**Perception as Controlled Hallucination**

“Perception is controlled hallucination,” according to certain proponents of *predictive processing* accounts of vision.[[1]](#footnote-1) I say they are right that something like this is a consequence of their view, but wrong in how they have developed the idea. In what follows, I advance my own analysis. In the process, I argue that the *causal theory of perception* should be understood in terms of a *productive* concept of causation, as opposed to a *difference-making* concept. On my view, predictive processing accounts entail that various putative instances of successful perception are instead cases of *veridical hallucination* because they do not satisfy the causal condition on perception, understood productively.[[2]](#footnote-2)

**1. Predictive Processing**

Predictive processing accounts can be understood as involving a hierarchical Bayesian computational model of vision implemented by a predictive coding algorithm.[[3]](#footnote-3) 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, 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.

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.[[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 is updated in accordance with Bayes’ rule, stating 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 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 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 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 a 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 the preceding is right. 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 “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 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 might instead be 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 disquieting conclusion that perception is controlled hallucination.[[15]](#footnote-15)

. In fact, if the import is just indirect realism, then all the 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. I will focus on one of the more famous ones, due to H. P. Grice.[[18]](#footnote-18)

**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. I take this verdict to be intuitively compelling and will not question it.

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. The subject does not successfully see the clock, she merely veridically hallucinates it.

I claim that if predictive processing accounts are correct, external 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 decent chance that your visual experience is *just like* that of the subject in the case of Grice’s **Clock**, at least in the relevant causal respects. To defend this claim, I now turn to competing accounts 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.

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 not often been careful to assess which causal concept should be operative in a given analysis. 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 of perception*?[[23]](#footnote-23)

**5. Absences**

Although the two concepts of causation often coincide, one place where they come apart is *absences*.[[24]](#footnote-24) 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 process involving the transference of a conserved physical quantity connecting the gardener to the plant. The gardener, we can even imagine, is hundreds of miles away, napping.

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![[25]](#footnote-25)

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.

Nico Orlandi and Geoff 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.[[26]](#footnote-26)

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 following section, I argue in favor of the productive theory of perception. If the productive theory is correct, then higher-level stages within the visual system are often not causally connected to distal objects, and so the Causal Condition on perception is 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 is reflected by *o*, as a result of its being *F*, and reaches *S*’s retina; the resulting incoming sensory signal reaches *L*. At this point, *L* does not pass the sensory 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). Given again that percepts are distributed across the visual hierarchy as opposed to located at any one stage, it seems to follow that the overall percept is not causally connected to the distal object, and so the Causal Condition is unsatisfied. *S*’s visual experience representing that *o* is *F* is therefore a veridical hallucination.

At this point, we can add that the hallucination is *controlled* in the sense that if *o* had not been *F*, *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” notion in counterfactual terms, even as we analyze the Causal Condition in productive ones.[[27]](#footnote-27) We thus arrive at the predictive processing slogan: perception is controlled hallucination.

**6. Productive Thought Experiments**

At this point I have articulated the productive theory of perception and explained what I take to be its consequences. I now turn to the positive case for the theory, which involves appealing to the same sort of thought experiments often used to motivate the causal theory of perception, but now designed to tease apart the two concepts of causation.

In general, difference-making theories of causation struggle with what is known as *late causal preemption*—“late” in that the preempted causal process runs all the way to completion.[[28]](#footnote-28) 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 to get it right. 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. Difference-making theorists thus seem forced into the absurd result that Suzy’s throw does not qualify as a cause of the shattering. Various solutions to the problem of late preemption have been explored, but it is fair to say that none are widely accepted or even regarded as very promising.[[29]](#footnote-29)

You can construct a late preemption scenario for any domain you happen to be theorizing about, and so you can construct one for perception. Here is one that involves a 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. But 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.[[30]](#footnote-30)

I take it to be intuitively obvious that the subject in **Censor** sees the clock, a verdict that I anticipate will be widely shared.[[31]](#footnote-31) (The idea is meant to be that the apparatus would have caused *S* to veridically hallucinate a clock on the shelf, just as in Grice’s original case, but it is delayed from doing so by the detector. Given this delay to the apparatus, *S* is able to successfully see the clock on the shelf using normal vision.) If this verdict is correct, it follows that the Causal Condition must be satisfied in the case. The productive theory of perception agrees that it is, since (we can suppose) energy is transferred from the clock to the neural realizer of *S*’s visual experience. In contrast, the difference-making theory of perception seems to entail that it is not, since even if there had been no clock on the shelf, *S* still would have had a visual experience as of a clock (since in that case the apparatus would be stimulating *S*’s brain, and would be doing so without delay). I thus take the productive theory of perception to have an advantage over the difference-making theory in its treatment of late preemption cases.

