stimulating apparatus does not by itself entail that *S* does not see the clock, and so difference-making theorists cannot point to this point alone to resist the conclusion. After all, *prosthetic vision* would seem to be possible in principle, as Lewis (1980) emphasizes in the development of his difference-making version of the causal theory of perception. [↑](#footnote-ref-37)
38. Because the three conditions of the causal theory of perception are not meant to be jointly sufficient, the premise that *S* in **Detector** does not see the clock does not deductively entail the negation of the difference-making theory of perception. We can still think the productive theory provides a better explanation of the truth of the premise than the difference-making theory does, however. [↑](#footnote-ref-38)
39. Schaffer (2000) and (2004). [↑](#footnote-ref-39)
40. It may be that late preemption cases do not arise in some domains. Hitchcock (2013) argues that they do not within causal decision theory, which would support Hall’s (2004) suggestion that the difference-making concept is what we want there. [↑](#footnote-ref-40)
41. Sartre (1956: 40-41). See Farrenikova (2013) and Sorensen (2008) and (2015) for recent defenses of absence perception that discuss Sartre’s case. [↑](#footnote-ref-41)
42. The name for the problem is due to Menzies (2004). [↑](#footnote-ref-42)
43. Or, suppose the contents of visual experience are sets of centered possible worlds. The set that is the content of Sartre’s experience presumably includes no world in which Wellington, Valéry, or Stalin are present in the café, just as it includes no world where Pierre is. [↑](#footnote-ref-43)
44. Lewis (2000: 196), Schaffer (2000: 295). [↑](#footnote-ref-44)
45. On the perception of holes see Casati and Varzi (1994: Ch. 11); on shadows, Sorensen (2011); on black holes, Goldman (1977); and on perfectly black objects, Tye (1982). Each of these authors makes use of a difference-making approach to allow for the perception of these objects. The term “perceptual ephemera” is taken from the Crowther and Cumhaill (2018) collection of essays. [↑](#footnote-ref-45)
46. On the proposal that holes are hole-linings, see Lewis and Lewis (1970). [↑](#footnote-ref-46)
47. The phrase is taken from Casati and Cavanagh (2019: 1). [↑](#footnote-ref-47)
48. “By its very nature, a black hole cannot be seen,” says NASA’s own website, reporting on the Event Horizon Telescope taking the “first photograph of a black hole,” as we might call it if we were to speak with the vulgar. What actually appears in the photograph is not the black hole itself, according to NASA, but the hot disk of material encircling it. [↑](#footnote-ref-48)
49. “Scientists develop a material so dark that *you can’t see it*” (emphasis added) says the headline of the tabloid online newspaper *The Independent*, speaking with profound philosophical wisdom. The article is about Vantablack, a material that absorbs 99.6% of visible light, making it one of the darkest substances known. See: Johnston (2014). [↑](#footnote-ref-49)
50. Kant (1998). [↑](#footnote-ref-50)
51. Rorty (1979). Hohwy (2013: Ch. 11) and Clark (2016: 305) both use Rorty’s term to mark a contrast with the predictive processing view. [↑](#footnote-ref-51)
52. Within contemporary epistemology, the view that perception requires Kantian spontaneity and not just receptivity is especially associated with McDowell (1994). McDowell connects the view to his thesis that perception has conceptual rather than nonconceptual content, but we can leave that thesis aside here. [↑](#footnote-ref-52)
53. See especially Swanson (2016). Passing references to Kant are also made in Hohwy (2013), Clark (2016), and Wiese and Metzinger (2017). Swanson suggests that there is a direct line of influence running from Kant to Helmholtz (1925) to predictive processing accounts, via Helmholtz’s Kant-inspired claim that perception involves “unconscious inference.” [↑](#footnote-ref-53)
54. Engel et al. (2001: 704). [↑](#footnote-ref-54)
55. Hohwy (2013), (2016), and (2017). [↑](#footnote-ref-55)
56. See for instance Friston (2005), Corlett et al. (2009), Fletcher and Frith (2009), Corlett et al. (2019). [↑](#footnote-ref-56)
57. On the flipside, it has been proposed that cases of autism spectrum disorder involves too much weight being assigned to prediction errors and not enough to top-down predictions, resulting in subjects who lose the incoming signal for the noise. See Friston et al. (2012). [↑](#footnote-ref-57)