In my view, difference-making theorists should just concede this point. That is, they should grant that their general struggles with late preemption arise in the special case of perception as well, and agree that this is a theoretical cost of their position. But they then should devote their energy to arguing that their approach also has various theoretical benefits that outweigh this cost. (Concede the battle to win the war.) One such potential benefit involves the ability to handle absence causation, so let’s consider a perceptual cases involving an absence.

**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 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 while **Censor** involved production without difference-making, **Detector** involves 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 the case operates by *not* sending the apparatus a signal—an absence.

I regard it as intuitively compelling that the subject in **Detector** does *not* see the clock. Here is an argument, if you don’t share my intuition. Start by assuming with Grice and others that the subject in the original **Clock** case does not see the clock on the shelf. Well, **Detector** is just like **Clock** except it includes a detector that registers the presence of the clock on the shelf and… *does nothing with this information*. The detector does not signal the apparatus or otherwise alter *S*’s brain, it does not alter the clock, it does not create new spatiotemporal or other physical relation between *S* and the clock that do not obtain in **Clock**.[[32]](#footnote-32) It just sits there, doing nothing. But in that case, the do-nothing detector seems entirely *extraneous* to whether *S* sees the clock in **Detector**. Taking the original **Clock** case and adding an inert detector to the scenario just does not seem like the kind of thing that could transform a hallucinating subject into a successfully perceiving one. And so I conclude that the subject in **Detector** does not see the clock.

**Detector** involves an event structure of the form (*c*, ~*d*, *e*), in which an initial positive event *c* (the clock’s presence on the shelf) makes a difference to a terminal positive event *e* (*S*’s visual experience) via an intermediary absence or negative event ~*d* (the detector not sending a shutoff signal). The same type of (*c*, ~*d*, *e*) event structure is found in predictive processing cases, where a distal object having the feature it does, *c*, makes a difference to a higher-level stage’s representation of the world, *e*, via the absence of an error signal as an intermediary, ~*d*. **Detector** is meant to be an analogue to the predictive processing case.

Jonathan Schaffer has compellingly argued that event structures of this (*c*, ~*d*, *e*) form are often regarded as causal in various scientific and commonsense domains, and so productive theories that deny such “causation by disconnection” in such domains are forced into an embarrassing causal revisionism.[[33]](#footnote-33) 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).

As a causal concept pluralist, I am happy to grant Schaffer that for many domains, what we want is a difference-making concept that allows for absences as intermediaries. 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 one that obtains between the dead plant and the gardener napping hundreds of miles away. The intuitive judgment that the subject in **Detector** does not see the clock, together with the judgment of predictive processing proponents that their own theory entails that perception is controlled hallucination, fits well with the productive theory I am defending.

Next consider event structures of the distinct form (~*c*, *e*), where ~*c* is an absence that is supposedly perceived and *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.[[34]](#footnote-34)

The description of the experience is “contentious” in that some deny we have perceptual experiences as of absences at all; I mean to leave the matter open in describing the case. Pierre’s absence makes a difference to Sartre’s visual experience (as the final sentence of **Pierre** entails) but does not produce it—absent Pierre transmits no energy to Sartre. More generally, the productive theory of perception entails that absences are never perceived, while the difference-making theory allows for absence perception. The question is which view has the upper hand here.

One argument that favors the productive theory derives from the problem of *profligate causation*: if there is any absence causation at all, there is a lot of it.[[35]](#footnote-35) 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. 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—after all, it’s not as though Pierre’s absence has a certain color or shape the others lack[[36]](#footnote-36)—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 all these absences, or (generalizing) that he sees the absence of everyone in the entire world who is not present in the café. The productive theory of perception avoids this absurdity by denying that any absences satisfies the Causal Condition—by denying any absence is seen.

The problem of profligate causation is not unique to 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 also qualifies as a difference-making cause of death. However, the problem seems worse for perception than it does for other domains.

Arguably the leading difference-making response to the problem concedes that absence causation is profligate but then appeals to pragmatics to explain why we cite only some absences as causes.[[37]](#footnote-37) So for instance, we do not mention the Queen’s failure to water the plant as a cause of death because nobody expected her to water it, and so citing it in conversation would impart no new information not already presupposed. I am willing to suppose that for many domains, this is a perfectly good solution to the problem. Not for perception though. To deploy the strategy in the case of perception would be to concede that Sartre really is perceiving every single person’s absence, he just has pragmatic reason not to mention the vast majority of them. But the notion that Sartre and all the rest of us are always and forever perceiving Stalin’s absence, it’s just that we have good pragmatic reason not to talk about it, is *totalitarianism*. Or, if not that, it is at least a phenomenologically implausible position. Nobody would arrive at such a view simply as a result of “returning to the phenomena themselves.” You could only get there if you were in the grips of a (causal) theory.

In short then, I claim that there are phenomenological costs to embracing profligate absence causation in the perceptual domain, costs that do not arise in other domains. There is no parallel phenomenological burden to saying the Queen’s omission caused the plant’s death, and so as a result the pragmatics-based solution seems more promising there than it does for perception. In fact, I think this feeds into an attractive view of absences as *unobservables*, like electrons and genes, that we posit to causally explain various observed effects (gunshot deaths, dead plants, other things that have nothing to do with death), even though we can’t see them. The productive theory of perception naturally supports such a view, as long as it is not combined with a productive theory of causation across all domains.

In conclusion, here is where I think things stand. Late preemption cases like **Censor** place a clear theoretical cost on the difference-making theory of perception. If you are a difference-making theorist, you might have hoped this cost would be outweighed by other benefits, but cases involving absences like **Detector** and **Pierre** seem if anything just to impose *further costs* on your view. This discussion has been too brief to be conclusive, but I think it has set out at least a pro tanto reason to accept the productive theory of perception. And so in turn it has given us a reason to agree that if predictive processing accounts are correct, perception is but controlled (veridical) 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 philosophical bad 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 (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-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) 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 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 subjects 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 factor 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 (1979), and Lewis (1980). Note that the three conditions are meant to be individually necessary but not jointly sufficient for successful perception. [↑](#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), and 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. There is a parallel question in the mental causation debate. Kim (2007) argues that what we want is productive mental causation, while Lower (2007) maintains that difference-making mental causation is good enough. Their focus is on 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-23)
24. On whether absences can be causes, see for instance Lewis (1986), (2000) and (2004); Dowe (2000) and (2004); Schaffer (2000) and (2004); Hall (2004), Paul and Hall (2013); Menzies (2004). [↑](#footnote-ref-24)
25. Clark (2012), emphasis added. [↑](#footnote-ref-25)
26. Lee and Orlandi (forthcoming). In the passage, the authors are describing a gloss on predictive processing models that they do not themselves endorse. [↑](#footnote-ref-26)
27. 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-27)
28. For discussion of late preemption, see Lewis (1986), Hall (2004), Paul and Hall (2013: 99-143). [↑](#footnote-ref-28)
29. 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-29)
30. Lewis (1980: 248). [↑](#footnote-ref-30)
31. Lewis (1980) denied that the subject in his version of **Censor** sees, but this is generally regarded as an embarrassment of his view; for critical discussion, see McLaughlin (1996). Lewis later granted 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; see Bruce Le Catt (1982). [↑](#footnote-ref-31)
32. Cf. Menzies (1996), who in his defense of a productive account of causation claims that causation is an *intrinsic relation* in the sense that it is determined by the intrinsic properties of the relata together with the natural relations they enter into. Difference-making theorists often reject the claim that causation is intrinsic—see Lewis (2000). [↑](#footnote-ref-32)
33. Schaffer (2000) and (2004). [↑](#footnote-ref-33)
34. Sartre (1956: 40-41). See Farrenikova (2013) and Sorensen (2008) for recent defenses of absence perception that discuss Sartre’s case. [↑](#footnote-ref-34)
35. The name for the problem is due to Menzies (2004). [↑](#footnote-ref-35)
36. 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-36)
37. Lewis (2000: 196), Schaffer (2000: 295). [↑](#footnote-ref-37)